

# EUR Research Information Portal

## Unity, plurality and explanation. The case of geographical economics and its neighbours

### Publication status and date:

Published: 11/11/2005

### Document Version

Publisher's PDF, also known as Version of record

### Citation for the published version (APA):

Marchionni, C. (2005). *Unity, plurality and explanation. The case of geographical economics and its neighbours*. [Doctoral Thesis, Erasmus University Rotterdam]. Erasmus Universiteit Rotterdam (EUR).

[Link to publication on the EUR Research Information Portal](#)

### Terms and Conditions of Use

Except as permitted by the applicable copyright law, you may not reproduce or make this material available to any third party without the prior written permission from the copyright holder(s). Copyright law allows the following uses of this material without prior permission:

- you may download, save and print a copy of this material for your personal use only;
- you may share the EUR portal link to this material.

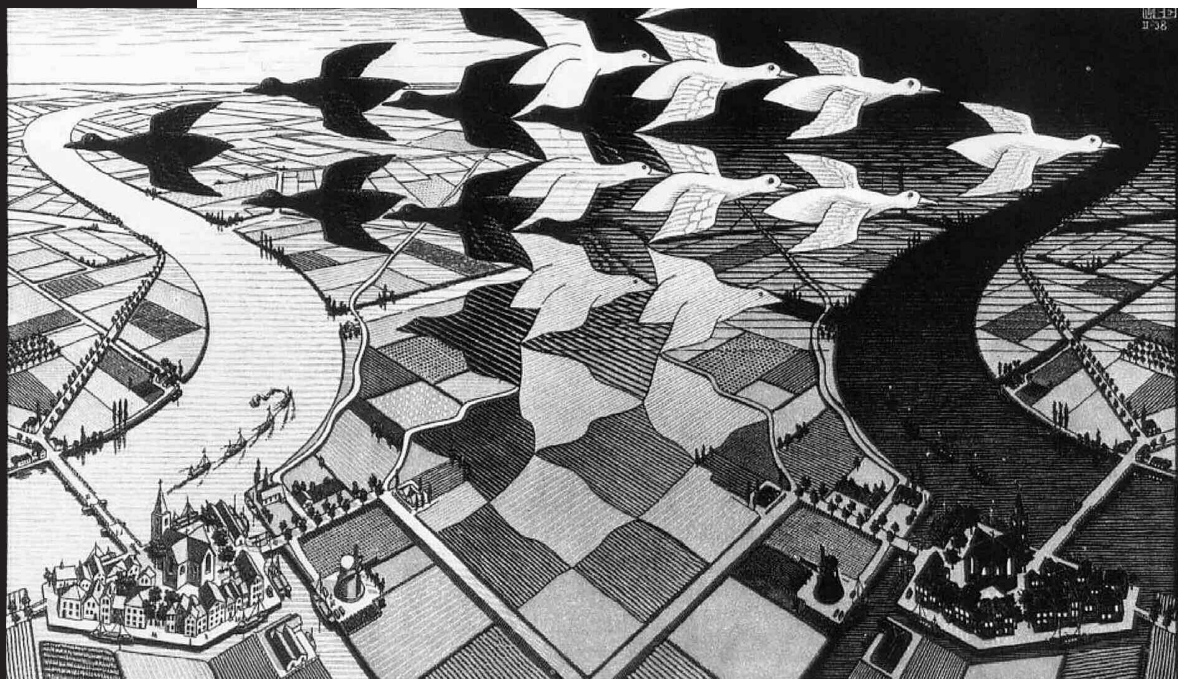
In case the material is published with an open access license (e.g. a Creative Commons (CC) license), other uses may be allowed. Please check the terms and conditions of the specific license.

### Take-down policy

If you believe that this material infringes your copyright and/or any other intellectual property rights, you may request its removal by contacting us at the following email address: [openaccess.library@eur.nl](mailto:openaccess.library@eur.nl). Please provide us with all the relevant information, including the reasons why you believe any of your rights have been infringed. In case of a legitimate complaint, we will make the material inaccessible and/or remove it from the website.

# UNITY, PLURALITY AND EXPLANATION

The case of geographical economics  
and its neighbours



Caterina Marchionni

# **UNITY, PLURALITY AND EXPLANATION**

*The case of geographical economics and its  
neighbours*

ISBN 90-9019966-7

© Caterina Marchionni, 2005

All rights reserved. No part of this book may be reproduced in any form or by any electronic or mechanical means (including information storage and retrieval) without permission in writing from the author.

See <http://eps.eur.nl/dissertations> for the electronic version of this thesis

Printed by Haveka BV, Alblasterdam, The Netherlands

# UNITY, PLURALITY AND EXPLANATION

*The case of geographical economics and its  
neighbours*

Eenheid, verscheidenheid en verklaring  
Geografische economie en haar bureu

Thesis

to obtain the degree of Doctor from the  
Erasmus University Rotterdam by command of the  
rector magnificus

Prof.dr. S.W.J. Lamberts

and in accordance with the decision of the Doctorate Board

The public defence shall be held on  
Friday 11 November at 11.00 hrs

by

Caterina Marchionni  
born in Padova, Italy

Doctoral Committee

**Promotor:**

Prof.dr. I.U. Mäki

**Other members:**

Prof.dr. S. Psillos

Prof.dr. G.A. van der Knaap

Prof.dr. C. van Marrewijk

**Copromotor:**

Dr. P. Oinas

*To Daniela, Franco and Stella*





# Contents

Preface and acknowledgements	ix
1 Unity, plurality and explanation: The questions, the premises, the contributions	1
2 On the structure of explanatory unification: The case of geographical economics	29
3 Geographical economics versus economic geography: Towards a clarification of the dispute	61
4 The many explanations of spatial clustering: Rival or complementary?	93
5 Geographical economics models: Contrastive approach in action	133
References	165
Appendix	181
Tables and figures	187
Summary (Dutch)	191
Curriculum Vitae	194



## **Preface and acknowledgments**

Geographical economics is a recent approach developed within economics. It aims to provide a unified framework for the study of spatial agglomeration: the spatial concentration of economic activity. This search for unity stands in sharp contrast with the present plurality of approaches which seek to explain the same phenomenon across a variety of disciplines such as economics, economic geography, and strategic management. More than any of the existing approaches, geographical economics has been contested by proponents of alternative theories.

This dissertation investigates the general issues of unity, plurality and explanation by analysing them in relation to the case of geographical economics and its neighbours. The thesis is comprised of four independent essays and addresses these issues from diverse perspectives. In addition, a fifth introductory essay maps the common philosophical premises, outlines the overarching philosophical framework, and puts forth the overall contributions of this work.

Academic research is to a great extent a social activity and an impressive amount of people have, in one way or another, contributed to its accomplishment.

Following a loosely chronological recollection of my intellectual debts, I owe much of my choice to continue studying and researching to Prof. Francesca Gambarotto. She has been a wonderful supervisor when I was writing my master's thesis, but also an endless source of inspiration during the PhD years. Above all, I am grateful for her enduring friendship and support, for her precious advise that has been a vital factor in maintaining my enthusiasm and positive mental health throughout the PhD years.

The Erasmus Institute for Philosophy and Economics (EIPE) is more than the sum of its parts, so while I wish to thank each of its constituent parts individually, a special thank goes to the whole institute. It has been a friendly and inspiring environment to work in and has often felt like a second home. As for the individuals, I wish to thank John Davis, John Groenewegen, Wade Hands, Gregory Dow, Mark Blaug, Jack Vromen, Uskali Mäki and Arjo Klamer for their very inspiring courses. I would also like to thank Loes van Dijk for her help in dealing with practical matters. A separate thank you goes to Arjo who has been the first to be brave enough to give me the opportunity to challenge myself in dealing with undergraduate students and their papers.

Emrah Aydinonat, Francesca Gambarotto, Frank Hindriks, Jan Lambooy, Alessandro Lanteri, Charles van Marrewijk, Frank van Oort, Ana Santos, Andrew Sayer, Jack Vromen, Petri Ylikoski have read parts of the manuscript and offered helpful comments for which I am most grateful.

Many friends both cheered me up when I was sad and had fun with me when I was happy. The list is so long that cannot be made to fit in here. Special thanks go to the *Pink Ladies* from my beloved home town, and to Ana, Anna, Bauke, Carlotta, Elena, Federico, Frank, Gulbahar, Mar, Roberta, Susana, and the EIPE's students from all over the world.

A special thanks goes to Päivi Oinas and Uskali Mäki, my PhD supervisors. Not only have they offered me the best possible guidance in my work, but they have also made me feel comfortable enough to look for their supporting presence when things were not progressing in the way I had hoped that they would. Their house certainly felt like another home. I also appreciate greatly the fact they have allowed me to include our joint papers, Chapter 2 with Uskali Mäki and Chapter 4 with Päivi Oinas, as parts of this dissertation.

I gratefully acknowledge the financial support of the Erasmus Institute for Philosophy and Economics and of the Academy of Finland. Vereniging Trustfonds contributed to my visits to conferences where I presented drafts of the chapters of this thesis.

Chapter 3 has been published in *Environment and Planning A*, 2004, 36 (10), 1737-1753. It is reprinted here with the kind permission of Pion Limited, London (<http://www.envplan.com>)

Finally, this thesis is dedicated to Daniela, Franco and Stella. More than a thousand kilometres distance has made many things difficult (and perhaps a few things easier) throughout these years. Despite this, it has always been, and always will be that my amazing family are the ones who have loved, supported and helped me in infinite ways. *Grazie*.



# **1 Unity, plurality and explanation: The questions, the premises, the contributions**

*They may seem to be different sorts of things. Some of the theories may address different phenomena or different realms of phenomena.*

*Some are genuinely competing, others can be reconciled with one another, while still others pass one another by, answering different questions. They fit together only in a very complicated and overlapping geometry (Garfinkel 1981: 1)*

## **1.1 Questions**

Spatial issues have been somewhat neglected in the history of the economics discipline (cf. Blaug 1996). Yet economic activity clearly does not take place in the proverbial head of a pin: space and distance do affect economic activity in a non-trivial way. Aimed at ending the long silence of the economics discipline on the “spatial economy”, a new approach to spatial issues was developed at the beginning of the 1990s. Almost by accident, Paul Krugman, at that time already well known for his contribution to new trade theory, noticed that a small modification to his new-trade-theory models would allow the endogenous derivation of spatial agglomeration: the New Economic Geography was set off. Krugman (1991a) employs a

general equilibrium model characterised by the presence of increasing returns, imperfect competition and transportation costs to derive a core-periphery pattern, that is, a situation where economic activity is fully concentrated in a core region. Later, partly in reaction to the complaints of “economic geographers proper” (cf. Martin 1999), the New Economic Geography became known as “Geographical Economics” (GE henceforth).

By now, GE is an established field in economics. The appearance of four monographic expositions confirms its present standing (Fujita et al. 1999; Fujita and Thisse 2000; Brakman et al. 2001; Baldwin et al. 2003). The same basic approach that was first developed in Krugman (1991a) has been refined, extended and suitably modified to be applied to a host of spatial phenomena such as the analysis of the emergence of cities and systems of cities, growth and development, international trade and the behaviour of multinationals.

There are many features of this intellectual episode that trigger the interest of a philosopher of economics. Although much of mainstream economics did neglect spatial issues, regional and urban economics are well-established fields of economics whose domains of inquiry overlap with that of GE. The emergence of GE also coincided with a renewed interest in spatial agglomeration and clustering<sup>1</sup> in neighbouring fields, such as economic geography, economic sociology and business studies.

Scholars from these different disciplines have criticised GE on a number of grounds. One of the main points of contention concerns the GE models and their unrealisticness. Debates over the unrealisticness of theoretical models,

---

<sup>1</sup> Clusters are geographical concentrations of firms operating in the same industry or related industries potentially connected through production or service-related ties.



especially (though not exclusively) those grounded in the neo-classical tradition, abound in economics. The dispute between economic geographers and GE provides a nice case study into the old but yet unresolved *issue of unrealisticness* [1]. In particular, it provokes us to ask the following questions.

[1a] What roles do different unrealistic elements play in the GE models?

[1b] If these elements play different roles, do proponents and opponents of the GE models correctly distinguish between these roles?

[1c] And more generally, what is the explanatory worth of unrealistic models?

Second, GE pursues *explanatory unification* [2]. The same basic approach is applied to explain a variety of phenomena that were previously attended to by separate theories. Among the questions posed by the pursuit of explanatory unification by GE are:

[2a] Is GE a genuine instance of explanatory unification?

[2b] If it is, what features does it display?

[2c] Do existing philosophical accounts mainly wired to the natural sciences shed light on this (alleged) social science instance of explanatory unification?

Third, at present we witness a plurality of approaches to spatial agglomeration, a plurality that apparently stands in sharp contrast with the unificationist ambitions of GE. Undoubtedly,

this is fertile terrain for an analysis of *inter-disciplinary and inter-theoretic relations* [3]. It requires us to address the following issues.

[3a] Are the approaches on offer incompatible and how should we go about finding this out?

[3b] Are they rivals or complementary and how should we go about finding this out?

Fourth, the presence of a plurality of alternative theories seeking to account for the same phenomenon leads to issues of *theoretical pluralism* [4]:

[4a] Should a unified approach be pursued and if obtained, should it replace the existing plurality?

[4b] What grounds are there for either accepting or rejecting the existing plurality?

From a philosophy of economics perspective, this host of questions offers enough stimulus and substance to motivate a philosophical and methodological study of the case of GE and its neighbours.

This thesis is a collection of essays that address the above issues using sophisticated tools developed in the philosophy of science. Two essays, Chapters 3 and 5, are primarily meant as contributions to the particular field constituting my object of study. They clarify features of GE and relate them to theories developed within neighbouring fields. In these two articles, the philosophical content has been kept to a minimum as their intended audience is that of practicing theorists, mainly (though not exclusively) interested in spatial agglomeration

issues. Chapters 2 and 4 intend to contribute in a more direct way to discussions carried out within the philosophy of economics and of the social sciences in general. In these, GE and its neighbours are analysed as a case study meant either to serve as a check on the fit between the chosen philosophical accounts and actual scientific practice or as an illustration for proposed refinements to available approaches.

These four essays share a common set of philosophical premises in terms of which some kind of overarching philosophical framework can be outlined. This introductory essay serves to clarify those premises and their relations. Thus, after a brief introduction to the case of GE and its neighbours, I outline the philosophical premises that informed this thesis. In particular, in Section 1.2, I discuss realism and the partial and approximate nature of theories. I take on the view that though there is only one world, phenomena are complex and multifaceted, and hence, our theories are only capable to capture aspects of those phenomena. One of the main epistemic tasks of theories is that of providing explanation of phenomena. In Section 1.3, I consider the view of explanation that permeates this thesis. I make extensive use of the contrastive approach to explanation, which makes it possible to precisely formulate what aspect of the phenomenon an explanation, and the theory that provides it, can actually explain. Finally, in Section 1.4, I discuss kinds of inter-theoretic relations and provide ontological, epistemological and pragmatic grounds for accepting a plurality of interrelated theories and explanations of the same phenomenon. What emerges is a realist-grounded pluralism. After having outlined the philosophical framework, I summarise the contributions of this thesis in Section 1.5.

## 1.2 Geographical economics and its neighbours

GE aims to explain the uneven distribution of economic activity in space, in particular, the phenomenon of spatial agglomeration at different spatial scales. The concept of agglomeration refers to phenomena such as the existence of industry clusters, the emergence of core-periphery patterns, the existence of cities and systems of cities, the pattern of international trade and specialization, the causes of economic growth and development.

Geographical economists hold that these apparently distinct phenomena might be (at least partly) brought about by the same kind of economic mechanisms. Motivated by this conviction, GE proposes a unified approach for the study of those phenomena.

The presence of increasing returns at the firm level and transportation/trade costs<sup>2</sup> are the building blocks of the GE framework. At the aggregate level, increasing returns and transportation costs give rise to pecuniary externalities. Pecuniary externalities<sup>3</sup> are transmitted through the market via price effects. Roughly, the presence of pecuniary externalities implies that the more firms and workers there are in a locality, the more the locality becomes attractive as a location for further firms and workers. This triggers a cumulative process whose end result might be that all economic activity turns out to be concentrated in one locality.

---

<sup>2</sup> Increasing returns to scale are defined as a decrease in the average costs per unit of output for the individual firm as the level of output increases. Transportation/trade costs are generally assumed to be of the “iceberg form” meaning that part of the good “melts” in transit, and hence, only a fraction of it arrives at destination.

<sup>3</sup> Externalities are defined as a decrease in average costs as a result of an increase at the level of output of the whole industry.

Pecuniary externalities are centripetal or agglomerating forces, that is, they push towards the concentration of economic activity. In the GE framework, these centripetal forces are counteracted by centrifugal forces, (such as the presence of immobile factors, of congestion and the like) which push towards dispersion. The GE models are characterized by the presence of multiple equilibria: whether or not, and where, agglomeration occurs depends on the relative strength of those forces and on initial conditions, that is, on previous locational decisions.

As I have already said, GE is not the only approach seeking to account for agglomeration phenomena. Fields such as urban and regional economics, economic geography and strategic management are also busy with investigating agglomeration and its effects on economic activity. Each of these disciplines and approaches has reacted differently to the emergence of GE. It seems fair to say that the fiercest critics of the GE approach have thus far come from within the field of economic geography.

At the basis of the rather hostile reaction of many economic geographers is a historically-rooted resistance against 'neoclassical economics' and its 'methodology'. Around the 1970s, many economic geographers abandoned the emphasis on mathematical formalism and the search for general laws and regularities that had characterised the field since the 1950s and began to look for alternative ways of theorising. One such influential alternative is provided by the version of Bhaskarian transcendental realism (Bhaskar, 1975) proposed by Andrew Sayer (1984) (cf. Mäki and Oinas 2004; Sayer 2004). At present, economic geography is a much-diversified field in terms of both philosophical and methodological commitments. As economic geographer Allen Scott (2004) put it, "[a] cursory count reveals" that, in recent years, the discipline has been variously

undergoing “an empirical turn (Smith 1987), an interpretative turn (Imrie et al. 1996), a normative turn (Sayer and Storper 1997), a cultural turn (Crang 1997), a policy turn (Martin 2001), and a relational turn (Boggs and Rantisi 2003), among others” (Ibid: 483).

The relation between GE and economic geography constitutes the subject of two of the subsequent chapters: Chapter 3 examines in some detail the ongoing dispute, and Chapter 5 discusses some of the economic geographers’ criticisms against GE as an illustrative case, the analysis of which may contribute to the literature on unrealistic models.

In this thesis I do not discuss all fields and approaches whose domain of inquiry in one way or another overlaps with that of GE<sup>4</sup>. In Chapter 4, GE is compared to three theoretical approaches to clustering proposed within different fields, namely economic geography, strategic management and regional planning.

### **1.3 Realism and the partial and approximate nature of theories**

The first philosophical premise that informs this thesis is a commitment to the view that there exists an objective world (metaphysical realism) and that theories can correctly capture its working (scientific realism) (cf. Psillos 1999).

The theories and approaches analysed in this thesis lend themselves to a realist interpretation, and their proponents seem to interpret them along these lines. This holds for GE as

---

<sup>4</sup> There is a host of other interesting approaches to clusters and agglomerations that I have not mentioned in this work (for example, Arthur 1994; Beccattini 1990; Bellandi 1996; Camagni 1995; Henderson 1977; Jacobs 1961; Piore and Sabel 1984; Putnam 1993).

well, contrary to the economic geographers' ascriptions of 'instrumentalism' and 'positivism' to it. In Chapter 3, I argue that these labels are misplaced: geographical economists appear to hold a realist attitude towards their theories<sup>5</sup>. The source of confusion very much rests on a conflation, noticed some years ago by Mäki (1992 a, b) between two different notions: those of realism and realisticness. Keeping these two notions distinct helps to see how the GE models may contain a host of unrealistic elements and yet be interpreted realistically. (A more extensive exposition of Mäki's view is given in Chapter 3).

Though theories can correctly capture the working of an objective world, they can only do so in a partial and approximate way. Since (i) most real-world phenomena are infinitely complex, (ii) our cognitive capacities are limited, (iii) theories are representations, and (iv) hence cannot be faithful to the represented object on every aspect with perfect precision, (v) our theories are by their very nature both partial and approximate.

In particular, it is in the nature of theories to be partial (in Mäki's (1992a) terminology this constitutes one sense of unrealisticness) for they do leave out many aspects of the phenomenon they investigate. Theories are also approximate, that is, they are similar to the objects they represent only to certain degrees (See Giere 1988; Teller 2004; Mäki 2001a)<sup>6</sup>. Very often, focusing on one or a few aspects not only requires neglecting others, but also introducing flatly false elements. In Mäki's terminology, we say that in order to isolate theoretically

---

<sup>5</sup> Recognising this, however, does not imply anything about the veracity of the GE theory, nor of any other for that matter. The point is simply that theories interpreted realistically can turn out to be either true or false and, whether they are true or false depends on objective states of affairs.

<sup>6</sup> Although many philosophers agree that it is in the nature of theories to be partial and idealized, the specific terminology employed and the implications they draw vary.

some aspects of the object under study, theories often employ idealizations, that is, plainly false assumptions made for the purposes of neutralizing, theoretically, the effects of “disturbing factors”. (Following Mäki (1992a), this constitutes a second sense in which theories and models are said to be unrealistic). It is generally agreed that this does not constitute a grave problem as far as those idealizations concern features which are irrelevant or secondary to the aspects which are meant to be truly described.

All this means that theories differ with respect to what aspects they include and to the degree to which they are “true” of those included aspects. Comparisons between theories, models and the explanations they provide can be carried out by looking at what, and how, they isolate as being in need of explanation and as doing the explaining. Questions by proponents of rival approaches about what is isolated as in need of explanation and as doing the explaining gives rise to what Mäki (2004b) terms the *dynamics of dispute*. Here “dispute” refers to the process of questioning the boundaries of those isolations by proponents of rival approaches, and “dynamic” refers to the role of these appraisals in driving theoretical change.

The dynamics of dispute creates the potential for *erotetic progress*. Erotetic progress amounts to asking better questions (potentially) leading to better explanations (Mäki, 2004b; cf. Kitcher, 1993). Erotetic progress is, in this way, parasitic on *explanatory progress*. Mäki (2004b) distinguishes between two notions of explanatory progress stemming from two conceptions of explanation. One sees explanation as unification: to explain is to unify a number of apparently diverse phenomena (cf. Friedman 1974; Kitcher 1981, 1989). The other sees explanation as a matter of representing the causal nexus within which the explanandum is embedded: to explain a



phenomenon is to cite the nexus of causes that brings it about (cf. Salmon 1984, 1998). Translating these into notions of explanatory progress, we obtain the two notions of progress as *scope expansion* and progress as improved *causal articulation or penetration*. The former occurs “when the domain of kinds of phenomena explained by a theory expands” (Mäki 2004b: 326). I call this *explanatory scope*: progress as scope expansion occurs when a widening of explanatory scope takes place. The second “is a matter of an improving grasp of the relevant details of the causal processes and mechanisms behind the phenomena to be explained” (Mäki 2004b: 326). I call this *explanatory depth*: progress as improved causal articulation occurs as explanatory depth increases.

#### **1.4 Causal explanation and contrastive questions**

Rather than theory or model I take explanation as the unit of analysis. Obviously explanation is only one among the goals a theory and its models are used for, but it is certainly one, if not the most, significant epistemic task. The strategy I employ is to reconstruct a given theoretical approach as giving an explanation of the phenomenon under investigation. In many cases, this means stripping the approach to its bare essentials, but when needed, other features of the theoretical approach are carefully considered.

The philosophical literature on explanation is vast (cf. Salmon 1989; Psillos 2002). It ranges from theories that investigate the nature of explanation to analyses of specific kinds and modes of explanation. A review of this exciting field is far beyond the scope of this work. Here I will discuss the most influential ideas about explanation that inform this work and how these fit with the rest of it.

An explanation is composed of three elements: the explanans (what does the explaining), the explanandum (what is to be explained) and the relation between them. Theories of explanation not only differ as to what the proper relation between explanans and explanandum is in order for an explanation to count as such, but also as to what the relations themselves are.

For purposes of generality, I take explanans and explanandum to refer to all sorts of things, such as laws, processes, mechanisms, phenomena, events, patterns, states of affairs, and so on. I am concerned with both singular explanations, that is, explanations of particular occurrences, and theoretical explanations, that is, explanations of laws, regularities, patterns and the like. In regard to the relation between explanans and explanandum, I take them as causally related. Roughly, I take it that explanation is a matter of citing causes, or better, of locating the explanandum in a causal nexus<sup>7</sup>. This view is thin enough to fit with a variety of substantive theories of causal explanation.

In addition to subscribing to the idea that explaining is (at least in the majority of cases) to cite causes<sup>8</sup>, I make extensive use of contrastive analysis (see Chapters 4 and 5). The contrastive approach to explanation (see for instance Barnes 1994; Garfinkel 1981; Lipton 1990; 1991; Woodward 2003;

---

<sup>7</sup> In the previous section, I have discussed two views of explanation: the causal and the unification view. The unification view takes it that to explain is to unify. In contrast, I here adopt the view that to explain a phenomenon is to cite its cause, and hence, that unification is parasitic on causation. Explanatory scope is a matter of explaining different classes of phenomena as brought about by same or similar causes.

<sup>8</sup> In Chapter 4 we encounter a case of functional explanation. Functional explanations are not at odds with the idea that explanation is a matter of citing the causes of a phenomenon. See for instance Kincaid (1996) for an exposition of the view that functional explanations are causal explanations.

Ylikoski 2001) takes an explanation to be, either explicitly or implicitly, an answer to a composite question of the following form: why  $f$  rather than  $c$ ? where  $f$  is an aspect of a phenomenon to be explained (for instance, leaves turning yellow in autumn) and  $c$  is an alternative to it (for instance, leaves turning blue in autumn or turning yellow in spring). Formulating an explanation-seeking question in contrastive terms allows us to make sense of a given selection of factors among the many that lie behind the occurrence of a phenomenon (see Chapter 4 for a more extensive discussion of the contrastive approach).

Although I do think that the contrastive approach to explanation points to a constitutive element of what explaining amounts to, the main insights of this thesis can be retained even if contrastivity is seen as merely a useful instrument to precisely formulate what a given explanation *can actually* explain.

The contrastive approach proves to be of help when assessing the explanatory potential of theoretical models and when investigating inter-theoretic relations (in terms of the set of questions posed at the beginning of this essay, the areas in which the contrastive approach is of help are [1] and [3], respectively).

First, as I argue in Chapter 5, contrastivity proves to be an aid in the solution of debates over unrealistic models in economics, namely, issue [1]. The point is that the degree to which a model is unrealistic is to be judged by paying attention to the contrast the phenomenon to be explained is compared to. If we take explanatory questions to be contrastive, then theoretical isolations and idealisations are, at least partly, dependent on the contrasts. Or to put it the other way around, isolations and idealisations determine the range of contrastive questions about a given phenomenon a theoretical model can actually explain (in Chapter 5, I call this “explanatory variety”). The contrastive approach serves well also in a dynamic context:

it helps to assess whether and how the explanatory potential of a given theoretical approach changes over time, and hence, it might assist when making judgements regarding explanatory progress.

Second, regarding inter-theoretic and inter-disciplinary relations, namely issue [3], as others have already observed (Garfinkel 1981; Ylikoski 2001; Day 2002), what might at first appear to be rival explanations, on closer examination display an intricate web of relations of complementarity and rivalry. Not only are theorists often not very specific about what they are explaining, they might even offer misleading hints regarding the actual explanandum. The *intended* explanandum, what a theorist intends to explain, and what we might call *actual* explanandum, that is, what the theory is actually able to explain, might diverge. The contrastive approach to explanation offers the resources to compare theories and their explanations in terms of their actual explananda, and it allows us to do so at a very fine grain level. Thereby, it permits to appreciate differences and similarities that a less fine-grained analysis would not. Chapter 4 confirms the usefulness of the contrastive approach in the analysis of inter-theoretic relations by applying it to (a selection of) explanations of spatial clusters.

## 1.5 Pluralism and inter-theoretic relations

Theoretical *pluralism* refers to a stance that favours the co-existence of a plurality of theories of the same phenomenon (cf. Mäki, 1997). Once we acknowledge that theories are both partial and approximate, ground is prepared for the co-existence of a plurality of theories about the same phenomenon, each interested in one or another of its various aspects. Addressing questions regarding the conditions under which a

plurality of theories is acceptable and even welcome, namely issue [4], is important in this respect. There is a variety of forms of pluralism and a variety of grounds on which pluralism can be advocated. Before assessing whether a given plurality of theories should be accepted or rejected and on what grounds, however, we are first required to deal with the question as to how those theories precisely relate to each other. Borrowing Garfinkel's words quoted at the beginning of the chapter, we are required to ascertain the overlapping geometry of their relations.

### 1.5.1 On inter-theoretic relations

Focusing on what has really been explained, that is, the actual explanandum, is crucial to a sophisticated analysis of inter-theoretic relations (issue [3]) and the contrastive approach allows us to do precisely such a fine-grained analysis. After having individuated the actual explananda of the theories under investigation, we are required to ascertain whether the explanations in question are compatible or not.

Given that I take explanations to (possibly) capture objective causal relations, then two sources of incompatibility can arise: a) incompatibility of the explananda b) incompatibility of the explanantia<sup>9</sup>. Incompatibility can arise because of factual disagreements, conceptual incompatibility, or because an explanation disputes the relevance of the explanans of the rival explanation (Day 2004).

When a plurality of explanations has the same explananda, we are confronted with the following possibilities. First, they

---

<sup>9</sup> I do not consider compatibility and incompatibility among explanatory relations as a distinct case. I take explanations to be possibly capturing causal relations and the explanantia to comprise the way in which a cause brings about the effect.

can be incompatible because of some form of incompatibility among the explanantia. Thus, at most one of them is correct. Second, the explanations are compatible. In this case, the explanations are either incomplete or dependent on each other and the factors they refer to can be causally or otherwise related to each other (Kim 1988, 1989; Day 2004). (Table 4.1 outlines some of these relations and Chapter 4 applies them to the plurality of explanations of spatial clustering)

In Chapters 3 and 4, compatible theories are said to be complementary and incompatible theories are said to be rivals. However, this represents a terminological simplification. Indeed, theories providing compatible explanations are not thereby complementary, and they might even be rivals. Let me explain. First, there is a distinction to be made between incompatible explanations and rival explanations. Incompatible explanations are those that cannot be true at the same time because of the way the world is. Rivalry might refer to incompatible explanations in the way defined above, but it needs not do so. Theories can be rival for instance when they compete in the pursuit of financial resources, dominance, prestige, that is, one or another scarce resource (it could be thought that incompatible theories compete in the pursuit of truth). The implication of making this distinction is that compatible explanations may still be rivals on these other dimensions. Some of the explanations and theories discussed in this thesis appear to be compatible from an ontological point of view, and yet they are perceived to be rivals on other dimensions.

Second, compatible theories and explanations are not *ipso facto* complementary. This might often be the case, but it needs not be so. Complementarity involves more than compatibility. Think for example of the case where the two explanations cite, employing different terminologies, identical causal factors. In

this case, the two explanations, though compatible, can hardly be said to complement each other. What then makes two compatible explanations complementary?

First, the two explanations should be about the same or closely-related aspects of a phenomenon: there is not much sense in attributing complementarity to two compatible explanations, one that explains why my pen just fell on the floor and the other why dinosaurs got extinct. Second, for explanations to be genuinely complementary, they have to provide *distinct (kinds of) information*: If one of the explanations contains all the explanatory information of the other, then the latter is redundant and should not be said to be complementary. Chapters 3 and 4 analyse alternative theories of spatial agglomeration and clustering and their various aspects. They suggest that some of these theories do provide different kinds of information, and hence are complementary, while others are compatible, but do not appear to complement each other in any interesting sense.

Compatible explanations could, at least in principle, be unified. Since unification is generally regarded as a scientific ideal (cf. Mäki 1990, 2001a), it might even be argued that the integration of the two explanations into one should be pursued. In this thesis I am dealing with explanations that are typically embedded in scientific theories; hence, the integration of explanations is usually achieved through the unification of theories. Unification of scientific theories generally takes place through the development of a more fundamental theory that shows that the objects of the unified theories are manifestations of the objects of the unifying theory (Mäki 1997; 2001b). There are a number of epistemological and pragmatic reasons that make theoretical unification in many cases difficult or even impossible to achieve.

For the sake of the argument, however, suppose that those difficulties are overcome and unification succeeds. Do all unifications entail the replacement of the original theories? The answer is negative because theoretical unification usually requires modifications of the original theories (Mäki 1997: 45). The unifying theory will also be partial and approximate, but it will be so in a way different from the original theories. Explanatory unification is typically achieved by way of developing a more abstract law or principle (Morrison 2000). This indicates that in most cases the unifying theory, by being more abstract, and partial and approximate in a different way, will not provide the same kind of explanatory information offered by the original theories.

Chapter 2 discusses unification in GE, namely issue [2], and shows that unification has been achieved by way of developing a more abstract and idealised framework. We do not extensively discuss the unified theories<sup>10</sup>, but GE theorists themselves regard their theory to be a complement, rather than a substitute, to the original theories. The main reason for this seems to be that the kind of mechanisms that accomplishes the unification is taken to bring about the phenomena the theory unifies, but it is not believed to be the only responsible mechanism. Other, different kinds of mechanisms, which constitute the objects of the unified theories, do operate as well in the production of those phenomena. Hence, the GE case appears to support the point that in most cases, the development of a unifying theory does not entail the replacement of the unified ones.

The larger implication of this is that either because of the impossibility of theoretical unification or because the theory

---

<sup>10</sup> These are theories in urban and regional economics, trade theory, growth and development theory (cf. Mäki and Marchionni 2005b for a discussion of this aspect of the GE unification).



that unifies does not fully replace the original theories, a certain degree of plurality generally remains. But are there grounds for accepting and even welcoming such a plurality?

### 1.5.2 What grounds for pluralism?

After having identified the philosophical premises that inform this thesis, I take a small step forward and show that those premises lead naturally to the endorsement of some form of pluralism. Thus, I address issue [4]. This has implications for GE and its neighbours where a plurality of theories and approaches prevails. What justifications might we have for accepting and perhaps welcoming this state of affairs?

The realist thesis that there is only one world, and hence only one way in which things in the world really are, is consistent with the idea that phenomena are multifaceted, and that different theories and explanations can capture only one or a few aspects of them. However, from the ontological viewpoint, a multiplicity of theories and explanations is acceptable only insofar as the theories and explanations in question are compatible.

Yet there are epistemological and pragmatic grounds on which even a multiplicity of incompatible theories can be accepted. These same grounds also provide further reasons for welcoming a plurality of compatible theories. The well-known problems of underdetermination and theory-ladenness which point to the difficulty (or even impossibility) of unambiguously proving which, if any, of the theories in question is the correct one provide strong reasons for accepting and in most cases welcoming a plurality of theories (whether they be compatible or not). Social epistemologists (for example, Kitcher 1994, 2001; Longino 1990, 1991; Solomon 1994, 2001) have discussed these problems at length. In general, they believe that an

environment favouring a variety of theories, methods and points of view, some of which might be incompatible, is more conducive to scientific progress, thereby endorsing some form of *epistemological pluralism* (cf. Feyerabend 1963). Erotetic progress, discussed in section 1.3 above, hinges on the dynamics of dispute, and hence is itself facilitated by the presence of a plurality of different theoretical approaches and viewpoints.

A plurality of compatible theories can also be retained for pragmatic reasons: some theories may serve certain purposes better than others. This means that theories and explanations of the same phenomenon may focus on different explanantia because of different pragmatic interests<sup>11</sup>, and that we could wish to retain them all because each meets a particular interest.<sup>12</sup>

The overall picture that emerges from this discussion is a realist pluralism of sorts, that is, a pluralistic view that nonetheless accords with realist tenets. There are strong reasons for accepting and even welcoming a plurality of *compatible* theories. These are ontological (the world is one, but the phenomena that make it up are complex and multifaceted), epistemological (theories are partial and approximate, and epistemic uncertainty prevails), and pragmatic (different theories meet different pragmatic interests). There are even epistemological and pragmatic grounds for welcoming a plurality of *incompatible* theories. This implies that though

---

<sup>11</sup> Pragmatic interests partly determine the selection of the contrast. Once the contrast is selected, pragmatic interests might also affect the isolation of certain factors among the ones that explain the contrastive explanandum.

<sup>12</sup> Notice, however, that accepting this does not imply that what counts as explanatory wholly depends on epistemic interests Cf. for instance van Bouwel and Weber (2002) and Weber and van Bouwel (2002) for a pluralist stance mostly based on pragmatic considerations.

disputes among proponents of alternative theories of the same phenomenon are likely to lead to theoretical progress, progress ought not to be interpreted as a tendency towards reducing theoretical variety and plurality. The development of the true and complete theory of a phenomenon is unlikely, if not impossible. On the other hand, acknowledging this does not amount to denying the virtues of theoretical and explanatory unification, especially if these are ontologically grounded (Mäki 2001a). The view I advocate here suggests that the way to theoretical progress is likely to be characterised by the coexistence of tendencies towards both unity and plurality.

The implication of this for GE and its neighbours should be clear. Spatial agglomeration research is characterised by a plurality of theories and approaches committed to different epistemological and methodological principles. Some of these theories might actually be compatible (see Chapters 3 and 4) and focus on different aspects; others, by contrast, might actually be incompatible. Yet, for the reasons just discussed, this variety of approaches should be accepted and even welcomed. Attempts to unify such as the GE's one should not necessarily be looked at with suspicion, at least as far as a certain degree of plurality and variety is preserved.

## **1.6 Summary of the contributions**

Having outlined the philosophical framework, it is now time to look in more detail at how it has been variously employed in the four essays to come, and hence, to summarise the contributions of this thesis. Notice that the order of the subsequent chapters does not entirely comply with the way in which I presented the issues in this introductory essay: the essay concerning unrealisticness, that is, issue [1], is the last

chapter in the thesis (Chapter 5). In addition, pluralism (issue [4]) is not explicitly discussed in any of the subsequent chapters.

Chapter 2 shows that GE represents an instance of explanatory unification [2a] and identifies its characteristic features [2b]. Unification in GE has been a matter of applying the same pattern of derivation to different kinds of explanandum phenomena. At the same time, GE aimed to show that these phenomena are really produced by the same kind of mechanisms. Although unification in GE has been achieved thanks to the application of an abstract mathematical structure, this does not imply that GE lacks explanatory power. Since it aims to explain stylized facts, its abstract mechanisms are possibly explanatory. Explaining particular occurrences, however, requires including not only details about the specific situation, but also mechanisms that are not included in the GE framework. We also ask whether the philosophical accounts we considered are able to shed light on this social science instance of unification [2c]. The analysis shows that although none of them is capable of fully capturing the distinctive features of the GE's unification, they contribute to illuminate its different aspects. We take this to suggest that unification in science is not uniform, but rather adopts different forms and strategies depending on disciplinary (and other) contingencies.

Chapter 3 discusses the dispute between economic geographers and geographical economists. It seeks to clarify this dispute by way of removing possible misunderstandings. In particular, this essay addresses the issue of inter-disciplinary and inter-theoretic relations [3]. It investigates whether the approaches of economists and geographers are incompatible [3a] and, if they are not, whether, and how, they complement each other [3b]. Once misunderstandings are removed, I show

that the two approaches might complement each other on both methodological and substantial grounds.

Chapter 4 analyses a selection of explanations of spatial clustering. We consider GE and theories proposed in economic geography, management studies and regional planning. This chapter, therefore, contributes to the issue of inter-disciplinary and inter-theoretic relations [3]. It answers questions concerning the way in which the theories in question relate to each other, whether they are compatible, complementary or rival, that is, questions [3a] and [3b] above. The formulation of the explanandum in contrastive terms helps to see as compatible explanations that would otherwise seem incompatible. We obtain an intricate web of relations: some explanations are partially incompatible, others are merely compatible, and still others appear to complement each other.

Chapter 5 also adopts the contrastive approach to explanation. It goes some way towards addressing the issue of the explanatory worth of theoretical models [1c]. In particular, I demonstrate that contrastivity pinpoints a distinctive role for some of the unrealistic elements in theoretical models, that is, the role of fixing the causal background shared by the fact and its foil. By identifying some of the roles unrealistic elements play in the GE models, [1a], I show that some of the unrealistic elements of the GE models can be interpreted as fixing the causal background. I investigate whether opponents of GE have correctly distinguished those roles [1b] and suggest that, once its explanandum has been precisely delimited, some of the criticisms of unrealisticness raised by economic geographers appear to be misguided.

\*\*\*

The contributions of this thesis can be divided into two categories: those which concern the specific social science case, that is, GE and its neighbours, and those whose relevance

concerns the philosophy of economics and of the social sciences more generally. In what follows I discuss the general conclusions, concerning both types of contributions, which can be drawn from the more detailed analyses.

### 1.6.1 For geographical economics and its neighbours

A number of general conclusions that primarily concern GE and its neighbours emerge from this thesis. I summarise them as follows.

(i) Explanations of spatial agglomeration and clustering have an implicit contrastive form. They compare agglomeration with alternative spatial arrangements such as dispersal. The contrast partly accounts for the type of agglomeration economies they isolate as explanantia. This holds for GE as well.

(ii) The GE models contain a host of unrealistic elements playing different roles. Some can be read as fixing the causal background between the fact (agglomeration) and the foil (dispersion). These unrealistic elements of the GE models restrict the GE explanandum which turns out to be: "Why is economic activity agglomerated, rather than dispersed, in cities, clusters, certain countries and groups of countries?"

(iii) Delimiting the GE explanandum in this way shows two things. First, the explanatory potential of the GE models, in terms of the range of questions it has the resources to explain, is very limited. Second, some misunderstandings affect the dispute with economic geographers, whose criticisms, in some cases, turn out to be misplaced. It appears that economic geographers and geographical economists are interested in different questions, and hence, in different aspects of agglomeration.

(iv) The above considerations open up possibilities for complementarity. The explanations of spatial clustering here examined are largely compatible, and in some cases, they complement each other.

(v) Recognising complementarity does not amount to a denial of the methodological differences among the approaches here analysed. However, the presence of different epistemologies and methodologies should be seen as a favourable ingredient for theoretical progress to take place.

### 1.6.2 For philosophy of economics

The general conclusions regarding the philosophy of economics and of social science more generally are the following.

(i) The contrastive approach proves useful in the discussion over the unrealisticness of theoretical models. Some of the unrealistic assumptions of theoretical models can be seen as serving the function of fixing the causal background between the fact and the foil. Those unrealistic elements do not endanger the truth of the explanation.

(ii) The contrastive approach shows that the difference between explanations given by theoretical models and explanations given by other means is not a matter of kind, but of degree. Theoretical models, typically built in order to isolate a given process or mechanism, have the resources to explain a fairly limited number of contrastive questions about the phenomenon they investigate.

(iii) None of the accounts of explanatory unification developed by philosophers of science here considered can alone fully capture the characteristic features of explanatory unification in GE. Together, they are able to illuminate its different aspects. This suggests that the philosophical accounts

are partial and/or that the structure of unification in science is not uniform, but rather takes on different forms and strategies depending on disciplinary (and other) contingencies.

(iv) It is a recurrent theme of this thesis that imprecision about what is being explained leads to misguided judgments on the explanatory value of theories and on how they relate to alternative theories. In practice, very often theories and the explanations they provide are expected to explain more than what they actually can, and critics often take this failure to imply that the theories in question cannot explain anything at all. In some cases, vagueness about what is being explained leads to considering alternative theories as rivals when in fact they are complementary.

(v) The contrastive approach to explanation helps to achieve clarity and precision on what is explained and by what. Because of this, its usefulness for inter-theoretic comparisons is confirmed.

(vi) Since theories are both partial and approximate, different theories of the same complex phenomenon typically account for different aspects of it. This provides grounds for accepting and even welcoming a plurality of theories. Further epistemological and pragmatic reasons reinforce those grounds. On the realist pluralism I advocate here, the way to theoretical progress is characterised by tendencies towards both unity and plurality.



## **2 On the structure of explanatory unification: The case of geographical economics\***

### **2.1 Introduction**

Amongst practising scientists, unification is widely perceived as a methodological virtue that should be pursued, and eventual achievements of unification are to be celebrated. Philosophers of science have long recognized this feature of scientific practice and have sought to develop accounts of it. They are accounts of what it is to “explain much by little” and of how to assess such explanatory unifications.

Most of the illustrations and case studies provided by philosophers come from the natural sciences, mainly physics and biology (for example, Kitcher 1989; Morrison 2000). Few philosophical studies have been recently concerned with unification in the social sciences (see, however, Kincaid 1997; Mäki 1990, 2001). This is a shortcoming, as unification is no less of a pressing issue in the social sciences. The urge to unify has shaped much of social scientific theory formation, and the dream of a unified social science has given rise to a long history of research and debate that is very much alive today.

---

\* This essay is co-authored by Uskali Mäki

In this paper, we offer a detailed examination of a contemporary episode of explanatory unification in the social sciences. Thus, we follow the tradition of looking at the actual practice of science to gain insights into the phenomenon of explanatory unification. Our case is a newly emerged field within economics, known as geographical economics (or new economic geography; GE henceforth). GE was recently developed to deal with the economic role of space, an issue largely neglected by earlier mainstream economic inquiry. Unification is a prominent theme that has motivated and shaped research within GE. This new field claims to have provided a unified approach to the study of phenomena of agglomeration at different spatial scales (local, regional, national and, international) by showing how these phenomena can be traced back to the same basic economic mechanisms.

We analyze our case to cast light not only on GE itself, but also on some of the philosophical accounts of explanatory unification. We ask what exactly is going on in GE when the practitioners say what they do is unification; and we use our empirical discoveries to assess key elements in some of the major contemporary philosophical accounts of unification. Or, perhaps more accurately, we use some of the issues in the philosophical accounts to guide our discoveries, and then use those discoveries to assess the philosophical accounts for their fit. In particular, we will discuss the role of premises and of argument patterns in unifying derivations; the role of ontological convictions and mathematical structures in shaping unifications; the question of how explanation and unification relate to one another; and the related issues of multiple mechanisms and degrees of unification.

The paper is organized as follows. Section 2.2 provides a first glimpse on unification in GE: the celebrated achievements of unification, the major phenomena and theories that are claimed

to be unified, and the building blocks of the unifying framework. Section 2.3 briefly lists some key issues in recent studies of explanatory unification in the philosophy of science which will guide the subsequent discussion. The main feast is served in Section 2.4. It discusses the issues identified in the previous section and gradually develops an account of the kind of explanatory unification that characterizes GE. Section 2.5 concludes the paper.

## **2.2 The explanatory framework of geographical economics**

GE is a theoretical approach to spatial issues developed within economics in the beginning of the 1990s. It seeks to explain the uneven distribution of economic activity in space, and more precisely, the phenomenon of spatial agglomeration taking place at different spatial scales. The concept of agglomeration refers to seemingly very distinct empirical phenomena such as the existence of industry clusters, the emergence of core-periphery patterns, the existence of cities and systems of cities, the pattern of international trade and specialization, the causes of economic growth and development. The question is whether these phenomena of agglomeration do have something in common.

The uneven spatial distribution of economic activity shows a fractal dimension, that is, it repeats itself at different levels of spatial aggregation. To geographical economists this suggests that these apparently distinct classes of phenomena might be (at least partly) brought about by the same *kind* of economic mechanisms, viz. "economic mechanisms yielding agglomeration by relying on the trade-off between various forms of increasing returns and different types of mobility costs" (Fujita and Thisse 1991: 1).

[...] GE is able to show that the same mechanisms are at work at different levels of spatial aggregation. The idea that at least to some extent the same underlying economic forces are relevant for explaining the spatial organization of cities, the interaction between regions within a nation, as well as the uneven distribution of GDP across countries is very important. In order to lay the foundations for a unified approach ... (Brakman et al. 2001: 323)

Instead of investigating each of those classes of phenomena separately, using a different theory for each, GE proposes a framework that unifies previously separate fields of inquiry – such as international, regional, and urban economics – by showing how what previously seemed to be disparate phenomena can be traced back to the same basic economic mechanisms.

The empirical phenomena [...] have been studied thoroughly from many different angles, based on different theoretical frameworks, for a long time. From what is primarily a location perspective there are urban economics, economic geography, regional economics, and regional science. The interaction between economic centers is addressed by international economics, development economics, and industrial organization. One way to proceed would be to investigate each empirical phenomenon separately, using the insights of those of the above fields, inside or outside economics, which are thought to be relevant for the issue at hand. (Brakman et al. 2001: 17-18)

Unification is regarded as a major virtue of GE, a virtue that it shares with theories in more mature fields such as physics.

Since the 1970s the top item on the agenda of theoretical physics has been to unify general relativity and quantum mechanics into a theory of quantum gravity. [...] A main purpose of this book [viz. Fujita, Krugman and Venables' "The Spatial Economy"] is to demonstrate that many of the stylized facts of urban and regional economics [...] can be derived from a set of common assumptions [...] In any case, there is a need to unify fields in economics just like in physics, and the book has contributed to this discourse (Urban 2001: 151).

It is thus evident that explanatory unification has served as a guiding ideal for the development of GE. This can be highlighted by considering a pressing objection to GE, namely that it lacks novelty: the kinds of mechanisms it postulates have been known since long by both geographers and economists (cf. Martin 1999; Berry 1999). Geographical economists respond by granting the point and arguing that there are at least two ways in which it contributes to advancing the literature on spatial agglomeration (Brakman et al. 2001: 323). First, it is argued, many of the previous theoretical attempts to explain agglomeration typically lacked economic micro-foundations, i.e. agglomeration was not derived from the location decisions of optimizing agents. GE instead claims to be capable of providing the missing micro-foundations. The second alleged contribution is that it shows that the same basic economic mechanisms might in fact operate so as to shape the distribution of economic activity at different spatial scales, thereby unifying the theories that had previously been used for studying those phenomena separately.

The framework of GE within which these mechanisms are modelled has two building blocks: the presence of increasing returns at the firm level and transportation/trade costs<sup>13</sup>. Increasing returns at the firm level requires dropping the traditional assumption of perfect competition and replacing it with that of imperfect competition which is modelled according to the Dixit-Stiglitz monopolistic competition model<sup>14</sup>. At the aggregate level, increasing returns at the firm level and transportation costs give rise to pecuniary externalities<sup>15</sup>. Pecuniary externalities are transmitted through the market via price effects and, simply put, their presence implies that the more firms and workers there are in a locality, the more the

---

<sup>13</sup> *Increasing returns to scale* are defined as a decrease in the average costs per unit of output for the individual firm as the level of output increases (cf. (2.3) in the Appendix). The presence of space implies that the cost of exchanging goods and services across locations increases as (both physical and cultural) distance increases due to physical transport, tariffs, cultural barriers and so on. *Transportation costs* are generally assumed to be of the “iceberg form” meaning that only a fraction of the good arrives at destination (cf. Appendix). This assumption is made in order to avoid modelling a separate transportation sector.

<sup>14</sup> The Dixit-Stiglitz model provides a way to model imperfect competition. The Dixit-Stiglitz is a general equilibrium model where there is a large number of firms producing differentiated products under increasing returns to scale. Firms take the price-setting behaviour of other firms as given and do not take into account the effects of changing their own price onto the price index. The products are symmetric, which means that consumers do not prefer one variety to another. However, they prefer variety for its own sake, which means that they always prefer to consume a unit of a new variety than an additional unit of a product they have already consumed (both features are captured by the CES utility function, see Appendix). The Appendix provides the basic features of the core model of GE where the supply and demand side of the economy are indeed modelled a la Dixit-Stiglitz.

<sup>15</sup> *Externalities* are defined as a decrease in average costs as a result of an increase at the level of output of the whole industry. *Pecuniary externalities* are externalities that are transmitted through the market via price effects for each firm, which, as a consequence, may decide to change its output decisions.

locality becomes attractive as a location for further firms and workers. This creates a cumulative process whose end result might be that all economic activity turns out to be concentrated in one locality. While pecuniary externalities are forces that push towards the concentration of economic activity (agglomerating or centripetal forces), the presence of immobile factors, of congestion and the like push towards dispersion (dispersing or centrifugal forces). Let us call this the pecuniary externalities mechanism of agglomeration.

The GE models are characterized by the presence of multiple equilibria: whether or not, and where, agglomeration arises depends on the relative strength of those forces and on initial conditions, that is, on previous locational decisions. The GE models represent the process as path-dependent: the cumulative nature of the process of agglomeration is such that a small advantage of one location due to locational chance events in the past can have snowball effects which turn that location into the centre of economic activity, even though this outcome might not be the optimal one.

To explain empirical phenomena as diverse as systems of cities, core-periphery patterns, international specialization and so on, locations are interpreted either as cities, regions or countries, and the pool of dispersing and agglomerating forces is accordingly modified.

### **2.3 Explanatory unification in philosophy of science**

Our strategy is to examine the claims to unification in GE from the point of view of a few core issues that have emerged in recent philosophical work on explanatory unification. This section lists those issues and outlines the key concepts for

dealing with them, while the next applies them to the case of GE.

Many explanatory unifications seem to involve inference, or derivation of large numbers of explanandum sentences from a small number of some other items. Our first issue has to do with the character of these other items. In his quest for what explaining and understanding amount to, Michael Friedman (1974) suggests that understanding of a phenomenon is achieved by reducing the total number of independent phenomena we have to accept as ultimate. Unification is achieved by minimizing the number of premises and maximizing the number of conclusions in explanatory arguments. Philip Kitcher (1981, 1989) revises Friedman's account, arguing that the unifying unit is rather a pattern or scheme of derivation: "Science advances our understanding of nature by showing us how to derive descriptions of many phenomena, using the same patterns of derivation again and again, and, in demonstrating this, it teaches us how to reduce the number of types of facts we have to accept as ultimate (or brute)" (Kitcher 1989: 432).

Both Kitcher's and Friedman's views imply that unification (and explanation) is a matter of inference and derivation. But does this capture the key feature of unification? This gives rise to our second issue. If it does, the view would be that unification is achieved merely by way of deriving large numbers of explanandum sentences from a small set of premises or patterns of derivation. If it does not, one may argue that explanatory unification essentially involves revealing a shared ontology underlying the diverse phenomena that are explained. This suggests a distinction between two variants of unification, derivational and ontological (Mäki 1990, 2001). While derivational unification is a matter of deriving large classes of explanandum sentences from a parsimonious set of



premises, theoretical structures or inferential patterns, ontological unification is a matter of redescribing a large number of apparently independent phenomena as forms or manifestations of a common system of entities or causes, thus revealing an underlying ontic unity between apparently diverse phenomena.

Another angle from which to consider the role of ontology is suggested by the observation that unifications often come about facilitated and shaped by mathematical structures. Drawing from a few cases of unification in physics and biology, Morrison (2000) notices that in most of them the unifying theory embodies a mathematical structure that plays a prominent role in the process of unification, sometimes at the expense of ontology. Does the prominent role of mathematics in enhancing unifications rule out ontological unification, or are they compatible? This is our third issue. We seek to identify the respective roles of mathematics and ontology in driving unification.

Our fourth issue concerns how unification relates to other epistemic notions such as explanation and confirmation. Kitcher's view is that explanation is unification: to explain is to unify (Kitcher 1981, 1989). Many others take issue with this and instead argue that unification and explanation are logically and conceptually separate – even though a good explanation may turn out to be capable of unification as well. The view is that the unifying power of a theory does not contribute to its explanatory power, but it may contribute to its confirmation (Whewell 1847; Mäki 2001). Since we have been unable to find evidence in the GE literature concerning a connection between unification and confirmation, we will not pursue this idea further. We will instead examine Morrison's (2000) view of how explanation and unification relate. She holds that not only does unification not imply explanation, but there is a trade-off

between them. In her view, unification does not typically translate into an increase in explanatory power because it typically occurs via a process of abstraction. More abstract and general laws may unify, but they have less explanatory power because they neglect details specific to the phenomenon to which they are applied.

Finally, our last issue deals with the twin ideas of degree of explanatory unification and of ontic unity by considering the questions of whether the unifying mechanism is always the only mechanism in operation, whether it is always the main mechanism, and whether it is in operation in all instances of the classes of phenomena it unifies.

## **2.4 The structure of unification in geographical economics**

In what follows we examine how the case of GE relates to the four issues discussed above. The analysis will help us to outline an account of what is going on in GE when the practitioners claim they seek and manage to unify. We show that none of the presently available philosophical accounts of explanatory unification fully captures the kind of unification that seems characteristic of GE. Rather than concluding from this that unification in this specific context is not a genuine instance of explanatory unification, we endorse the view that unification is a complex phenomenon taking on different features in different contexts.

### **2.4.1 Premises and conclusions versus argument patterns**

On Friedman's early account (Friedman 1974), unification is conceived in terms of the number of premises and conclusions in explanatory arguments. Explanations are arguments whose

premises are the explanantia and whose conclusions are the explananda. Unification is a matter of minimizing the number of premises and maximizing the number of conclusions (Kitcher 1989: 431). Kitcher extends and refines Friedman's account. He suggests that explanatory unification is a matter of applying the same patterns of derivation over and over again to derive descriptions of different kinds of explanandum phenomena.

To Kitcher, argument patterns are composed of the following elements. First, schematic sentences are expressions in which most if not all non-logical expressions are replaced with dummy letters. Second, a set of filling instructions for a schematic sentence indicates how dummy letters are to be replaced in specific applications. A schematic argument is a sequence of schematic sentences. Third, a classification for a schematic argument provides the inferential characteristics of the argument. The triple composed of a schematic argument, a set of sets of filling instructions, and a classification for a schematic argument constitutes the argument pattern. Unification and hence progress is achieved by deriving descriptions of many types of phenomena using the same or similar patterns of derivation over and over again<sup>16</sup>.

In line with the trend in economics more generally (cf. Mäki 2001: 495), Kitcher's account appears to be more suitable than Friedman's for characterizing unification in GE. Here is an expression of the idea in terms of models that implement a theoretical pattern:

---

<sup>16</sup> Similarity between argument patterns is defined in terms of stringency of a pattern. This in turn is determined by the conditions on the logical structure imposed by the classification and by the conditions on the substitution of expressions for dummy letters.

Across a variety of models that seem quite different on the surface, a suitable redefinition of variables leads to the same expressions for break point and sustain point. [...] In this sense we can claim to have developed a theory of spatial concentration that is broader than any particular model, and that helps us to see a number of different models as particular cases of a more general approach (Fujita et al. 1999: 10)

The “more general approach” can be interpreted as the argument pattern which is used over and over again to derive agglomerations of different kinds with “a number of different models as particular cases”. The “suitable redefinition of variables” yields those particular cases; those variables can then be seen as the dummy letters in the schematic sentences. An instance of how variables are to be re-interpreted according to the specific application is given by the following quote.

In a world of countries—which we think of as geographical units that can trade, but among which labor does not move—agglomeration in the sense of population concentration cannot occur. However, linkages among industrial sectors can still lead to a process of industrial concentration that is conceptually very similar to the stories about agglomeration we told in our regional and urban analyses. [...] So far we have shown that a geographical approach to world trade can indeed be conducted using the same tools we have applied to cities and regions, and that the structure of this analysis is indeed at an abstract level almost exactly the same as what we have done before” (Fujita et al. 1999: 259)

The same abstract structure of analysis is retained across applications. In each, the basic expressions are to be redefined so as to refer to cities, regions or countries. This requires redescribing a number of variables so as to fit them with the application to those distinct levels of aggregation.

To exemplify, consider the core model of GE, namely Krugman (1991a)<sup>17</sup> (The basics of the core model and of an extension are provided in the Appendix). This derives the conditions for the existence and sustainability of a core-periphery pattern, that is, a situation in which the economic system is characterized by the presence of a core where the bulk of economic activity is concentrated. In this model, workers are mobile between locations and are assumed to move in response to differences in real wage<sup>18</sup>. This movement is at the basis of the forces that can bring about and sustain a core-periphery pattern. Consider now the establishment of patterns of international trade as the phenomenon to be derived. In this case, the assumption of labour mobility specific to the core model is dropped (as the extent of international mobility is much lower with respect to mobility across regions or cities), and hence locations are to be interpreted as countries. Labour mobility is replaced by the presence of inter-industry linkages

---

<sup>17</sup> Krugman (1991a) is considered the first core model of GE. It is the *core* model in the sense that it provides the basic apparatus for subsequent models which are extensions and refinements to it. It is also generally regarded as the *first* core model because recently a second kind of core model has been introduced (Fujita, Krugman and Venables 1999; Krugman and Venables 1995, 1996; Puga 1999). By replacing inter-regional labour mobility with inter-industry linkages (grounded on the presence of intermediate inputs) the second core model and its refinements solve a few problems which affect the first core model such as its bias towards deriving complete agglomeration and its failure to account for the empirical fact that concentration of manufacturing production exceeds concentration of manufacturing labour (Brackman et al. 2001; 2004)

<sup>18</sup> Cf. equation (2.6) in the Appendix

among firms that employ each other's products as intermediate inputs in production, as a force pushing towards agglomeration<sup>19</sup>.

The argument patterns that yield either agglomeration or dispersion are composed of expressions referring to characteristics of the supply and the demand side of the economy (according to the Dixit-Stiglitz model of monopolistic competition). Once the equilibrium solutions are obtained, expressions that represent the role of backward and forward linkages can be derived. These expressions provide the conditions (range of transportation costs) under which "an economy with agglomeration becomes possible" (the sustain point) and those under which "an economy without agglomeration becomes unstable" (the break point) (Fujita, Krugman and Venables 1999: 10). These expressions turn out to be very similar across a variety of different models.<sup>20</sup>

#### 2.4.2 Derivational versus ontological unification

On the interpretation just given, GE achieves unification by using similar patterns of derivation to derive descriptions of

---

<sup>19</sup> In the Appendix, the basics of the core model and the extension to international trade are outlined. Notice the similarity between (2.17)-(2.20) of the international specialization models with (2.7)-(2.10). Their differences are essentially due to the presence of the intermediates for production in the international specialization model. Also the centrifugal force may change depending on the phenomenon to be derived. For instance, the immobile factor can either be land or the agricultural workers. In the core model, the latter are immobile, whereas when GE seeks to explain the existence of cities and systems of cities, the immobile factor is typically land.

<sup>20</sup> Appendix reports expressions (2.15) and (2.21) for the sustain points and (2.16) and (2.22) for the break points of the core model and the international specialization model respectively. Notice that they are basically identical expressions except for a change in parameter.

apparently diverse phenomena. The next question we consider is whether GE yields nothing but derivational unification, in contrast to ontological unification. Derivational unification is a matter of deriving large classes of explanandum sentences from a parsimonious set of premises, theoretical structures or inferential patterns. It is based on the derivational capacities of theories. Explanations are construed as arguments and theories are regarded as logical formulae, possibly devoid of truth-value, serving the task of generating implications and saving the phenomena. Per se derivational unification does not imply anything about unity among the phenomena themselves.

By contrast, ontological unification is based on the representational capacities of theories in depicting underlying systems. Explanations are construed as descriptions of the order of things in the world. Theories are regarded as purportedly true pictures of the simplest mechanisms and processes of the world's workings; phenomena are regarded as manifestations thereof. Ontological unification is a matter of redescribing apparently independent and diverse phenomena as manifestations (outcomes, phases, forms, aspects) of one and the same small number of entities, powers, mechanisms, processes. (Aronson 1984; Mäki 1990, 2001)

In GE, rather than excluding one another, derivational and ontological unification seem to be conjoined. Unification is not solely a matter of derivation. The aim is rather to establish a significant fact about the world: agglomeration at different spatial scales can in fact be caused by the same kind of mechanisms.

Talk of “mechanisms” and “forces” is frequent in GE as the following quotations exemplify.

The main reason for looking at these different levels of aggregation is that in explaining clustering, GE shows

that to a large extent the *same basic forces* apply at all levels of aggregation. (Brakman et al. 2001: 2; emphasis added)

By using highly stylized models, which no doubt neglect a lot of specifics about urban/regional/international phenomena, GE is able to show that *the same mechanisms* are at work at different levels of spatial aggregation. (Brakman et al. 2001: 323; emphasis added)

This talk about the “same mechanisms” and the “same forces” supports a view of unification that is consistent with the ontological variant. It also connects with the issue of the role of real mechanisms in explanation. In this respect, it seems to speak against Kitcher’s position and for views such as Salmon’s on which explaining a phenomenon is a matter of laying bare the mechanisms or causal processes that underlie the phenomena we observe and wish to explain (Salmon 1984, 1998; see also Skipper 1999). The mechanistic view of explanation is grounded on the realist thesis that the world has a causal structure independent of our explanatory efforts; Kitcher’s view of explanation instead leaves no room for independent metaphysical concepts such as that of mechanisms.

Viewing GE from yet another epistemic angle suggests that it indeed does subscribe to common mechanisms that are independent of explanation and unification. This can be expressed as an implicit commitment to both realism (about the existence of common mechanisms) and fallibilism (about the possibility of one’s models not getting those mechanisms quite right). It also runs counter to the view that unity in the world is a function of a unifying theory.



The use of highly unrealistic assumptions made for reasons of mathematical tractability (“modelling tricks” as geographical economists call them) has prompted geographical economists to raise questions about how the models relate to the real world. While the employment of radically unrealistic assumptions is often fine for realism, or for the acquisition of significant truths about the world (see for example, Mäki 1994, 2004), geographical economists grant that some of their assumptions are motivated by mathematical tractability rather than ontology – and that the resulting models might therefore not be fully adequate. This is expressed in the following lengthy quote from a major representative.

It has become apparent, however, that while new geography models do make it possible for the first time to put spatial considerations into models rigorous enough to become part of the analytical canon, those models are too simple, too stylized to reproduce the real economic geography of the world very well [...] We might note in particular that the new economic geography [...] suffers to some extent from the temptation to focus on what is easiest to model rather than on what is probably most important in practice [...] Still, there are good reasons why mainstream economics does place a high value on being able to produce tightly specified models—if only to provide the backdrop for less tight, more empirically motivated study. (Krugman 2000: 59)

Rather than as such speaking against realism, this fallibilist attitude reveals a realist attitude: though there are pre-theoretic reasons to believe that there is unity among the phenomena studied, the theoretical constraints employed in order to capture that unity may in the end fail to do the job. The

statement recognizes the very possibility that the unifying theory might fail in describing a real unity or in picking out the correct mechanism or in depicting its precise working or relative relevance. As said, this runs counter to Kitcher's position on which unity amongst phenomena is a function of a unifying theory.

### 2.4.3 Mathematical structure versus ontological convictions

The abstract mathematical framework employed by GE has been crucial in the process of unification. GE is said to be part of "the increasing returns revolution" in economics, which started unfolding with the appearance of the Dixit-Stiglitz model of monopolistic competition in the field of industrial organization. The Dixit-Stiglitz model offered a mathematical tool with which to deal with the presence of increasing returns at the firm level. New trade theory, new growth theory and GE itself all developed as applications of the Dixit-Stiglitz model to their specific fields of study. In particular, a sequence of successive extensions to models of new trade theory resulted in the appearance of the first core model of GE which only adds mobility of workers to its international trade antecedent.<sup>21</sup>

---

<sup>21</sup> Trade theory enjoys a special role in the dynamics of unification of GE. It was first proposed in order to account for the observation that much of trade between countries is of the intra-industry rather than of the inter-industry kind as predicted by its neoclassical antecedent. Krugman (1979) seeks to explain intra-industry trade in terms of a framework of increasing returns to scale and imperfect competition modelled a la Dixit-Stiglitz. The presence of increasing returns explains intra-industry trade between countries thanks to the assumption that consumers have taste for variety. Given the assumption that each variety is produced at only one location, in presence of increasing returns to scale the taste for variety creates an incentive for both countries to trade those varieties of goods they produce thereby giving rise to intra-industry trade. To this model, Krugman (1980) adds the presence of transportation costs and later

Geographical economists maintain that the reason why space had previously been neglected by the mainstream of the economics discipline was the lack of appropriate mathematical tools to deal with increasing returns at the firm level: it is these that are held to be crucial for an explanation of agglomeration. It was then thanks to the Dixit-Stiglitz model of monopolistic competition that economists could deal with spatial issues within a framework that they would find satisfactory.

This is in line with Morrison's observation that in many instances of unification the unifying theory embodies a mathematical structure that plays a prominent role in the process of unification, sometimes even at the expense of ontological considerations (Morrison, 2000). That the abstract mathematical framework has played a fundamental role in the unifying process is obvious. On the other hand, we just discovered that ontological convictions have also been relevant in the development of GE. How do the two relate? To realise this, consider the following passage.

What we find remarkable and gratifying in all of this is the extent to which we are able to use the same modeling "architecture" to address so many issues in seemingly disparate fields. But then our point is precisely that these fields are not that disparate after all: be it urban economics, location theory, or international trade, it's all about where economic activity takes place - and why" (Fujita et al. 1999: 12)

---

Krugman and Venables (1990) adds uneven distribution of economic activity, thereby bringing new trade theory even closer to what would become the GE approach. The first core model of GE, Krugman (1991) only adds mobility of workers to Krugman and Venables (1990) (See Appendix where (2.6) gives the ad hoc dynamics of workers' migration.). As a consequence of this assumption, the size of the market can be endogenously determined by the locational choices and hence agglomeration is explained rather than assumed.

What is here referred to as the “same modeling architecture” relates to the abstract mathematical structure that Morrison talks about and what Kitcher refers to as a pattern of derivation. However, at the same time, Fujita et al. (1999) seem to entertain the idea that unification achieved by applying the same modelling architecture to “seemingly disparate fields” is ontologically grounded: “the fields are not disparate after all” in the sense that they all describe the same generic aspects of how things are in the world, namely “where economic activity takes place - and why”. A similar point is made even more clearly in the following quote.

The empirical phenomena touched upon above have been studied thoroughly from many different angles, based on different theoretical frameworks, for a long time.[...] We have already mentioned that the fractal nature of location phenomena [...] suggests that *similar forces* might be relevant in explaining them. We *therefore* use throughout the book a common structural approach to help understand the phenomena [...] (Brakman et al. 2001: 17-18; emphases added)

The unification of the phenomena studied separately by previous theories has been achieved with the aid of a “common structural approach” or the same pattern of derivation. Yet unification is grounded on the possibility that “similar forces might be relevant” in explaining the unified phenomena, a possibility suggested by the observation that location phenomena have a fractal dimension.

We conclude that both ontological and mathematical considerations have played a role in the unification process of GE. It seems there need not be any conflict between them. We

might put their respective roles by saying that unification has been *facilitated* by mathematics and *motivated* by ontology (even though we admit there is no sharp dichotomy between facilitation and motivation)<sup>22</sup>.

Even though the GE unification has been motivated by ontology, does this allow us to draw conclusions about ontic unity? Morrison (2000) makes a distinction between reductive and synthetic unity. Reductive unity is established when two phenomena are identified as being of the same kind (for example, electromagnetic and optical processes; Ibid: 5). Synthetic unity involves the integration of two separate processes or phenomena under one theory (for example, the unification of electromagnetism and weak force; Ibid: 5). Whereas synthetic unity does not have implications for claims about ontic unity, the ontological implications of reductive unity depend on whether the unification has been merely the product of applying an abstract mathematical structure or whether there are good reasons for believing that the unified processes are really one and the same. What we have said above suggests that the GE unification appears to be of a reductive kind. But are there good independent factual reasons for believing in such a unity as Morrison requires in order to draw justified conclusions about unity amongst the phenomena? Although there seems to be a consensus on the existence of the mechanisms described by GE and on how they work by and large, doubts remain (cf. Martin 1999; Sjöberg and Sjöholm 2002) as to whether they operate in the way depicted by GE and whether, and to what extent, they bear explanatory role in respect to all classes of phenomena purportedly unified

---

<sup>22</sup> Notice that while geographical economists hold ontological convictions about the unity of the phenomena they investigate, they envisage the possibility that the constraints posed by the mathematical structure may

by GE. In this sense, though GE might be seen to pursue reductive unity, a cautious attitude should be adopted when considering whether those phenomena are in fact united precisely in the way depicted. Empirical investigation, which is still at an infant stage in GE (cf. Fujita et al. 1999 and Brakman et al. 2001), should be helpful in settling this issue.

#### 2.4.4 Explanation versus unification

The relationship between unification and explanation is one of the thorny issues in the literature on unification. By considering this issue we will get a better picture of some aspects of unification in GE. In particular, this will help us see the nature of the relevant explanantia and explananda. While Kitcher believes that to explain is to (derivationally) unify, one rival view takes explanation to be a matter of identifying causal mechanisms, whether singular or general (Salmon 1984; Barnes 1992; Skipper 1999; Morrison 2000). On this account, unification may come as an extra bonus. We have seen above that the GE's case seems to support the latter view. Geographical economists believe that explaining phenomena requires invoking unification-independent mechanisms. This suggests that in GE explanation is not the same as unification.

Morrison claims that not only are explanation and unification separate, but there is often a trade-off between the two: "the very project of unifying two theories is, for the most part, at odds with the procedures necessary for obtaining detailed explanatory knowledge." (Morrison 2000: 192). Indeed, she argues,

---

not be conducive to the correct identification of the unifying mechanisms and of their precise workings.

The more general the hypothesis one begins with, the more instances or particulars it can, in principle, account for, thereby “unifying” the phenomena under one single law or concept. However, the more general the concept or law, the fewer the details that one can infer about the phenomena. Hence, the less likely it will be able to explain why particular phenomena behave as they do. If even part of the practice of giving explanation involves describing *how and why particular processes occur*—something that frequently requires that we know specific details about phenomena in question—then the case for separating unification and explanation becomes not just desirable but imperative. (Morrison 2000: 20)

The “how” element of the explanation involves the description of the process that has brought about the phenomenon, and in this sense, the grounds for separating unification and explanation are similar to those suggested by the advocates of mechanical conceptions of explanation. We will next focus on Morrison’s additional idea that unification usually requires the development of a more general concept or law. By virtue of its abstract character, this concept or law includes fewer details about the phenomenon, and therefore has limited explanatory power.

No doubt unification in GE has been achieved at the expense of neglecting details specific to the phenomena under investigation as geographical economists themselves acknowledge:

By using highly stylized models, which no doubt neglects a lot of specifics about urban/regional/international phenomena, GE is able to show that the same mechanisms are at work at different levels of spatial aggregation. [...] In order to lay the

foundations for a unified approach, there is a price to be paid in terms of a neglect of institutional and geographical details. (Brakman et al. 2001: 323)

This supports Morrison's view about the way in which unification functions. But does abstraction from specific details entail that there is a trade-off between unification and explanation, and that a higher degree of unification is responsible for a decrease in explanatory power? As reviewers have also noted (Ladyman 2003; Muller 2001), Morrison's lack of a precise notion of explanation makes her argument about the dissociation between explanation and unification a bit vague. However, two ideas seem to motivate her claim that there is a trade-off between explanation and unification. First, when making the claim she seems to have in mind explanation of particular phenomena. Second, she seems to subscribe to the view that increased understanding is achieved by describing how particular processes occur and that such a description (often) requires knowing specific details about the phenomenon to be explained. We propose both ideas need qualification.

First, insisting on detailed descriptions of processes is too strong. Explaining always implies selection among the many details of the causal history of the phenomenon to be explained. It follows that description of processes can be given at various levels of abstraction and resolution, depending on the kind of explanatory information we are looking for. As many have already pointed out (for example, Garfinkel 1981; Lewis 1986; Jackson and Pettit 1992), one kind of valuable explanatory information concerns what is common across instances of a phenomenon. This requires selecting common details and factors rather than instance-specific ones. As argued elsewhere (Chapter 3 of this thesis), GE can be interpreted as searching for those economic mechanisms that are possibly in operation



across instances of agglomeration independently of spatial and temporal particularities and specificities.

Morrison's second ground for her claim of a trade-off between unification and explanation seems to be associated with explanation of particular phenomena. That is, she says, "the more general the concept or law, the fewer the details one can infer about the phenomena. Hence, the less likely it will be able to explain why particular phenomena behave as they do" (Ibid.: 20). It is not clear whether she holds the view that explanatory power is ultimately to be judged against explanation of particular occurrences. Be that as it may, we submit that if explanation of generic kinds, phenomena and regularities is considered, then a decrease of explanatory power is likely not to occur since these "theoretical" explanations typically require information about what is common across instances of the generic phenomenon.

This is an important point to make, as the characteristic thrust of social science theorizing is to provide theoretical explanations rather than singular ones. In economics—and GE is no exception—theories typically aim to account for what economists refer to as "stylized facts". (A few examples of the stylized facts GE is concerned with are the following: "At the global level, there is a strong correlation between the degree of urbanization and per capita income" (Brakman et al. 2001: 6); "Trade is large between similar countries and dominated by intra-industry trade" (Brakman et al. 2001: 245); "There are frequent changes in economic ranking, known as leap-frogging" (Brakman et al. 2001: p. 186); "In nearly all countries, cities largely specialized in a few activities coexist with much more diversified cities" (Duranton and Puga 1999: 2).

In general, stylized facts are abstractions and modifications of more particular bodies of information: they are generic patterns rather than particular occurrences. To theoretically

explain stylized facts is different from explaining particular bodies of data, as the latter requires more information about the details of the particular case. Therefore, failure to explain particular occurrences does not imply failure to explain generic patterns. While abstract mechanisms, laws and processes might have great explanatory power when explaining stylized facts, the same mechanisms, laws and processes are not sufficient for explaining particular occurrences.

This implies that while it may well be true that in many instances, unification is achieved at the expense of explanatory power with respect to singular occurrences, this is not necessarily so when explanatory power is judged in terms of stylized facts, patterns, regularities and the like. The abstract framework of GE is to be supplemented with a host of specific details when employed to explain particular events such as “the establishment of a hierarchical system [of cities] in the US by the 1870” (Fujita, Krugman and Mori 1999) or the existence of particular industry clusters, such as Silicon Valley. The abstraction from these specific details may very well limit the explanatory power of GE with respect to those particular occurrences; and their introduction might require major modifications of the theoretical framework, or they might even resist integration within the straightjacket of the GE framework. Yet, GE might still be capable of explaining the kinds of stylized facts mentioned above.

#### 2.4.5 Multiple mechanisms and degrees of explanatory unification and of ontic unity

We have seen that unification in GE proceeds by way of isolating the pecuniary externalities mechanism that is believed to be common to cases of agglomeration at different scales, and thus to constitute the unity among these cases. Given the

concession of the possibility of multiple effective mechanisms, including those not modelled by GE, we can now develop the twin ideas of *degrees of explanatory unification* and *degrees of ontic unity*. We can identify the following possibilities.

- The unifying mechanism is always the only mechanism in operation.
- The unifying mechanism is not always the only mechanism in operation, but it is always the main mechanism in operation.
- The unifying mechanism is not always the only mechanism in operation, and is not always the main mechanism, but is always in operation.
- The unifying mechanism is not always the only mechanism in operation, is not always the main mechanism, and it is not always in operation in all instances of the classes of phenomena it unifies.

As we descend from the top to the bottom of this simple taxonomy, two things happen. On the one hand, the unifying power of the mechanism (or the theory depicting it) will increase in the sense that an increasing number of kinds of phenomena is likely to be governed by it. On the other hand, the strength of ontic unity amongst the phenomena will decrease in the intuitive sense that the firmness of the hold with which the mechanism governs them gets weaker.

In GE, the unifying mechanism, namely the pecuniary externalities mechanism of agglomeration, might, at least in principle, be sufficient to bring about and sustain agglomeration even in the absence of other agglomerating

factors. However, in reality, pecuniary externalities rarely (if ever) act in isolation. Moreover, the GE mechanism may not even be the main mechanism at work in every case since, as geographical economists themselves point out, in certain cases, for instance cities, technological externalities are likely to be more relevant. This covers the first three cases in the above taxonomy. In regard to each of them, we are prepared to say that unification (of varying degrees) takes place.

It is the fourth case that is problematic. Pecuniary externalities might indeed not be necessary for agglomeration, because, in principle, agglomeration may be a chance occurrence or perhaps the outcome of the design of a dictatorial planning agency. Naturally, we may be hesitant to say that in the fourth case unification happens at all: the mechanism simply fails to unify those phenomena with respect to which it plays no causal role. In other words, if it fails to govern particular members of a class of phenomena, then it fails to unify that class with other classes. On the other hand, we might try to accommodate this case by extending the idea of degrees of unification. We may just say that the higher the proportion of members of a class of phenomena that are governed by the mechanism, the higher the degree of unification, and of unity between this class and others. We should add that there is nothing very dramatic about this situation, given that in economics, and the social sciences more widely, general claims characteristically admit of exceptions. What this suggests is that the kind of unification pursued and achieved by GE unifies large classes of phenomena while permitting relatively weak ties of ontic unity among them.

If phenomena of agglomeration are governed by multiple mechanisms, then their explanation will require multiple theories supplementing GE. The set of such supplementary theories is likely to include also the unificandum theories: those

whose domains GE has sought to unify. The various inter-theoretic relations that emerge are bound to be somewhat complex: a subject for another study.

## 2.5 Conclusion

We have shown that explanatory unification has played an important role in the development of GE. With the aid of existing accounts of unification in the philosophy of science we have attempted to provide a picture of the structure of unification in this specific context.

In particular, we discovered that unification in GE has been a matter of applying similar patterns of derivation to different kinds of explanandum phenomena, rather than a matter of minimizing the number of premises and maximizing the number of conclusions. The GE pattern of derivation is the abstract mathematical structure that enabled the process of unification. As such, unification in GE is closer to Kitcher's account than to Friedman's one. On the other hand, two features of the GE unification cannot be captured by Kitcher's view of unification. First, unification in GE has proceeded by showing how the same mechanisms and forces, being at work at different spatial scales, can produce different kinds of phenomena. Second, geographical economists are prepared for the possibility that the mathematical structure has imposed excessively severe constraints on unification such that the outcome might not adequately capture the real unity in the world. This gives us a picture of unification in GE that is in terms of mechanisms that exist independently of explanation and unification.

Both mathematical and ontological considerations have played a role in driving unification in GE. In particular, unification has been motivated by ontology and facilitated by

mathematics. Our case study thus lends support to Morrison's claim that unification in science often proceeds with the aid of an abstract mathematical structure. Yet, no such support is found regarding her claim that unification often proceeds at the expense of explanatory power. Since GE aims primarily at explaining stylized facts, the abstract mechanisms captured by the GE mathematical structure might well be explanatory.

On the other hand, GE's capability of explaining particular occurrences might be very limited. Explaining particular occurrences not only typically requires the inclusion of a host of details about the specific situation, but also that of other mechanisms not encompassed by the unifying theory. This appears to be the case in GE where the unifying mechanism is not always the only mechanism in operation and in some cases it is not even the main one, implying a high degree of explanatory unification, but a low degree of ontic unity.

With some conceptual adjustments, we have been able to identify and describe a special kind of unification in one recent branch of social science. From our analysis it follows that although none of the philosophical accounts we drew upon fully captures the kind of unification characteristic of GE, each of them has enabled us to highlight some of its aspects. This provides further confirmation to the view that unification in science is not uniform but rather adopts different forms and strategies depending on disciplinary contingencies and on the peculiarities of the unified domain (cf. Morrison 2000).

### **3 Geographical economics versus economic geography: towards a clarification of the dispute\***

#### **3.1 Introduction**

Geographical economics is an attempt to incorporate the role of space into conventional economics, initiated by Paul Krugman at the beginning of the 1990s under the label of 'new economic geography' (see for example, Fujita and Thisse 1996; Fujita et al. 1999; Krugman 1991a, 1991b, 1996; Krugman and Venables 1996; Puga and Venables 1997; Venables 1996a, 1996b). Economic geographers who have traditionally been concerned with similar issues, rather than welcoming this 'new' interest by economists in spatial issues, strongly criticise geographical economics on a number of grounds (for example, Dymski 1996; Martin 1999; Martin and Sunley 1996; Sheppard 2001; Sunley 2001).

I will put such criticisms under scrutiny taking Ron Martin's 1999 contribution as representative. His critical review of the 'new geographical turn in economics' has had a remarkable impact on geographers partly because it reflects their common view on geographical economics and on neo-classical economics more generally. Martin claims that GE "represents a

---

\* This paper has been published in *Environment and Planning A* (Marchionni 2004)

case of mistaken identity: it is not that new, and it most certainly is not geography" (1999: 67).<sup>23</sup> This claim is accompanied by three fundamental objections to GE: its 'myopic view of theory', to which I refer as the *methodological criticism*, its 'commitment to mainstream mathematical economics' and 'its neglect of geography', which I lump together as the *isolation criticism*. Examining these objections in some detail involves touching upon key issues such as those of realism, realisticness<sup>24</sup>, concepts of theory and model and the method of isolation. One purpose of this exercise is to elicit certain features of GE and by so doing to partly contribute to the clarification of the ongoing dispute between geographical economists and economic geographers.

To do this, I first introduce the dispute between GE and economic geography (Section 3.2). Then I examine Martin's *methodological criticism* and argue that it rests on a narrow conception of realism which prevents an appreciation of how realism can be compatible with the use of unrealistic assumptions (Section 3.3). On a more abstract notion of realism attuned to the wider philosophical discussion, GE turns out to be compatible with realism. I show this in section 3.4. Once it has been established that both economic geography and GE are compatible with realism, the dispute emerges as one between

---

<sup>23</sup> I take Martin as representing the stance of the economic geographers towards GE. In the SSCI (May 2004), Martin's 1999 paper is cited in fifty-four articles. It is also discussed by The Economist (1999) and, as I will show in section 3.2.2, similar points are made by other economic geographers. Though Martin's criticisms are levelled against GE and new growth theory, my discussion is on GE alone. Within the latter, Krugman plays a prominent role primarily because of his commentaries upon methodological aspects of GE.

<sup>24</sup> For a detailed discussion of (un)realisticness see Mäki (for example, 1992a, b, 1998c). As will be seen later, 'realisticness' refers to "a set of properties of theories and their constituent parts" (Mäki 1998c: 409) and is to be distinguished from realism which refers to a collection of philosophical doctrines.



distinct theoretical isolations. In fact, the other two sets of objections Martin levels at GE are directed at what factors GE isolates as explaining spatial agglomeration. Martin's point is that GE does not take into account factors that economic geographers deem important for understanding spatial agglomeration. I show in section 3.5 that the main thrust of the dispute revolves around the concepts of horizontal and vertical isolation. I notice that the theoretical isolations of GE depend on and in turn determine what aspect of spatial agglomeration it can possibly explain. I hence suggest a way to formulate the explanandum of GE which helps to make sense of its theoretical isolations (section 3.6). The precise delimitation of the GE explanandum suggests that Martin's criticisms fail to hit the target, but more importantly, it allows us to look at the relationship between economic geography and GE in a new light. Taking this as a point of departure, in section 3.7, I ask whether beyond the methodological divide there are grounds for complementarity between the two approaches.

## **3.2 The dispute**

In this section I summarise the intended contribution of GE, the critical reactions of economic geographers and a response from the GE perspective.

### **3.2.1 An overview of geographical economics**

The location of economic activity in space had been a subject somewhat neglected by (mainstream) economists (cf. Blaug 1996). According to Paul Krugman, the reason is that economists had been unable to handle imperfect competition, a feature of markets where increasing returns to scale at the level of the individual firm are present. Yet increasing returns to scale are precisely what enforce concentration of some activities

in space. This is why GE "may be regarded as the fourth [...] wave of the increasing returns-imperfect competition revolution" in economics (Krugman 1999a: 93) that began in the 1970s in the field of industrial organisation. The same analytical tools (namely, models of imperfect competition in presence of increasing returns), had been applied first to international trade and then to technological change and economic growth and finally to the development of GE (Ibid. 93). Hence it has been the availability of certain modelling techniques that cleared the way for the final recognition of economic geography by the economics discipline.

In Krugman's characterisation, GE is a

"genre" or a style of economic analysis which tries to explain the spatial structure of the economy using certain technical tricks to produce models in which there are increasing returns and markets characterised by imperfect competition (Krugman 1999a: 93).

The explanandum of GE is the spatial structure of the economy, more precisely, the "concentrations of population and of economic activity: the distinction between manufacturing belt and farm belt, the existence of cities, the role of industry clusters" (Fujita et al, 1999: 4). The modelling tricks that characterise GE are the use of the Dixit-Stiglitz model of monopolistic competition, the modelling of transport costs as the fraction of a good shipped that melts away in transit (Samuelson's iceberg costs), the employment of ad hoc dynamics and the use of computer simulations to supplement analytical results. Modelling tricks are understood as "assumptions that reflect not so much a realistic view of how the world works as a judgement about what will make the analysis of geographic issues manageable without doing too much damage to the relevance of that analysis" (Fujita et al.

1999). GE thus faces a trade-off between tractability and realisticness seemingly solved in favour of the former. Because this is one of the charges economic geographers level at GE I will have more to say about this later. For now let us look at how the models of GE explain the spatial structure of the economy.

The spatial structure of the economy is analysed as the result of the interaction between agglomerating and dispersing forces. Table 3.1 lists (some of) these forces: the left-hand column represents the sources of external economies, the centripetal forces responsible for the agglomeration of economic activity (the Marshallian local external economies); the right-hand column lists the forces that act against the local external economies, that is the centrifugal forces that push for the dispersion of economic activity.

Table 3.1. Centripetal and centrifugal forces

<i>Centrifugal forces</i>	<i>Centripetal forces</i>
Market size effects (linkages)	Immobile factors
Thick labor market	Land rents
Pure external economies	Pure external diseconomies

Source: Krugman (1999b: 91); Fujita et al. (1999a: 346)

Though “in the real world not only agglomeration in general, as well as any example of it, typically reflects *all* items on the menu” (Krugman 1999b: 91), GE typically focuses only on the interaction between market-size effects as the centripetal force and the presence of immobile factors as the centrifugal force (first line of table 3.1).

Market-size effects are held possibly to give rise to cumulative processes of agglomeration when increasing returns and transportation costs are both important. The basic idea is that firms typically want to locate close to their customers and

suppliers (or to the suppliers of goods required by their workers). If for some reason a certain place already has a concentration of producers, then it will offer a larger market because of the demand of the producers and workers (backward linkages) and a good supply of inputs and consumer goods (forward linkages) (Fujita et al. 1999: 7). Even abstracting from advantages inherent to particular localities (that could be identified at the local, regional or national scale), a slight difference in the size of the market in one locality, through the operation of forward and backward linkages<sup>25</sup>, could trigger and sustain cumulative and self-reinforcing processes of agglomeration and thus 'lock in' a certain pattern of evolution in that locality. The systems are usually characterised by the existence of multiple equilibria; hence whether and where concentration will take place depends on the relative strength of the two counteracting forces and on initial conditions.

This basic framework depicting the tension between market-size effects and immobile factors, is employed, with little variations, for providing insights into apparently diverse empirical phenomena such as the distinction between manufacturing belt and farm belt, the existence of cities, the role of industry cluster, the dynamics of the product cycle in international trade.

---

<sup>25</sup> The concepts of backward and forward linkages and their role in producing processes of cumulative causation have been proposed by Hirshman (1958). See also Myrdal (1957) and Kaldor (1978) for the concept of cumulative causation, especially in the context of regional development. Krugman offers an interesting interpretation of how these insights, though important, were neglected by the discipline of economics because they were not mathematically formalised (and not even formalisable at the time) (See Krugman 1995).

### 3.2.2 The reaction of economic geographers

Many economic geographers hold a negative attitude towards neo-classical economics. In a recent intervention Amin and Thrift (2000) warn economic geographers of the dangers of a rapprochement with economics for economic geography would most likely become its "prey". It is perhaps not surprising that, though GE aims at bringing space and geography onto the research agenda of economists, economic geographers are generally not persuaded. GE is blamed for its imperialistic tendencies (Johnston 1992); for being founded on an inadequate rationalist justification (Barnes 2003); and for inspiring the feeling of *déjà vu* (Berry 1999; Hoare 1992; Martin 1999). On methodological grounds, GE is often criticised because it employs formal mathematical models based on neo-classical economics that do not allow the complexity of real-world phenomena to be captured (Clark 1998; Dymski 1992; Martin 1999). For Clark "there is a suspicion that analytical elegance and tractability drive the focus of analysis rather than empirical problems" (Clark 1998: 75; see also Martin 1999). On more substantial grounds, economic geographers have identified a series of issues that are poorly treated or even omitted altogether in the models of GE and that geographers believe are important. For instance, GE is held to downplay the key role of technological and knowledge spillovers in explaining agglomeration (cf. Olsen 2002) and the importance of "macro-structural factors" in determining the location decisions of firms (Dymski 1992). It is also charged with paying too little attention to the issue of "spatial contingency" which results in a failure to acknowledge "spatial interdependence" (Sjöberg and Sjöholm 2002). And finally, economic geographers noted that GE deals poorly with the problem of aggregation in the sense that it overlooks the possibility that different mechanisms may be at work at different spatial scales (Martin

1999; Sjöberg and Sjöholm 2002). Though certainly not exhaustive, this overview should hint at the kind of criticisms economic geographers have directed at GE.

### 3.2.3 The response of geographical economists

Geographical economists have virtually ignored the criticisms coming from the economic geographers' camp. An interesting exception is Brakman, Garretsen and van Marrewijk (2001) who explicitly address Martin's 1999 criticisms. They concede that *geography* and its sociological, psychological and institutional components do play an important role in explaining agglomeration. They also recognise that some of the theoretical and empirical insights of GE are in fact not new. Yet they argue that the novelty of the approach lies in the way in which it deals with economics and geography rather than in its research topic. To them, the contribution of GE is twofold. First, GE has been capable of showing that the same mechanisms are at work at different spatial scales, an achievement reached at the cost of neglecting "institutional and geographical details". And second, it is "the only field of economics which provides a micro-economic foundation in a consistent general equilibrium framework for the spatial distribution of economic activity" (Ibid.: 323).

### 3.3 The methodological criticism

Martin (1999) observes that the works of geographical economists and economic geographers are "quite distinct methodological and epistemological genres" (Martin 1999: 81; see also Amin and Thrift 2000; Clark 1998, Dymski 1992). From the methodological standpoint, he criticises GE for its

employment of “positivistic accounts” unlike geographers who have abandoned them “in favour of realist approaches”.

A realist approach is held to be one in which “explanations are built ‘from below’, often relying upon close dialogue with individual agents and organisations, and linking this ‘local’ knowledge with wider, larger stylised facts and conceptual frameworks” (Ibid. 81). This is opposed to the “positivistic accounts” of GE characterised by “deductivist, mathematical demonstration” in which “mathematical tractability drives the focus rather than the apparent diversity of economic systems” and “the complexity of the economic landscape is one of mathematical solvability, rather than empirical messiness and particularity” (Ibid.: 81; Clark 1998).

Although there is no doubt that the methodological commitments of economic geography and GE set them apart, I argue that Martin’s characterisation of the former as realist and the latter as positivist is based on a rather narrow conceptions of both philosophical positions<sup>26</sup>. Relevant for the subsequent discussion is the notion of realism. As it is generally understood in philosophy, realism refers to a collection of theses about what exists (and how) and about what relationship holds between our representations and the world and not to a method whereby theories are formulated (see Mäki and Oinas 2004). Accordingly, confusion can be avoided by saying that certain methods of theory formation are compatible with realism while others are not, rather than talking of “realist methods”. And more importantly, despite emphasis on formalism and

---

<sup>26</sup> Martin’s usage of the terms is influenced by the way in which positivism and realism are perceived in human geography. Johnston (1985) and Sayer (1985) give a historical account of the reception of positivism and realism in economic geography. Mäki and Oinas (2004) examine the “narrow” notion of realism of economic geographers and how that differs from the one discussed by philosophers of science.

tractability, models in GE might be compatible with this view of realism.

Martin's argument proceeds by noticing that these two methods lead to different notions of theory. GE endorses a "myopic view of theory" apparently meaning that theory "is assumed to be synonymous with formal, mathematical model building" (Ibid.: 81). In economic geography, instead, "the dominant mode of theorising is one of discursive persuasion" that "permits the construction of much richer maps or representations of reality" (Ibid.: 82). In Martin and Sunley (1996: 268), the contrast is posed in terms of the "aim at being realistic" of economic geographers and the "unrealistic assumptions" of GE.

Although it may well be correct that the GE models are unrealistic in a variety of senses, their representations of the core of real-world phenomena might nonetheless be true (or approximately true). And, if GE employs unrealistic assumptions to get at the core of real-world problems, then it is quite compatible with realism as a philosophical doctrine about theories. In order to understand whether GE's philosophy and method are actually incompatible with realism, as suggested in Martin (1999) and Martin and Sunley (1996), we need to distinguish between the concepts of realism and realisticness. The implications of this distinction go well beyond terminological clarity for this sheds a new light on GE and perhaps reduces its distance from economic geography, without, however, downplaying the methodological divide between the two approaches.

### **3.4 Unrealistic assumptions and realist philosophy**

On several occasions Uskali Mäki (1992a, 1992b, 1998a, 1998b, 1998c, 2003) has insisted on the need to distinguish between realism and realisticness, two concepts that though



partly related are in fact quite distinct. The distinction proves useful because it serves to show that realism and unrealisticness are compatible in a way that Martin may have failed to recognise.

The notion of *realism* we are interested in here is realism as a philosophical theory about scientific theories. Roughly, it holds that scientific theories can refer to real-world entities, that they can represent those entities by attributing properties to them, and that they can be held to be true or false representations of those entities on the basis of how they relate to the way the world works (for example, Mäki 1992a).

The concept of *realisticness* instead refers to "a property or a set of properties of theories and their constituent parts" (such as observability, truth, plausibility and partiality) (Mäki 1998c: 409). Unrealistic theories or assumptions can be compatible with the above notion of realism. In order to appreciate the senses of unrealisticness relevant to our discussion, a further distinction is to be employed, that between the *whole truth* and *nothing but the truth* (Mäki 1992a, 1992b, 1998a, 2004a). First, scientific theories are bound to be unrealistic in the sense of being partial representations: they isolate one part of reality from the influence of the rest of reality. As such they do not tell the whole truth. Second, one way to theoretically bring about theoretical isolations is by way of assumptions that are unrealistic in the sense of being false about what they refer to; therefore failing to convey nothing but the truth. Idealisations, (when a certain variable is falsely represented as having value 0 or  $|\infty|$ ), exaggerations and simplifications are unrealistic in this latter sense (Mäki 1992a, 1992b).

As Mäki's framework clarifies, the systematic violation of the whole-truth and nothing-but-the-truth criteria can be compatible with realism when falsehood is about inessential or irrelevant features of the phenomenon under investigation, or

about features only temporarily isolated, and hence when it serves the pursuit of truth about essential or relevant features. (Mäki 1998a; for the framework in action as applied to von Thünen's theory see Mäki 2004a)

### 3.4.1 Realism in geographical economics

Krugman is “at once theoretician, prescriptive methodologist and historian of the theory he produces” (Meardon 2002: 233). His writings ‘as a methodologist’ show that despite emphasis on formalism and tractability he endorses realism as a theory about theories<sup>27</sup>.

For Krugman every science needs to find a way to deal with complexity. The fact that the only fully correct model of a system is the system itself leads to the necessity of “leaving out many aspects of reality” (Krugman 1995: 69), that is, to violate the whole truth. The decision of what to put into a model is constrained by 1) modelling techniques and 2) other resources (such as money and patience). These two constraints delimit a wide variety of possible models and the one that is actually built “depends on educated guessing”. We will see that this educated guessing may be said to be about the

---

<sup>27</sup> The focus here on Krugman's writings is motivated by the fact that he has provided quite extensive and sophisticated commentaries on these issues. His views are nonetheless taken as representative of geographical economists. There are good reasons to suppose that most geographical economists (and perhaps many economists in general) hold similar views. Allegations such as, “the economic geography approach [...] is by no means the only tools for *understanding* the shaping of the space-economy. It is our contention, however, that the qualitative results presented [...] are fairly robust and are *representative* of general tendencies at work in *contemporary economies*” (Fujita and Thisse 2003: 40, emphasis added), are fairly common and suggest that geographical economists aim at understanding real-world phenomena by way of models thought to be capable of isolating real tendencies. This reveals a realist understanding of the activity of theorizing.

relevant/essential/primary forces that bring about the phenomenon to be explained and as such it provides reasons for ascribing to Krugman a realist attitude towards models and theories<sup>28</sup>.

To begin, consider the following passage.

[E]volutionary theorists find what is sometimes derisively called ‘bean-bag genetics’ a tremendously *useful* fiction: literally untrue though it may be, it is a powerful intellectual tool, offering a way to cut through the essence of complex higher-level phenomena in a way that a ‘realistic’ appreciation of the genetic mechanisms never could. The same is true of neoclassical economic analysis (Krugman 1996: 135-136)

Krugman argues that neoclassical economic analysis, to which he fully subscribes, is a *useful* fiction, or rather it employs useful fictions, and in that regard it is literally untrue, in the sense that it violates the nothing-but-the-truth criterion. Yet it is precisely thanks to those falsehoods that economics is capable of getting at the *essence* of complex phenomena.

Try to understand why countries specialize and trade in terms of realistic descriptions of human psychology, and without abstracting from the disequilibrium that characterizes global markets most of the time, and you will end up with mush—or with something that you imagine is a sophisticated analysis but is really a crude set of misconceptions hidden beneath literary pretensions. (Krugman 1996: 136)

---

<sup>28</sup> Models are to be distinguished from theories. For simplicity, we can think of the GE theory as a collection of models where each model is a more specific and more isolative version of the theory (cf. Mäki, 2004b).

The point is that for Krugman “there is no alternative to models. We all think in simplified models, all the time. The sophisticated thing to do is not to pretend to stop, but to be self-conscious—to be aware that your models are maps rather than reality” (Krugman 1995: 79). Though, for Krugman, models and their assumptions are useful fictions, he does not endorse the view that untrue models and assumptions are useful in the sense instrumentalists would have it—as the geographer Eric Sheppard (2001) seems to suggest in ascribing a Friedmanian-like methodology to GE. While instrumentalists would be happy to retain untrue assumptions and models if they were good instruments for, say prediction, to Krugman they are to be used primarily insofar as they enhance understanding. When an untrue assumption fails to serve the purpose of understanding, then it ought not to be retained at all costs. The following passage is suggestive in this respect.

[...] it would be better if economists were more self-aware—if they understood that their use of maximization-and-equilibrium [...] is an useful fiction rather than a principle to be defended at all costs. If we were more modest about what we think our modelling strategy is doing, we might free ourselves to accommodate more of the world in our analysis. (1999b: 27)

Krugman’s methodological writings thus reveal reasons for ascribing him a realist stance. He believes that the very act of understanding real-world phenomena implies the need for isolation, that is the violation of the whole truth; and, moreover, he maintains that employment of false assumptions, violating nothing but the truth, is a means by which truth about the relevant aspects of the phenomenon can be attained. On the other hand, realists may feel uncomfortable with Krugman’s strategy: recall that for Krugman the decision of what to include

in the model (the isolated features) seems to depend in the first place on modelling techniques and available resources. Only within these constraints do ontological considerations of relevant versus irrelevant features determine the choice of the isolated terms. Such a view may conflict with a realist stance. Though on a realist view both tractability and ontological considerations may drive the use of false assumptions, the latter should predominate over the former (Mäki 1998a: 310). And in Krugman's case, it may well be that, if the relevant forces operating in the real world were intractable or if turning them tractable was overly costly, then they would be isolated away in the first place. This should prompt questions about his realism.

To qualify as realist, however, Krugman does not have to believe that each and every model actually isolates the relevant forces operating in the real world insofar as he holds that it is the goal of scientific inquiry to pursue truth about relevant aspects within the constraints imposed by pragmatic considerations. This means that an individual model may well fail to isolate the relevant forces, but that the progress of the scientific inquiry along with a widening of the available modelling techniques should aim at the identification of the essential elements of real-world phenomena.

#### 3.4.2 Realism and the choice of the isolated forces

GE typically focuses on the interaction between only two of the forces held responsible for shaping the spatial structure of the economy (see Table 3.1 above), namely market-size effects and immobile factors. If those forces are isolated because they are regarded as the relevant causal factors, then we can safely conclude that the GE strategy of theorising is compatible with a realist position.

Geographical economists admit that “[i]n the real world not only agglomeration in general, but any particular example of

agglomeration, typically reflects *all* items in table 1" (Krugman 1999a: 4, emphasis in the original). The choice of the forces to model however is dictated "less by empirical judgement than by [...] strategic modelling considerations" (Krugman 1999a: 5).

This last consideration lends support to the economic geographers' point: given that tractability considerations predominate over ontological ones, GE' strategy of theory construction and model building appears not to be compatible with a realist stance. Yet consider the following passage:

Still, we believe that it would be useful to carry out a more systematic exploration of the implications of our menu, to inquire into the behavior of models in which multiple centripetal and centrifugal forces are operating, to ask how the predictions of those models depend on the relative importance of those forces. Only by carrying out such an exploration will we be in a position to interpret the results of the obvious next step: empirical research (Fujita et al. 1999: 346)

It appears quite clearly that though the exigencies of modelling has led to the analysis of only two forces actually operating in the real world, it is important, especially for interpreting empirical results, to take the omitted forces into account. Such a position may not contrast with a realist philosophy after all if the omission of the other forces is regarded as an early-step assumption<sup>29</sup> required for exploring the real working of the isolated forces, but to be relaxed at later stages of the theoretical endeavour<sup>30</sup>. The other forces will be introduced later for assessing their respective relevance in accounting for

---

<sup>29</sup> On the employment of this terminology, see Mäki (2000).

<sup>30</sup> Recent developments indeed take into account the operation of other centripetal or centripetal forces such as for instance knowledge spillovers (Baldwin and Forslid 2000) or congestion (Brakman et al. 1996)

agglomeration. The three sets of forces are, at least for the time being, all regarded as relevant for explaining agglomeration.

It thus appears that, though employing different methods, both geographers and economists aim at explaining real-world phenomena. If this is so, then one important point of disagreement would be about the factors deemed relevant for explaining the phenomenon under study. This is in fact the issue to which the other two sets of Martin's criticisms refer and to which I now turn.

### **3.5 The isolation criticism**

Martin objects to GE on two further grounds: it is strongly committed to mathematical mainstream economics and neglects real places. Though Martin keeps the discussion of the two objections separate, I will discuss them together as the *isolation criticism*. Both sets of criticisms in fact refer to factors that economic geographers find relevant for understanding the phenomenon of agglomeration but are generally ignored in GE. In brief, economic geographers suspect that GE has isolated too few explanatory factors.

Martin maintains that because of the commitment to mathematical mainstream economics, the "'messy' social, cultural and institutional factors [...] are assumed to be of secondary importance" (Ibid. 75, see also Amin and Thrift 2001 for a similar point in regard to economics more generally). But, he argues, "it is precisely the social, institutional and political embeddedness that can play a key role in determining the possibilities for or constraints in development, and thus why spatial agglomeration of economic activity occurs in particular places and not others" (Ibid. 75). The heavy dependence on modelling has also led to "serious misrepresentations of processes that are deemed to be important" by GE, viz. history and path-dependence. In Martin's view GE has neglected the

“real and context-specific periods of time over which actual spatial agglomeration have evolved” and “the locally-embedded and emergent socio-historical process of technological, institutional and social evolution” (Ibid. 75-76). Moreover, “structural change and hence dynamic, qualitative aspects of spatial development” (Ibid. 76-77), which are alleged to be key to today’s economic landscape do not find place in the GE models because of its commitment to mainstream economics. And finally, what lacks in the GE models is *geography*, intended as “real communities in real historical, social and cultural settings with real people, going about the ‘ordinary business of life’” (Ibid. 78).

Such criticisms, however, may fail to hit the target once we delimit precisely the explanandum of GE. This is not to argue that those factors are actually not important for understanding agglomeration. The point is instead that their omission is determined and in turn determines what aspect of spatial agglomeration GE can possibly explain. To show this I employ a further distinction in the framework of theoretical isolation, that between horizontal and vertical isolation (Mäki 2004a).

*Horizontal isolation* is an isolation that occurs at a given level of abstraction and refers to the fact that a theory focuses on a limited set of objects or properties. *Vertical isolation*, by contrast, is isolation that involves a change in the level of abstraction. In vertical isolation, a universal or quasi-universal is isolated from all its particularities and specificities in time-space. An example may help clarifying the distinction. Horizontal isolation is involved for instance when one or a few industries are isolated from all the other industries making up the economy as a whole. Vertical isolation implies the isolation of an abstract industry from the particularities of concrete industries in time and space (cf. Mäki 2004a).



In this terminology, GE has isolated “agglomeration” from contextual and local specificities and particularities in order to focus on those features that have a chance to be common across instances of the phenomenon. Martin’s criticisms instead ask for the opposite operation, that is, vertical de-isolation: to Martin what geographical economists should include in their models are (historical and spatial) specificities of concrete cases of agglomeration. Nonetheless to do this would require a decrease of the level of abstraction and generality of the GE explanation.

Again, this is not to deny that the omission of historical and context specific factors can in fact mean that GE is “unable to tell us *where* it [industrial localisation and specialisation] actually occurs, or why in *particular places* and not in others” (Ibid.: 78), as Martin claims. But then one should ask whether these are the only questions worth posing.

Consider the following reflections of Fujita et al. (Ibid: 1) on St. Martin’s Court, a street where sellers and second-hand books and prints are clustered:

Why, then, have the shop’s owners chosen to be there?  
To be near each other. No doubt there is some interesting story about how that cluster (...) originally became established, but what sustains it now is a sort of circular logic.

Two distinct questions can thus be asked about a cluster: how it originally became established and what is it that sustains it now. According to geographical economists, although the answer to the former question tells a story about that particular cluster (‘real history and real place’), the second involves abstracting from that story and focusing on the mechanism supposed to be common to many, if not all, instances of agglomeration.

Suggestive in this respect are Sayer's considerations (1992) on how the rise of Silicon Valley could be explained:

A narrative would take us through a series of events, leading in quasi-teleological fashion to its conclusion: the chance location of Stanford University, with its strong electrical engineering interests and the presence of Frederick Terman [...] On the other hand, analysis would abstract from the historical accidents and apply concepts like 'agglomeration economies' and 'vertical disintegration' which purport to be applicable to a whole range of cases of industrial development (Sayer 1992: 259).

Both narrative and analysis explain the rise of Silicon Valley, but they serve different purposes<sup>31</sup>. I suppose Sayer would agree that questions such as "where agglomeration occurs and while in particular places and not in others" cannot be fully addressed by analysis. The reason is that an answer to these questions is likely to depend not solely on generic mechanisms but on local and historical specificities as well.

### **3.6 Towards a clarification of the dispute**

When addressing the *methodological criticism* I argued that, though Martin is right in observing the methodological differences between geographers and economists, these need not be put in terms of different philosophical doctrines. Geographical economists could adopt a realist attitude towards

---

<sup>31</sup> "The economizing view of theory [analysis] is more appropriate to abstraction of objects (relations, mechanisms, concepts) which are stable and pervasive, while thick description [narrative] is more appropriate for accounts of concrete situations in which there is considerable historical specificity and change" (Ibid. 262).

theories while employing highly unrealistic assumptions in formal mathematical models. Once this was acknowledged, the dispute was interpreted in terms of competing isolations of what are regarded as the relevant features of the phenomenon under study.

A fundamental point needs to be made in this respect. The role of isolations in one's theory is strictly related with one's method of theory formation and with one's conception of theory. If economic geographers and geographical economists do entertain different conceptions of how theory is to be constructed this has consequences for their respective isolative strategies and styles. The GE models are very isolative and therefore unrealistic (violating both the whole truth and nothing but the truth) partly in virtue of the role tractability plays in them. It may well be that an analysis conducted through formal mathematical models (necessarily involving a greater degree of idealisation) is less suited for understanding the role of certain relevant causal factors (perhaps those of a context-specific kind) because of the stronger need to isolate and idealise.

The last remark is in fact one of the fundamental objections levelled by economic geographers: it is precisely because of the commitment to mathematical mainstream economics that GE neglects relevant aspects of real-world phenomena. Interpreted in terms of rival theoretical isolations determined by different beliefs on how the world works (and not solely by tractability considerations), the problem for geographers is twofold: (a) certain factors cannot be ignored if the working of real-world phenomena is to be understood; (b) mathematical formal modelling does not allow us to capture the working of those very factors.

In principle space-time specificities can be incorporated in an abstract model without a decrease in the level of abstraction.

For instance, as GE actually does, history could be incorporated by letting it determine the initial parameters of the model. Yet this is not what geographers seem to have in mind when they urge geographical economists to take into account contextual and local specificities and particularities. Recall Martin's dissatisfaction with the way in which history and path-dependency are dealt with in the models of GE. Martin's criticism therefore calls for a decrease in the level of abstraction and hence in the generality of the GE models.

Yet, as suggested in the previous section, this kind of criticism may fail to hit its target if the GE explanandum is looked at in light of its vertical and horizontal isolations. Hence:

- GE aims to investigate the influence on agglomeration and thus on the shaping of the spatial structure of the economy of a certain mechanism, namely, the interplay of market-size effects and immobile factors. In order to do so, the mechanism has been horizontally isolated from the operation of other causal factors, in particular the other centrifugal and centripetal forces.
- GE aims to investigate how that mechanism could be taken to operate across instances of agglomeration independently from particularities and specificities of concrete cases. In order to do so, the explanation is given at a high level of abstraction by way of vertically isolating the general mechanism from those specificities and particularities that in fact may affect the actual outcomes of its operation.

And it has been precisely thanks to these horizontal and vertical isolations that GE can claim to be capable of (a) explaining why and how agglomeration could emerge and then persist only as a result of the tension between market-size effects and immobile factors and (b) suggesting that a variety of different kinds of phenomena (such as the distinction between manufacturing belt and farm belt, the existence of cities and the role of industry cluster) may be explained as the result of the

same set of causal factors, thus resulting in explanatory unification<sup>32</sup>.

These observations seem to be in line with the geographical economists' own conception about the explanatory purposes of their theories. Brakman et al. (2001) describe what they regard as the contributions of their field. The passage is worth being quoted here at length.

By using highly stylized models, which no doubt neglect a lot of specifics about urban/regional/international phenomena, geographical economics is able to show that the same mechanisms are at work at different levels of spatial agglomeration [...] In order to lay the foundations for a unified approach, there is a price to be paid in terms of a neglect of institutional and geographical details, as the aforementioned criticisms make clear. But this is a price worth paying initially, certainly in view of our optimism that a number of these voids will be filled as geographical economics keeps developing in the future. (Brakman et al. 2001: 323)

---

<sup>32</sup> Martin (1999) does not seem to value such an attempt because for geographers "scale matters" (see also Sjöberg and Sjöholm 2002). In response, geographical economists argue that "a few general principles seem to govern the formation of distinct agglomeration even though the content and intensity of the forces at work may vary with place and time" (Fujita and Thisse 2002: 3). Whether the general principles postulated by GE are actually in operation needs to be decided by means of empirical investigation. To realists, theoretical unification of apparently diverse phenomena increases the explanatory power of a theory only so far as the unified phenomena are in reality the result of the same forces (see Mäki 2001a).

GE seems to be willing to expand the explanatory scope of its models in the future, thus conveying the impression that even the omission of context-specific factors could be regarded as an early-step assumption. For now, however, this seems no more than a promissory note and doubts remain on whether this will be done in a way that economic geographers would find satisfactory.

### **3.7 Grounds for complementarity**

GE and economic geography appear to be committed to irreconcilable methodologies. These extreme characterisations, however, look more like caricatures than accurate representations of the two approaches. Many geographers, for instance, though critical of the specific assumptions on which models in mainstream economics are built have raised their voices against banning mathematical modelling from the discipline altogether (for example, Plummer 2003; Plummer and Sheppard 2001; Sheppard 2001); others have been critical of the consequences that the methodological tendencies in economic geography can generate in the extreme (for example, Markusen 1999, Martin and Sunley 2001, Rodríguez-Pose 2001). On the other side of the divide, Krugman's critical attitude towards discursive theorising at times seems too arrogant to be representative of all geographical economists<sup>33</sup>. Once such extreme positions have been tempered, we might be optimistic about the possibility for the two approaches to peacefully co-exist, learn from each other, and perhaps even complement each other, as some commentators have, more or less forcefully,

---

<sup>33</sup> The difficulty for the two approaches to coexist peacefully is often imputed to Krugman's attitude towards methods and modes of theorising other than those expressed in mathematical formal models. Moreover, geographical economists seldom comment on geographers' work. This attitude clearly does not favour a constructive debate.

already acknowledged (cf. Brakman and Garretsen 2003; Dymiski 1996; Martin and Sunley 1996; Sjöberg and Sjöholm 2002).

Whether economic geography and GE are rival or complementary endeavours is still an open question. To reach an answer we would need to single out the assertions of one among the variety of theoretical approaches to issues of agglomeration in economic geography (for example, Malmberg and Maskell 2002; Martin and Sunley 2003; Scott 2000) and compare them with the relevant GE models, something which cannot be done here. Yet I hope that what follows will indicate what kinds of issues require special attention when such comparisons are carried out.

As I have said, GE seeks to show how the (abstract) phenomenon of agglomeration may come about and be sustained through the operation of scale-independent economic mechanisms. These scale-independent economic mechanisms can hardly be thought of as being capable of accounting for all aspects of spatial agglomeration. A complete explanation of all aspects of spatial agglomeration—if possible at all—is likely to call for a wider set of explanantia than that of GE. Geographical economists themselves admit that contingent and locally specific factors of a social, historical, and institutional kind do play a role in explaining agglomeration. At the same time, economic geographers, while emphasising the role of local specifics and historical contingencies, concede that “the explanation of local agglomerations, of spatial economic differentiation necessarily involves close explication of locally specific and contingent factors as well as deeper, more general processes” (Martin *ibid.*: 80).

Table 3.2 reiterates the main differences between GE and economic geography in terms which have framed our discussion so far. It should prove useful as a guide for the

following reflections upon the relation between the two approaches which proceeds by looking at their respective explanantia.

Table 3.2. Geographical economics versus economic geography

	<b>Geographical economics</b>	<b>Economic geography (Martin)</b>
<b>Realisticness</b>	Use of unrealistic assumptions (partial and false) to build mathematical formal models	Comprehensive (closer to satisfying the whole truth) and realistic (satisfying nothing but the truth) discursive descriptions
<b>Realism</b>	Compatible with an abstract version of realism	Grounded on a narrow version of <i>realism</i>
<b>Explanandum</b>	How agglomeration can come about and be sustained	Where agglomeration occurs; why in particular places and not in others
<b>Level of description</b>	High degree of abstraction	Low degree of abstraction
<b>Explanantia</b>	General and abstract economic mechanisms operating at different spatial scales	Locally specific and contingent factors and deeper general processes

The general economic mechanisms GE investigates might be seen as (a subset of) the "deeper, more general processes" Martin refers to (1999: 80). It is clear that by focusing only on this set of factors, the GE models are necessarily partial<sup>34</sup>. What kind of relationship holds between economic geography and GE, that is to say, whether they are rivals or complementary,

---

<sup>34</sup> Whether the theories of economic geographers do include both locally specific and contingent factors and general processes is not investigated here. Notice also that the distinction between specific and general does by



will then depend on the range of the respective set of explanatory factors and on their overlapping.

If (a) economic geography offers a comprehensive theory of spatial agglomeration that includes the whole set of relevant locally specific and contingent factors as well as of deeper general processes and if (b) their claims about deep general processes are incompatible, then (c) economic geography and GE are rivals.<sup>35</sup>

Instead (as it seems more plausible) if (a) neither economic geography nor GE offers such a comprehensive theory and if (b) they focus on different explanatory factors, regardless of whether this difference is in terms of the general versus specific divide, then (c) economic geography and GE are complementary<sup>36</sup>. Yet note that their complementarity is subject to the constraint that the different mechanisms and processes they postulate are compatible with each other, that is, they can be thought to operate in the same reality.

If we accept that economic geography and GE are complementary, then further issues need to be explored: Should one theory supplement the other? Or should an encompassing theory be developed which integrates the two? Or alternatively, should they simply co-exist side-by-side and perhaps be used separately for addressing different sorts of questions?<sup>37</sup>

---

no means imply that investigating the former does not serve to gain insights into the workings of the latter and vice versa.

<sup>35</sup> If their claims on general processes are compatible, then economic geography would replace GE by virtue of the fact that its set of explanatory factors includes the other. The same would hold for GE in the event its set of explanatory factors include the economic geographers' one.

<sup>36</sup> A similar conclusion is reached by Phelps and Ozawa (2003: 600): "if important historical continuities and discontinuities are to be recognised", then the two perspectives (GE being concerned with "invariable principles of agglomeration", economic geography with "historical patterns of restructuring") need to be reconciled.

<sup>37</sup> Mäki (1997) formulates three strategies (vector addition, unification, and division of labour) that parallel the three questions above.

Supplementing one theory with the insights of the other presupposes not only the compatibility of the sets of forces and mechanisms each identifies, but also that their separate effects can be added up. Following John Stuart Mill (1843), we should distinguish between *mechanical* and *chemical composition*. When different mechanisms combine mechanically, it is possible to study each of them in isolation and then add up their separate effects. But when mechanisms combine chemically, their joint effect is different from the sum of each of them separately. An analogous idea is captured by the concept of emergent powers, that is, “powers or liabilities which cannot be reduced to those of their constituent parts” (Sayer 1992: 119). This means that we cannot for instance explain “the power of water to extinguish fire by deriving it from the power of its constituents, for oxygen and hydrogen are highly inflammable” (Ibid.: 119).

If the forces affecting the spatial structure of the economy combine chemically, then GE (economic geography) cannot be simply supplemented by incorporating the insights of economic geography (GE). In such a case supplementation ceases to be a possibility. Integration into a unifying, more fundamental, theory would be a worthwhile pursuit but also overly difficult (if feasible at all). In face of all this, it seems that the best strategy is for the two approaches to co-exist side-by-side and be used for addressing different aspects of the phenomenon of agglomeration.

### **3.8 Conclusions**

The aim of this paper has been to elucidate certain features of GE and by doing so to contribute to the clarification of the

---

Groenewegen and Vromen (1996) distinguish between supplementary theories, complementary theories that apply under different conditions and complementary theories that address different aspects of the same phenomenon.

ongoing dispute between GE and economic geography. Despite the heavy dependence of its models on unrealistic assumptions, GE can be compatible with realism insofar as unrealistic assumptions are subservient to the pursuit of truth about relevant aspects of spatial agglomeration. Once this has been acknowledged, the dispute between GE and economic geography is to be read in terms of what are regarded as relevant features of spatial agglomeration. While geographical economists seek to analyse mechanisms that might be common across a diverse set of instances of agglomeration, economic geographers are also concerned with the effects of contingent and locally specific factors. Yet this is not sufficient to make them rivals once the boundaries of the GE explanandum are correctly identified. This suggests that many of the criticisms of economic geographers are to some extent misplaced.

This conclusion does not intend to score a point in favour of GE. Rather my argument is intended to be neutral. Removing possible misunderstandings in the dispute should favour both parties in that criticisms and responses can become far more incisive. The work of clarification attempted here can benefit geographical economists as well for it may be helpful in precisely identifying the explanatory potential of their models.

Finally, I suggested that without downplaying obvious methodological differences, complementarity might be a possibility provided that the two sets of explanatory factors are compatible. If none of the available theories include all the relevant factors affecting agglomeration, then the selection of a certain subset of those explanatory factors as explanantia will determine what aspect of agglomeration they can possibly be explained.

## **4 The many explanations of spatial clustering: rival or complementary? \***

### **4.1 Introduction**

Spatial clusters are geographical concentrations of firms operating in the same or related industries, potentially connected through production or service related ties. Alfred Marshall (1890) already drew attention to the phenomenon, but with the rise of the large corporate organisation the interest in spatial agglomeration withered in economics and related disciplines. Along with the study of “postfordism”, and the “new” or “knowledge-based” economy, since the 1980s, interest in spatial clustering re-emerged in a range of disciplines such as economic geography (Malmberg and Maskell 2002; Cooke 2002; Storper 1995, 1997), geographical economics (Krugman 1991a; Fujita et al. 1999; Brakman et al. 2001), political economy (Piore and Sabel 1984), public policy (Putnam 1993), and business strategy (Porter 1990, 1994, 1998). Due to their economic potential, clusters and their study have had a remarkable impact on industrial and regional policies designed to promote economic development and the competitiveness of firms, regions and countries (see for example, Cooke 2002).

A multitude of approaches have been proposed apparently to explain different aspects of clusters. Those approaches are

---

\* This essay is co-authored by Päivi Oinas.

concerned with what clusters are (Martin and Sunley 2003); how they emerge (Fujita et alii 1999; Cooke 2002; Saxenian 1994); why they are beneficial for business firms (Porter 1990, 1994, 1998; Storper 1997; Malmberg and Maskell 2002); and why they persist (Fujita et al. 1991; Malmberg and Maskell 2002; Saxenian 1994). Alas, the literature suffers from theoretical ambiguities (Martin and Sunley 2003), and it is far from clear how the various explanations relate to each other. As a result, theoretical disputes have emerged across disciplines (see Chapter 3). Given the amount of intellectual resources devoted to the study of spatial clustering, and the widespread adoption of cluster-based policies, it is high time to try to bring some clarity in this area of research. For this purpose, it is necessary to disentangle the elements of the various explanations and to identify their explanatory scope in precise terms. This is the task we take upon in this paper with the aid of simple tools drawn from the philosophy of science.

As examples, we select the writings of a few core contributors (and some closely related follow-up literature) in four disciplines: geographical economics initiated by Paul Krugman, strategic management research represented by the work of Michael Porter, economic geography represented by Anders Malmberg and Peter Maskell's outline of a knowledge-based theory of clusters, and a cultural approach to clusters as advocated by the regional planner Annalee Saxenian. Even though these are all well-known approaches, even across disciplines, detailed analyses of how they relate to each other have not yet been carried out.

We identify the explanatory structures of these approaches to clustering: *what they explain* (the explanandum) *by what* (the explanans or explanantia) in Section 4.2. Having provided our interpretive reconstructions of the basic elements of the explanations, in Section 4.3 we present theoretical tools that

assist us in determining how the explanations of different theories relate, in particular whether they exclude or complement each other. In Section 4.4, we take a close look at the explanations that seem to explain the same aspect of spatial clustering and offer a comparison by reformulating the explananda in contrastive terms. We suggest that this detailed analysis allows us to discuss the ways in which they relate more precisely than presented in the literature thus far. Section 4.5 discusses and qualifies the results of the analysis, and finally, Section 4.6 concludes and points at further research questions.

## **4.2 The four explanations of spatial clustering: what explains what**

All theories involve incomplete representations of complex reality. Theories narrow down complexity by focusing on selected aspects of phenomena. We can specify the explanatory scope of theories by asking: What is explained by what? Or, what is the explanandum and what is (are) the explanans (explanantia)? The fixing of explananda and explanantia involves theoretical isolations: theories isolate certain features of the phenomenon under study while explicitly or implicitly leaving out others (Mäki 1992b, 1994, 2004b). Isolations can be divided into horizontal and vertical isolations (*ibid.*). A *horizontal isolation* takes place at a given level of abstraction; it amounts to focusing on a limited set of items (for example, when a particular industry is isolated from all other industries). A *vertical isolation* occurs when a universal or quasi universal is isolated from the particularities and specificities of its manifestations (such as isolating the general characteristics that define what it is to be an industry from all other characteristics of specific industries at given times and places). Furthermore,

theoretical isolation is effected by means of omissions (leaving certain items unmentioned) and idealizations (explicitly assuming certain variable(s) to be zero or infinite, such as in assumptions of perfect information or homogenous space).

Theories differ in terms of what they isolate as explananda and explanantia, and in the ways in which these isolations are effected. It often appears that the elements of the explanation and their relations are not clearly specified. Therefore, we have to offer interpretive reconstructions of the explanatory elements in the four approaches to spatial clusters we analyse. (A detailed analysis of the isolations involved in them falls beyond our purposes here, yet examples of those are identified in Section 4.5.)

#### 4.2.1 Porter's strategic management explanation

Michael Porter is a strategic management scholar, known beyond the boundaries of his discipline for his work on factors creating competitive advantage at the level of firms, nations and regions. Since 1990, his work has addressed the significance of clusters for enhancing competitiveness. Porter observes that "most past theories address aspects of clusters or clusters of a particular type" (Porter 1998: 208.). His theory seeks to address all types and aspects of clusters, and hence to provide a unified account (Porter 1998: 208)

Porter understands clusters as "geographic concentrations of interconnected companies, specialized suppliers and service providers, firms in related industries, and associated institutions (for example, universities, standard agencies, and trade associations) in particular fields that compete but also cooperate" (Porter 2000: 253). The widespread presence of clusters, especially in advanced economies, suggests that "much of competitive advantage lies outside a given company

or even outside its industry, residing instead in the *locations* of its business units” (Porter 1998: 198). Porter believes that clusters may be formed at different spatial scales (Porter 2000: 254), and that the same factors can be invoked in explaining them.

Porter integrates his theory of clusters with his broader theory of competition and competitive advantage (1998: 198). He identifies four interrelated elements that affect competitiveness, constituting what he refers to as “diamond”: factor (input) conditions; demand conditions; firm strategy and rivalry; related and supporting industries (Porter 1990). These elements come together in clusters: “Clusters contain one facet of the diamond (related and supporting industries), but they are best seen as a manifestation of the interactions among all four facets” (Porter 2000: 258; cf. 1998: 213).

Porter’s theory seeks to integrate explanations of competitiveness at the levels of firm (1985), nation and region (1990, 1994, 1998, 2000). More specifically, the competitiveness of firms adds up to the competitiveness of more aggregate levels of analysis (cf. Porter 2000: 257). At any scale, “[f]irms located within a cluster are more likely to attain competitive advantage” (ibid.: 257). As to the connecting mechanisms, “[l]ocation affects competitive advantage through this influence on firm *productivity* and especially on *productivity growth*” (ibid.: 256). Our interpretation suggests that clusters *per se* are not the target of his explanation. His focus, instead, continuously lies on the competitive advantage of firms (cf. Porter, 1985), which translates into the competitiveness of higher levels of aggregation. We propose to formulate Porter’s explanandum as referring to competitiveness<sup>38</sup>:

---

<sup>38</sup> We follow Mäki (2004b) in denoting M for explanandum (what is explained) and S for explanans (what explains). Throughout the paper M



**[M<sub>P</sub>]** The competitiveness of clustered firms

His theory isolates as parts of his explanans those *features of clusters that lead to increase productivity and productivity growth* and hence explain why firm's competitiveness is sustained by clustering. The explanans is made up of the following set [S<sup>1-4</sup><sub>P</sub>] (Porter 2000: 259-265):

- [S<sup>1</sup><sub>P</sub>]** *Features of clusters that increase (static) productivity.* Proximity, face-to-face contacts, close and ongoing relationships and "insider" access to information enable superior or lower cost access to specialised inputs, information, institutions and public goods; facilitate complementarities between the activities of the cluster participants (in production, marketing, buying, alignment of activities), improve the incentives within firms and facilitate the measurement of performance of in-house activities.
- [S<sup>2</sup><sub>P</sub>]** *Features of clusters that enhance firms' capacity to innovate and thus productivity growth.* Both rivals and demanding customers located in its proximity prompt firms to innovate. Swiftness in innovation is facilitated by flexibility and the ability to combine different capabilities in clusters.
- [S<sup>3</sup><sub>P</sub>]** *New business formation.* Thanks to easier access to information about opportunities, together with lower barriers to entry and lower perceived risks of entry, clusters stimulate new business formation and attract firms from elsewhere. Increased rivalry strengthens surviving firms in clusters.

---

and S will at times denote phenomena and at others the theoretical representations. In each case the context should clarify the usage.

[S<sup>4</sup><sub>p</sub>] *Favourable social structure.* Shared features of social relations a) facilitate interaction within clusters so that the potential in the other explanantia is realized, and b) by so doing exclude outsiders so as to increase the competitiveness of those included in the cluster (cf. 1990: 129).

Note that “social structure” or “social glue”, that is, various features related to social relations (cf., Porter 2000: 264), has a distinctive explanatory role. As Porter explains, “[t]he mere presence of firms, suppliers and institutions in a location creates the *potential* for economic value, but it does not necessarily ensure the realization of this potential” (ibid.). He seems to suggest that cluster-specific social features are manifested in “relationships, networks, and a sense of common interest” (Porter 2000: 264). He has pointed out earlier that *cultural factors* “work through the determinants [of competitive advantage], not in isolation from them” (Porter 1990: 129); “such influences are important ones to competitive advantage [...] because they change slowly and are difficult for outsiders to tap or emulate” (ibid.). We conclude that features of social relations assist in competitiveness creation in clusters because they make interaction possible, and they serve to include (individuals and firms in a specific cluster) and exclude (competitors outside the cluster). Notice that “social structure” can only influence some of the other explanantia, namely those that work through social relationships, but not those stemming for instance from the available pool of capital (Porter 2000: 226). Obviously, social structure can take on many forms and can also prevent the realisation of the potential in the other explananda. Thus, it should be *favourable* to contribute to enhanced competitiveness.

#### 4.2.2 The explanation of geographical economics

Geographical economics – initially labelled “new economic geography” (Krugman 1991a) – is a theoretical approach to spatial issues developed within economics since the early 1990s. It is firmly grounded in the economics tradition, applying the general equilibrium framework to analyse the location decisions of maximising agents.

Geographical economics (GE) aims to explain agglomeration defined as “the clustering of economic activity, created and sustained by some sort of circular logic” (Fujita et al. 1999: 1). Agglomeration refers to very distinct real-world phenomena including industry clusters, core-periphery patterns, and cities (Fujita and Thisse 2000: 1). GE takes these diverse phenomena to be influenced by the same kind of economic mechanisms, viz. “economic mechanisms yielding agglomeration by relying on the trade-off between various forms of increasing returns and different type of mobility costs” (ibid.: 1).

GE typically assumes homogenous and neutral space in order to isolate, from the influence of spatial differences, the effects of “purely economic mechanisms” on agglomeration. The optimizing agents in the GE models make their location decisions in the presence of *increasing returns at the firm level*, *imperfect competition* (typically employing the Dixit-Stiglitz (1977) model of monopolistic competition) and *transportation costs* (modelled as iceberg transportation costs<sup>39</sup>). At the aggregate level, the interaction between transportation costs and increasing returns give rise to pecuniary externalities (typically market-size effects and labour market pooling). These are counteracted by forces pushing towards dispersion, namely

---

<sup>39</sup> Iceberg transportation costs “imply that a fraction of the manufactured goods does not arrive at the destination when goods are shipped between regions” (Brakman et al. 2001: 80)

centrifugal forces, such as the presence of immobile factors, congestion and the like.

As an example of the kind of mechanisms postulated by the models of GE consider the operation of market-size effects. In the presence of increasing returns and transportation costs, a firm's location affects the local market by increasing local demand for upstream activities (market expansion) and local supply for downstream ones (market crowding) (Ottaviano 2003: 667). The entry decision of a firm gives rise to pecuniary externalities as its decision is based on its own profit. This does not reflect modifications in the payoffs of upstream and downstream activities because of imperfect competition. Agglomeration arises when the former effect (market expansion) predominates over the latter (market crowding) (Ottaviano 2003: 667).

The literature distinguishes pecuniary and technological externalities as the potential explanantia of agglomeration. *Technological externalities* – such as Marshall's (1890) knowledge or information spillovers – impact on the individual firms' production function: if the output of the industry increases, the technological relationship between input and output for the firm is altered. *Pecuniary externalities* resulting, for instance from labour market pooling and market-size effects (Marshall 1890) operate via the market by affecting prices at the level of the firm, possibly but not necessarily leading to an altered output level. GE mainly focuses on pecuniary externalities; this limits what GE can possibly explain about agglomeration. This choice is based on tractability requirements: technological externalities cannot be endogenously derived from a rational-choice *cum* equilibrium framework. Yet, the proponents admit that since these are likely to be more important than pecuniary externalities when the explanation concerns spatial clusters at smaller spatial scales (Krugman 1991b: 54), their account

appears more adequate for larger scales of spatial resolution. In addition, pecuniary externalities are taken to be causes of agglomeration at all times and places, whereas technological externalities are believed to be more important in contemporary advanced economies. This specifies the range of clusters the GE explanation is concerned with: clusters at all scales and at all times.

GE seeks to explain agglomeration at different scales. To them, “all these concentrations form and survive because of some form of agglomeration economies, in which spatial concentration itself creates the favourable economic environment that supports further or continued concentration” (Fujita et al. 1999: 4). The main contribution of GE has been to open the black box of agglomeration economies and hence to derive them from more fundamental considerations (Ibid: 4). In particular, GE investigates two issues. First, the conditions under which “a spatial concentration of economic activity is sustainable”; when the advantages created from agglomeration are sufficient to keep it in place. Second, the conditions under which small differences among locations trigger a cumulative process whose result is the break down of the dispersed equilibrium; when an equilibrium without spatial agglomeration is unstable (Fujita et al. 1999: 9)<sup>40</sup>. Answering both questions amounts to showing how agglomerations can form and be maintained, and hence, to explaining their existence. We formulate the general form of the explanandum of GE as:

[M<sub>GE</sub>]      The existence of clusters at all scales and at all times

---

<sup>40</sup> Although we talk of a “process”, the ad-hoc dynamics in the models falls short of genuinely describing the cumulative process leading from even dispersal to agglomeration. The models only *assume* that workers gradually migrate to locations where real wage rates are higher.

It can be decomposed into two parts.

- [M<sup>1</sup><sub>GE</sub>] The formation of clusters at all scales and at all times
- [M<sup>2</sup><sub>GE</sub>] The sustenance of clusters at all scales and at all times

The general form of the explanans in each case is:

- [S<sub>GE</sub>] The agglomeration mechanisms of pecuniary externalities that is relatively stronger than centrifugal forces.

This abstract explanatory framework is variously modified to generate more specific accounts for the occurrence of agglomerations at different spatial scales.

#### 4.2.3 Malmberg and Maskell's "knowledge-based" explanation

The economic geographers Malmberg and Maskell (2002) point out that despite the considerable attention paid to spatial clustering over the years, "its causes and effects remain elusive" (ibid.: 431); a satisfactory theory of spatial clustering is still missing. These authors observe that existing accounts, both the traditional ones concerned with costs reduction and the more recent ones concerned with knowledge spillovers, can be seen as relying on the reduction of costs of interaction generated by spatial proximity as determinants of spatial clustering (ibid.: 432, 434). Yet empirical evidence does not confirm that the degree of interaction among firms in clusters is higher than

among firms outside, and hence, does not support the idea that clusters exist because they reduce the costs of interaction. To remedy this explanatory failure, they propose that a “satisfactory theory of spatial clustering”, first, “must include an explanation for *the existence of the cluster*”; second, it should “address *the internal organization of the cluster*”; and third, it should be “dynamic in the sense that it can encompass the possibility of and reasons for the decline of formerly successful clusters” (ibid.: 438). We discuss the first and the third of these explananda<sup>41</sup> (see Maskell 2001 for the original formulation; and Bathelt et al. 2004, and Maskell and Lorenzen 2004 for extensions). We begin by discussing their explanation of the *existence* of spatial clusters.

Like many authors (ibid.: 430, fn 1), they believe that “location close to similar or related firms contributes to the competitiveness of an individual firm” (ibid.: 430). In line with recent work – among which they include Porter’s and Saxenian’s contributions – they focus on knowledge creation and learning as the main advantages of clustering in a “knowledge-based” economy. These advantages operate most obviously at the local scale (p. 443). This specifies the kinds of cluster they attempt to explain: local clusters in the knowledge-based economy. In contrast to recent accounts, which typically focus on processes of knowledge creation among vertically-related firms, they identify the “*horizontal dimension*” of clusters, that is, horizontally related (rival) firms, as the most important

---

<sup>41</sup> The second requirement regarding the internal organization of the cluster “implies that it [the theory] should provide a framework for understanding the division of labour taking place between and among firms within the cluster” (ibid.: 438). We found it difficult to specify the elements of this explanation, and will not analyse it in detail. It seems that this explanandum concerns the division of labour between horizontally and vertically related firms in clusters, and the explanantia are in line

locus of knowledge creation (cf. Porter's [S<sup>2</sup><sub>P</sub>] above). They invoke variation, observation, comparison, selection and rivalry as the "knowledge-enhancing mechanisms" (ibid.: 442) that take place in clusters among horizontally related firms. These mechanisms permit "processes of parallel experimentation and testing of a variety of different approaches" (ibid. 439) to related problems in different firms, and lead to the growth of business-specific knowledge.

Then, *the existence of spatial clusters* [M] is accounted for by "knowledge-enhancing mechanisms" along the "horizontal dimension of clusters" [S]. Yet, on a closer examination of what exactly explains what, the story becomes somewhat more complicated. Considerable effort is required to disentangle the various elements of their account. Our reconstruction suggests that two explanations are included. The first is a *functional explanation* of the existence of clusters. We refer to it as explanation [1<sub>M&M</sub>]:

[M<sup>1<sub>M&M</sub></sup>] The existence of local clusters in the knowledge-based economy

[S<sup>1<sub>M&M</sub></sup>] The competitiveness of clustered firms

According to [1<sub>M&M</sub>], clusters exist because they perform a function: the creation or enhancement of firm competitiveness. For Malmberg and Maskell, "an explanation for *the existence of the spatial cluster* [...] means that it must specify the process or processes that impel similar and related firms to cluster at one place and by doing so thrive" (ibid.: 438). Note, however, that they themselves do not specify such a process: they don't explicitly describe "how the cluster emerged as a *consequence* of

---

with the explanation for the existence of the cluster discussed below, plus the growth of the number of horizontally and vertically related firms.



these benefits [of clustering]" (p. 434). Instead, their theory describes the processes that impel similar and related firms to thrive and that attract new firms to a cluster, *once the cluster is already in place*. The second explanation, [2<sub>M&M</sub>], instead, provides a *causal explanation* of the explanans of [1<sub>M&M</sub>]:

- [M<sup>2</sup><sub>M&M</sub>] The competitiveness of clustered firms
- [S<sup>2a</sup><sub>M&M</sub>] Efficiencies in knowledge creation among horizontally related clustered firms due to mechanisms of variation, observation, comparison, selection and rivalry
- [S<sup>2b</sup><sub>M&M</sub>] Efficiencies in knowledge creation among vertically related clustered firms due to mechanisms of specialization, interaction, exchange, coordination and collaboration
- [S<sup>2c</sup><sub>M&M</sub>] Cost reductions
- [S<sup>2d</sup><sub>M&M</sub>] Shared institutional environment

This is a causal explanation describing the mechanisms that enable clustered firms to be competitive. The explanantia appear to have different degrees of importance in explaining competitiveness. In Malmberg and Maskell's words (*ibid.*: 440), "[t]he cluster exists [...] primarily because of the benefits of enhanced knowledge creation that occur when many colocated firms undertake similar activities", that is [S<sup>2a</sup><sub>M&M</sub>].

They then move on to discuss "the important, but secondary, types of knowledge creation stemming from the vertical dimension of the cluster, once its existence has been secured" (*ibid.*: 440): efficiencies in knowledge creation among vertically related firms, [S<sup>2b</sup><sub>M&M</sub>]. The advantages from knowledge creation along the vertical dimension begin to play a role only during later stages of cluster development, either because the

cluster attracts specialised suppliers and sophisticated customers or through task partitioning taking place among specialised firms in the cluster (ibid.: 440).<sup>42</sup>

Furthermore, that “a theory of spatial clustering should play down the role of cost efficiencies” (ibid.: 444), that is  $[S^{2c}_{M\&M}]$ , is here taken to mean that cost efficiencies are not decisive, but they should still be included among the explanantia of  $[M^2_{M\&M}]$ .  $[S^{2c}_{M\&M}]$  includes the kinds of items discussed in traditional theories of clusters: costs-savings due to (a) collective resources; (b) well-functioning markets for specialized skills and (c) co-location with partners (ibid.: 432-433). Largely in line with Porter’s analysis, the main thrust in Malmberg and Maskell’s theorising lies in the explanation of why clustering makes firms competitive.

Malmberg and Maskell claim that ‘institutional fit’, that is, the emergence of local institutions meeting the needs of the types of firms constituting the cluster, is not in itself part of an explanation of the existence of the cluster (Ibid.: 442). However, their discussion in fact seems to suggest that institutional factors contribute to an explanation of the competitiveness of clustered firms (and thereby to an explanation of the existence of clustering). In discussing “the internal organization of the cluster”, they suggest that bridging cognitive distances in transferring knowledge is “less difficult when the transfer takes place within a community that shares the same language, beliefs, judgements and values” (p. 440). This additional explanans comes very close to Porter’s fourth explanans (social

---

<sup>42</sup> Although is not fully clear whether in the later stages the advantages of knowledge-creation along the vertical dimension still remain secondary vis-à-vis those along the horizontal dimension, we reconstruct them as such. The discussion of early and later stages seems to belong to what Malmberg and Maskell themselves see as their explanation of the internal organisation of the cluster (see footnote 39 above). We fail to reconstruct this explanation as separate from the other explanations.

structure, [S<sup>4</sup><sub>P</sub>]): institutional factors influence the way in which the other factors operate by moulding the institutional (socio-cultural) bases and patterns of interaction and thereby differentiating clusters from each other. Similarly as in Porter's [S<sup>4</sup><sub>P</sub>], insiders are favoured vis-à-vis outsiders who do not share the same institutional environment. This makes specialized suppliers and sophisticated customers (p. 440) and entrepreneurs (p. 441) attracted to clusters. We therefore formulate a *shared institutional environment* as an additional explanans to Malmberg and Maskell's explanation of competitiveness, [2<sub>M&M</sub>].

Their third explanation concerns "the reasons for the decline of formerly successful clusters" (ibid.: 438). A cluster's persistence over time involves the development of an institutional fit. The institutional fit is said to contribute to the "explanation of the successful path-dependent development trajectories of clusters - and to the lock-in situations in which clusters sometimes end up" (ibid.: 442). The locality-specificity of institutions contributes to *how* learning takes place in a cluster and thus distinguishes learning processes in one locality from those in other localities. The effect is that it shortens cognitive distances among agents in a cluster, and makes presence in a cluster attractive for actors pursuing success in an innovative business (ibid.: 441).

This third explanation has a temporal dimension to it: A *rigid* "institutional fit", (preventing firms from renewing) leads to reduced competitiveness through time, which accounts for the decline of clusters. We take it that the opposite also applies: a *flexible* "institutional fit" sustains firm competitiveness in time, and hence, accounts for the continued success of clusters. For later comparison, we formulate [3<sub>M&M</sub>] as an explanation given at the firm level rather than at the level of the cluster and as an explanation of the success of clusters instead of their decline.

This is because continued competitiveness of firms in clusters seems to add up to the successful continuation of clusters in Malmberg and Maskell's theorising. Hence, we get the following explanation:

[M<sup>3</sup><sub>M&M</sub>] Continued competitiveness of firms in clusters

[S<sup>3</sup><sub>M&M</sub>] Flexible institutional fit

In order to distinguish it from the previous explanation, we add "continued" as an attribute meant to capture the temporal dimension characterising this explanation.

#### 4.2.4 Saxenian's cultural explanation

Saxenian's (1994) well-known study compares two leading high-tech regions of the United States (up till the early 1990s): the surroundings of Route 128 on the East coast and Silicon Valley on the West coast. Some have regarded Saxenian's book as "path-breaking" in regional science (Waldorf 2004: 62), and there is a continuous flow of references to her study across disciplines. It seems that her powerful narrative has convinced broad audiences about the fact that it is something about *regional culture* that made Silicon Valley firms succeed and that made Route 128 firms do less well.

The two regions stood out as symbols of economic and technological success in the U.S. (Florida and Kenney 1990: 98). By the 1970s, they were also regarded internationally as the world's leading centres of innovation in electronics (Saxenian 1994: 1). In the 1980s, these regions went through severe downturns. By the early 1990s, Silicon Valley was able to regain its former vitality, and Route 128 was still in trouble. "Why?" is the question the book seems to set out to answer. In Saxenian's

analysis, firms are embedded in regional environments that provide the basis for their success, or lack of it, in competition. Much of the answer lies in characterizations of Route 128 firms as “autarkic” and Silicon Valley firms as members of a “community”.

Saxenian’s book may be effectively characterised as an ethnographic narrative, a thick description of the production and innovation environments in the two regions. The beginning of the book provides a light conceptual apparatus that loosely guides the detailed narrative of the key moments and industrial practices that gave rise to considerable differences in regional economic success. While Saxenian discussed those practices only in terms of “regional culture”, her narrative incorporates aspects of regional, organisational and industrial cultures (Oinas 1995, 1998). Saxenian shows how the characteristics of the regional cultures that had evolved in the two regions in the course of history (strong puritanist culture in the East Coast vs. the “Wild West”) had given rise to very different organisational cultures (closedness, rigidity and hierarchy in Route 128 vs. openness, flexibility and informality in Silicon Valley) and *regional industrial cultures* (co-located autarkic corporations in Route 128 vs. collectively learning community in Silicon Valley). Details in Saxenian’s conceptualisation and narrative suggest that there were feedback loops among the various cultural features: they seemed to co-evolve. Silicon Valley’s cultural features induced adaptability which went hand-in-hand with entrepreneurship and innovativeness, and resulted in continued competitive success of an increasing number of Silicon Valley firms. The “localized accumulation of technical knowledge enhanced the viability of Silicon Valley start-ups and reinforced a shared technical culture” (Saxenian 1994: 37). The fate of Route 128 was to suffer from rigidity and lack of innovation, and decreasing competitiveness among its large

autarkic firms. From this reconstruction we extract Saxenian's explananda, which we formulate at the firm level rather than at the cluster level.

[M<sup>SV</sup><sub>Sax</sub>] The continued competitiveness of firms in Silicon Valley

[M<sup>R128</sup><sub>Sax</sub>] The declining competitiveness of firms in Route 128

These in turn are explained (albeit via a more complex chain of causal relations than the one sketched below) by:

[S<sup>SV</sup><sub>Sax</sub>] Cultural features of Silicon Valley (openness, flexibility and informality; collectively learning community)

[S<sup>R128</sup><sub>Sax</sub>] Cultural features of Route 128 (puritanism, closedness, rigidity, hierarchy)

From these specific explanations of concrete cases, we can extract a more general one that can be presented schematically as a causal chain: (regional, regional industrial and organisational) *cultural features* are reflected in *adaptability* which leads to *innovativeness* which leads to *competitiveness*. From our reconstruction of Saxenian's concrete explanation we can derive a generalized explanation (the successful case), where the explanatory elements are:

[M<sup>G</sup><sub>Sax</sub>] The continued competitiveness of firms in certain clusters

[S<sup>G</sup><sub>Sax</sub>] Those cultural features that facilitate adaptability

We can see that Saxenian's explanation closely resembles Malmberg and Maskell's third explanation as formulated above.

### 4.3 Inter-theoretic relations

#### 4.3.1 Rival and complementary explanations

Identifying the isolations involved in specific theories and the resulting configuration of explananda and explanantia provides us with the basic elements to carry out inter-theoretic comparisons. Such comparisons seek to answer the question of whether different theoretical explanations of a phenomenon are complementary or rival to each other. Two explanations are rival when their claim about either the explanantia or the explananda are incompatible, that is, they cannot be true at the same time (cf. Day 2004). For instance, if one were to propose that the only reason for firms to cluster is the attainment of cost reductions, then any explanation citing some other advantage as exclusively responsible for firm clustering would be rival in relation to the first explanation because their explanantia are incompatible. When explanations are rival, at least one of them is mistaken. Partial rivalry is also possible, for example, when two explanations take the same factors as explanantia but ascribe different explanatory relevance to them.

[Table 4.1 around here]

Table 4.1 schematically identifies the possible ways in which theories differ in terms of what they isolate as explanantia and explananda (the first and second columns) and lists potential sources of rivalry and complementarity (the literature on

explanatory exclusion has given inspiration here; Day 2004; Kim 1988, 1989). It will turn out that our comparisons concern two of the cases

In Case I, the same explanandum is accounted for by different explanantia,  $S_1$  and  $S_2$ . In this case the two explanations are rival if  $S_1$  and  $S_2$  explanantia are incompatible. If they are not,  $S_1$  and  $S_2$  can either be both necessary causes, be causally related, operate under different conditions, or be related otherwise. In Case II, different explananda of two explanations,  $M_1$  and  $M_2$ , are explained by the same explanans,  $S$ . Here rivalry arises from the incompatibility of  $M_1$  and  $M_2$ . The explanations can complement in the following cases:  $S$  explains both  $M_1$  and  $M_2$ ;  $S$  operates under different conditions to bring about either  $M_1$  or  $M_2$ ;  $M_1$  and  $M_2$  are causally or otherwise related. Finally, in Case III, both the explanantia and explananda of two explanations differ. This latter case is relevant only when the  $M$ 's and  $S$ 's differ but relate to the same phenomenon. Here, the explanations are rival in case of incompatibility either between the  $S$  of one and  $M$  of the other, between the  $S$ 's or between  $M$ 's. They are complementary when the  $S$ 's and the  $M$ 's are somehow related to each other.

#### 4.3.2 Explanation and contrastives

In order to determine whether explanations are rival or complementary, we employ what philosophers of science call the "contrastive approach to explanation" (Barnes 1994; Day 2004; Garfinkel 1981; Lipton 1990, 1991; Morton 1990; Woodward 2003; Ylikoski 2001). The crux of the contrastive approach is that explanations do not explain plain phenomena, but those phenomena in contrast to alternatives. This insight makes it possible to formulate the explananda in precise terms – in terms that enable us to see why different explanations invoke



the explanantia that they do. A fictional example illustrates the power of the contrastive approach. A detective investigating a murder case asks a witness: “Why did the victim die?” The witness replies: “We all have to die sometime”. Obviously this is not the answer the detective expected to receive. While the detective had asked, why the victim died *rather than lived on as expected*, the witness answered the question why he died *rather than lived forever*? The detective and the suspect had contrasted the victim’s death with different alternatives (cf. Garfinkel 1981: 22) and this explains the miscommunication between them.

Explanation-seeking questions are taken to have a contrastive form, whether explicit or not, and whether they concern singular events (as typically discussed in the philosophical literature), like the victim’s death, or theoretical explanations. The contrastive explanandum is usually expressed as a why-question of the form: “Why X *rather than* Y?” Our notation below is of the form: “[CM<sub>Author</sub>] X [Y]” (cf. Ylikoski, 2001). For example, “[CM<sub>GE</sub>] The existence of spatial agglomeration [*spatial dispersion*]” means that the GE explains why spatial agglomeration rather than spatial dispersion occurs. Given that phenomena or aspects thereof are ultimately brought about by a vast number of *causally relevant* factors, the effect of the contrastive formulation is that it allows for picking out the *explanatorily relevant* factors among them. The contrast narrows down the number of explanantia compared to the number of explanantia for a non-contrastive explanandum. As illustrated by the fictional example above, a different contrast usually requires a different explanans. Notice that the contrastive approach to explanation is neither about what the person who asked for an explanation had in mind nor what the one who is giving the explanation means. It is instead about what a given explanation can actually explain (cf. Ylikoski 2001:

19), that is, what the explanantia of a given explanation can actually account for about a given phenomenon.

The contrastive approach to explanation is a useful tool in comparing apparently competing explanations of the same phenomenon. It allows realising whether the explanations are really explaining the same thing and if not, to spell out the way in which they differ (cf. Garfinkel 1981; Ylikoski 2001; Day 2004). In the following section, we employ the contrastive approach to compare explanations that our previous analysis revealed to deal with the same phenomenon, that is, the competitiveness of clustered firms and the existence of spatial clusters. It appears that once we formulate the explanations in contrastive terms, their relations correspond to Cases I and III of Table 4.1 (cf. Table 4.2 in Section 4.5).

#### **4.4 The four explanations of spatial clustering: how they relate**

##### 4.4.1 Explanations of competitiveness

We reconstructed Malmberg and Maskell's and Porter's accounts as explanations of the enhanced competitiveness of clustered firms, and we reconstructed Malmberg and Maskell's and Saxenian's accounts as explanations of the continued competitiveness of firms in clusters. Let us first discuss the relations of the former two by employing the insights of the contrastive approach.

To begin with, the two accounts differ in terms of the scale of clusters which they deal with. Whereas Malmberg and Maskell's explanation is geared to the local scale, Porter's

explanation supposedly concerns all spatial scales<sup>43</sup>. Both explanations, therefore, explain the local scale.

When we try to formulate Malmberg and Maskell's explanandum in contrastive terms, it is not self-evident what "competitiveness of clustered firms" is contrasted with. A closer look at Malmberg and Maskell's different explanantia identified in Section 4.2.3, suggests us how to formulate their contrastive explananda.

As we have seen, Malmberg and Maskell take competitiveness to be primarily explained by the advantages that arise from horizontal relations among rivals. These are advantages that accrue to many related firms "when co-located but which are not available to a hypothetical single firm carrying out precisely the same activities, even at the same location, using the same suppliers, customers and workforce" (ibid.: 438)<sup>44</sup>. That is to say, when advantages of clustering are contrasted to the potential advantages of a single uni-locational firm, the explanans is  $[S^{2a}_{M\&M}]$  (and, we presume,  $[S^{2d}_{M\&M}]$ ).

This gives us the following contrastive explanandum:

---

<sup>43</sup> The benefits Porter identifies as explanantia mainly arise from spatial proximity even though his explanation is concerned with clusters at all scales. We can think of two ways of remedying this inconsistency. One option is to suggest that a) clusters at all scales are accounted for by factors that draw on his "static productivity" explanantia (especially cost reductions in clusters), and b) lower scale clusters are accounted for by "dynamic efficiency" explanantia. The second option is that the explanantia Porter identifies for clustering at any scale work most efficiently at the local scale but many aspects of them, even if not all, can be seen as operating at higher scales as well (for example, frequent face-to-face interaction is less likely over long distances, yet this can be compensated by telephone and e-mail interactions).

<sup>44</sup> Note a small slip in the argument. If the point is precisely that more activities can be carried out and more knowledge can be created in multiple firms in clusters through mechanisms of variation, observation, comparison, selection and rivalry than in single firms, "carrying out precisely the same activities" in clustered firms and single unilocational firms would not be possible.

[CM<sup>2A</sup><sub>M&M</sub>] The competitiveness of clustered firms [*of single unilocational firms*]

[S<sup>2a</sup><sub>M&M</sub>] Efficiencies in knowledge creation among horizontally related firms

[S<sup>2d</sup><sub>M&M</sub>] Shared institutional environment

Moreover, Malmberg and Maskell identify *dispersal* – or isolated locations – as the contrast in “[t]he more theoretically oriented part of the literature of agglomeration [... that] aims at explaining the existence of spatial clustering by identifying and analyzing those permanent advantages that may accrue to firms located close to other similar and related firms, rather than being located in isolation” (ibid.: 432). The explanantia include cost efficiencies as covered in the earlier theoretical literature (ibid.: 432), efficiencies in knowledge creation among vertically related firms, and the facilitating effect of the shared institutional environment. Since existence is explained by competitiveness, and this is explained by the advantages from clustering, we formulate the contrastive explanandum as follows.

[CM<sup>2B</sup><sub>M&M</sub>] The competitiveness of clustered firms [*of dispersed firms*]

[S<sup>2b</sup><sub>M&M</sub>] Efficiencies in knowledge creation among vertically related firms

[S<sup>2c</sup><sub>M&M</sub>] Cost reductions

[S<sup>2d</sup><sub>M&M</sub>] Shared institutional environment

Similarly, Porter asks, what is it in clustering that confers competitiveness to firms, but is not available to firms otherwise

organized? Porter recognises clustering as an alternative to a range of other arrangements including vertical integration, formal alliances with outside entities, sourcing inputs from distant locations, outsourcing, dispersal, isolated firms or locations (Porter 1998: 214-225, cf. Porter 2000). We formulate his contrastive explanandum as:

- [CM<sub>P</sub>]      The competitiveness of clustered firms [*firms otherwise organized*]
  
- [S<sup>1</sup><sub>P</sub>]      Features of clusters that lead to increase (static) productivity
- [S<sup>2</sup><sub>P</sub>]      Features of clusters that enhance firms' capacity to innovate and thus lead to productivity growth
- [S<sup>3</sup><sub>P</sub>]      New business formation
- [S<sup>4</sup><sub>P</sub>]      Favourable social structure

Malmberg and Maskell's and Porter's explanations are very similar in substantive terms. In both, the explanantia include a range of advantages that account for the superiority of clusters vis-à-vis other forms of spatial organization. Even the equivalent of Porter's [S<sup>3</sup><sub>P</sub>] can be found in Malmberg and Maskell's description of the later stages of cluster development when firms are attracted to join the cluster from elsewhere and the cluster processes create opportunities for the emergence of new firms (Malmberg and Maskell 2002: 440). The main difference lies in Malmberg and Maskell's emphasis on knowledge creation among horizontally related firms. Porter does not ascribe explicit priority to the advantages giving rise to dynamic productivity over those giving rise to static productivity and, within the former, to those stemming from the horizontal dimension. Thus, the two explanations differ in

the claims made about the explanatory relevance of the various explanantia. This makes them partially rival.

Summing up, we conclude that Malmberg and Maskell's and Porter's explanations of competitiveness are related as follows:

- (i) Both can be applied to explaining firm competitiveness in local clusters in the knowledge-based economy.
- (ii) They explain similar contrastive explananda with very similar explanantia, which makes them closely related.
- (iii) However, they seem to be partially rival: in Porter we could not distinguish different explanantia specifically explaining different contrastive explananda. By contrast, Malmberg and Maskell explicitly regard the knowledge creation advantages along the horizontal dimension (or advantages from rivalry) as explanatorily primary.

Consider now the explanations of the continued competitiveness of firms in clusters. The contrast of Saxenian's explanation is easy to identify as it can be derived from her narrative on the continued competitiveness of Silicon Valley's firms vis-à-vis the *declining competitiveness* of Route 128's firms. The contrastive formulation of Malmberg and Maskell's third explanation [ $CM_{M\&M}^3$ ] is virtually the same.

[ $CM_{Sax}^G$ ] = [ $CM_{M\&M}^3$ ] Continued [*declining*] competitiveness of firms in certain clusters

The explanantia are respectively:

[ $S_{Sax}^G$ ] Cultural features facilitating change

[ $S_{M\&M}^3$ ] Flexible institutional fit

Saxenian's narrative depicts the cultural features of clusters that support adaptability and innovativeness and that lead to competitiveness ( $[SC_{Sax}]$ ). Similarly, Malmberg and Maskell seem to suggest that cluster-specific institutions "such as social capital" (p. 441) have to remain adaptive to support processes of continued learning and innovation. Thus, their explanantia are also closely related. Both explanations refer to institutions, but Saxenian's narrative especially highlights the broader cultural foundations of the institutionalised business practices (and Malmberg and Maskell, 2002: 440, also point at them in passing). While for Saxenian culture facilitates adaptability and collaboration (even among competitors, although her narrative concerns the earlier stages of Silicon Valley development), Malmberg and Maskell's emphasis is on adaptability and rivalry.

We draw the following conclusions, suggesting that the two explanations are complementary.

(i) The two explanations seem to agree that what discriminates between successful and unsuccessful (declining) clusters in time is the presence of suitable (flexible) institutions that enable processes of adaptation and innovation and hence account for the continued competitiveness of firms.

(ii) Saxenian puts considerably more emphasis than Malmberg and Maskell on the cultural foundations of the cluster-specific institutions.

#### 4.4.2 Explanations of existence

Geographical economists and Malmberg and Maskell seek to explain the existence of clusters. GE explains the existence of all kinds of clusters at all times, whereas Malmberg and Maskell focus on local clusters in a "knowledge-based economy". The

latter is a subset of the former and hence both explanations apply to local clusters in a knowledge-based economy which is where they can be compared more easily. The formulation of their explananda in contrastive terms reveals a significant difference, however.

The geographical economists' models show that economic activity becomes either (fully or partially) agglomerated or dispersed, depending on the relative strength of the forces pushing towards agglomeration (for example, market expansion effect) vis-à-vis those pushing towards dispersion (for example, market crowding effect). From this, we derive their contrastive explanandum:

**[CM<sub>GE</sub>]** The existence of clusters [*of spatial dispersion*] at all scales and at all times

**[S<sub>GE</sub>]** The agglomerating mechanism of pecuniary externalities that is relatively stronger than centrifugal forces

Malmberg and Maskell explain the existence of clusters functionally by invoking their role in creating competitiveness. It is not evident what existence is contrasted with, but, from Section 4.4.1 above, we can infer that the alternatives included in the contrast can be at least *organisation of economic activities as a single uni-locational firm and dispersal of (horizontally or vertically related) firms in several locations*. Thus, we take the contrast to be the *non-existence of clusters*, that is, all other states of affairs in which economic activities are organised in space in some other way. The contrastive formulation is:

**[CM<sup>1</sup><sub>M&M</sub>]** The existence of local clusters [*of other forms of spatial organisation*] in the knowledge-based economy



[S<sup>1A</sup><sub>M&M</sub>] The competitiveness of clustered firms

The contrast reveals a key difference between the two explanations. The rationale of Malmberg and Maskell's explanation is that the benefits of clustering *vis-à-vis all other forms of spatial organisation* explain why clusters exist. Those benefits are given in explanations of [CM<sup>2A</sup><sub>M&M</sub>] and [CM<sup>2B</sup><sub>M&M</sub>] regarding how firm competitiveness is created in clusters (*vis-à-vis* single uni-locational firms and dispersal), and in [CM<sup>3</sup><sub>M&M</sub>] regarding how firm competitiveness continues to be sustained (*vis-à-vis* declining) in institutionally flexible clusters. These seem to be the reasons why firms keep clustering, and hence why clusters keep existing. The argument is, essentially, that no other form creates firm competitiveness quite like clustering, and that is why clusters exist *rather than not exist at all*.

GE only compares two possible scenarios: agglomeration and *dispersion*. Under certain ranges of parameters, benefits of co-location offset benefits of dispersion. This explains how clusters may come about and be sustained. When GE models compare clustering to the *dispersal of firms and workers over two or more locations*, 'existence' refers only to the formation and sustenance of clusters under the specific conditions specified in the models. Thus, they explain why clustering rather than dispersion takes place and not why clusters exist rather than not existing at all – this is because clustering and dispersal are not the only two organizational options available to firms.

Summing up, the explanations relate as follows

- (i) The contrastive formulation reveals that the theories provide two different explanations and are complementary.

- (ii) The contrastive formulation unravels two different meanings of 'the existence of clusters'.

#### 4.5 The intricate web of explanatory relations

We have seen that explanations differ in what they isolate as explananda and explanantia, and in how they isolate them (cf. Section 4.2 above). For instance, to explain merely the ever-present phenomenon of clustering – not, for example, the regional agglomeration of innovation (cf. Caniels and Romijn 2003) – GE abstracts (isolates vertically) clusters from place-specific and scale-related particularities. To isolate the “purely economic mechanisms” supposed to always explain a tendency of certain types of economic activity to cluster, it employs a host of horizontal isolations (for example, the neglect of technological spillovers) as well as vertical isolations (e.g., abstract agents), effected through strong idealizations (e.g., maximizing agents) and various omissions. Their explanation can be seen as complementary to those of other scholars who have set out to explain different aspects of clusters or specific types of clusters. In Porter’s explanation, the degree of horizontal isolation regarding the explanantia is lower than in GE as Porter includes a wider range of factors. Porter focuses on what makes clustered firms competitive. With the same aim, Malmberg and Maskell include similar factors as Porter. By focusing on *local* clusters they horizontally isolate them from clusters at other scales. Two of the explanations agree that various socio-cultural and institutional features account for the capability of clusters to remain successful in time through their firms’ continued competitiveness. Saxenian’s rich historical narrative requires fewer isolations and idealisations. When we reconstructed it as a more generic explanation to compare it

with Malmberg and Maskell's explanation, we vertically isolated the cultural characteristics from the specific features they take on in concrete clusters.

The implicit isolations determine those aspects of the clustering phenomenon that are being explained and how broad a variety of explanatory features are drawn upon. Whether some of these isolations are more adequate than others cannot be decided a priori. The correctness of a given explanation, and hence of its isolations should be ultimately settled by empirical investigation. Difficulties in identifying the explanatory strategies in the theories we have analysed suggest that some explanations do this more clearly than others. It may be more customary in some traditions to be fuzzy as to the precise explanatory scope (cf. Markusen 1999). Our analysis suggests that clarifications regarding one's explanatory goals will be helpful in assessing and comparing explanations presented in the literature.

The explananda in the four approaches we analysed seemed to fall into three categories: a) the existence of clusters and b) the competitiveness of clustered firms, and c) their continued competitiveness. In Table 4.2 we summarise the results of our comparisons. The classification corresponds to the cases of Table 4.1 that are of relevance here, that is, Cases I and III. The contrastive formulation of the explananda revealed that even if the explanations explain the same aspect of the clustering phenomenon, the precise explananda were different. Thus, most of the comparisons ended up representing Case III in Table 1. Comparisons of the reconstructed explanations suggest that they are complementary (except for Porter's and Malmberg and Maskell's case of partial rivalry).

[Table 4.2 around here]

The three explananda represent different aspects of clustering and are obviously related. Although the contrastive analysis has been of assistance in unearthing the aspects of clustering that each account can possibly explain, it does not provide the means to understand how those different aspects relate to each other. Identifying their intricate explanatory relations remains a challenge for theorists. We propose that our analysis prompts the following considerations, however.

Both the existence of clusters and the competitiveness of clustered firms are explained by invoking the benefits accrued to firms by clustering, although the types of benefits differ. For geographical economists, the benefits come in the guise of pecuniary externalities, and those are sufficient to bring about clusters and keep them in place *vis-à-vis dispersal*. For Porter, and for Malmberg and Maskell, the benefits from clustering do include something akin to pecuniary externalities (cost savings), but the focal benefits, in their view, relate to features of clusters that increase innovativeness, and thus enhance the competitiveness of clustered firms.

The comparison of Porter's and Malmberg and Maskell's theorising shows that formulating a precise contrastive explanandum helps in disentangling the role of various explanatory elements in the theory. They invoke very similar explanantia for the competitiveness of clustered firms. But while Porter's theory seems to offer only a generally formulated explanandum with a list of undifferentiated explanantia, Malmberg and Maskell's theory could be dissected into more detailed explanations concerning the competitiveness of firms in cluster *vis-à-vis their activities being organised as a uni-locational single firm* and *vis-à-vis being dispersed*. Additionally, Malmberg and Maskell are explicit about how the explanations of firm competitiveness and of existence of clusters relate: competitiveness and its sustenance functionally explain

clusters' existence. Yet, it remains unclear whether firms knowingly cluster in view of obtaining those benefits and hence competitiveness, or whether it is the market mechanism that ensures that clustered firms survive in competition – or both.

Furthermore, the benefits from clustering can only arise when a certain number of firms are already co-located. The realisation of these advantages then attracts more firms in the locality and thereby further reinforces these advantages. This brings us to definitional issues about the very concept of cluster: at what point does a spatial concentration of firms turn into a cluster?

#### **4.6 Conclusions**

Many real-world phenomena are multifaceted and complex, and theories and models, by nature, can only provide partial representations of them. Hence theories typically isolate some aspects of the phenomenon they study, while neglecting others. Vagueness about what is being explained and by what leads to misguided judgements of the explanatory value of theories. In practice, very often theories and the explanations they provide are expected to explain more than what they actually can (cf. Ylikoski 2001), and critics often take this failure to imply that the theories in question are false. In some cases vagueness of what is being explained leads to considering alternative theories as rivals when in fact they are potentially complementary. (Misunderstandings of this sort seem to have afflicted the dispute between economic geography and GE, cf. Chapter 3 and 5). The contrastive approach helps in making sense of the different theoretical isolations which we find in explanations that may, at first, appear to be rival. Ours is then a call for clarity and precision on what is being explained by what and on its relations to alternative approaches. Clarity about

explanatory strategies allows avoiding misguided judgements of rivalry between explanations, eases cross-fertilization and communication, and enhances the possibility of attaining explanatory progress.

Recognition of these features should also ground attempts at comprehending the plurality of theories of spatial clustering. Spatial clustering is a complex phenomenon and a variety of explanation-seeking questions can be asked about it. The selection of theories and explanations examined here appear to focus on closely related aspects of clusters, namely the existence of clusters due to various benefits they bestow and the competitiveness of firms as an outcome of clustering.

The different theories in spatial clustering research are built according to disciplinary conventions and for specific purposes. This influences what aspects of the phenomenon each theory isolates as in need of explanation and as doing the explaining. The theories analysed here come from different disciplinary traditions and, accordingly, offer different types of explanation, construe those according to different methodologies and with different purposes in view. We have largely neglected these differences. Abstracting from these features means neglecting other forms of rivalry. Yet this abstraction serves the purpose of showing that despite differences in style, methodologies and foci, the theoretical approaches are not explanatorily rival.

## 5 Geographical economics models: contrastive explanation in action

### 5.1 Introduction

Disputes over the unrealisticness of models and their assumptions abound in economics. They typically centre on the question of whether a given model is realistic enough so as to explain the real-world phenomena it is used to investigate. These disputes are often carried out without paying enough attention to the various roles that the different unrealistic elements of the model play. In the methodological literature, three main categories of assumptions have been identified: negligibility, applicability and tractability assumptions (Musgrave 1981; Mäki 2000; Hindriks forth). Here, I spell out a distinct role for unrealistic elements of theoretical models which cannot be reduced to any of the above categories. Their role is derived from the application of the contrastive approach to explanation (Garfinkel 1981; Lipton 1990; 1991; Woodward 2003; Ylikoski 2001).

The contrastive approach takes an explanation to be, either explicitly or implicitly, an answer to a composite question of the following form: why  $f$  rather than  $c$ ? Here  $f$  is the aspect of the phenomenon to be explained (for instance, leaves turn yellow in autumn) and  $c$  is an alternative to it (for instance, leaves turn blue in autumn or turn yellow in spring).

Formulating an explanation-seeking question in contrastive terms allows us to make sense of a given selection of factors among the many that lie behind the occurrence of a phenomenon: the explanatory factors to be selected are those which discriminate between  $f$  and  $c$ . This implies that those causal elements common to the causal nexuses of  $f$  and of  $c$  will not be part of an explanation of why  $f$  rather than  $c$ . These common elements constitute the shared causal background. I suggest that at least some of the unrealistic elements of economic models serve the function of fixing the causal background. These unrealistic elements have an ontological justification and, when correctly selected, do not endanger the truth of the explanation of why  $f$  rather than  $c$ .

I will show that looking at the explanations offered by economics models in contrastive terms helps to be precise about the explanatory potential of those models, and hence, to assess the adequacy of their unrealistic elements. It is by this means that contrastive analysis enhances more sophisticated disputes over the unrealisticness issue.

I will use the geographical economics (GE henceforth) models as an illustration. GE is a recent approach aimed at introducing the long-neglected role of space into (conventional) economics (for example, Baldwin et al. 2003; Brakman et al. 2001; Fujita and Thisse 2002; Fujita et al. 1999; Krugman 1991a,b). It employs a general equilibrium framework in which optimizing agents take their location decisions in presence of increasing returns, imperfect competition and transportation costs. Criticisms from neighbouring fields, in particular from economic geography, have been mainly directed at the unrealisticness of the assumptions of the GE models. Economic geographers who have traditionally been interested in agglomeration and spatial phenomena have adopted a



particularly critical attitude towards GE. The ongoing and intense dispute between geographical economists and economic geographers over the worth of these models provides a fertile terrain to illustrate the usefulness of the contrastive approach. A contrastive analysis of the explanandum of GE shows that many of the unrealistic elements of the models serve to fix the causal background, and by so doing, that most of the charges of unrealisticness are misguided. The analysis also makes the explanatory ambitions of the GE models more explicit and suggests that, as it stands, GE can at most explain a fairly narrow aspect of real-world spatial phenomena.

The paper is comprised of three sections. Section 5.2 presents the contrastive approach and what I take to be its contribution to the unrealisticness issue. Section 5.3 applies the contrastive approach to the case of GE. It discusses a few theoretical isolations of the GE models and the critiques directed at them. The contrastive approach is shown to clarify the explanatory potential of the GE models, and accordingly, in what ways the criticisms of unrealisticness are misguided. Section 5.4 concludes the paper.

## 5.2 Contrastivity and unrealisticness

The common way to illustrate the idea behind contrastive explanation is by telling the story of Willie Sutton, the bank robber. When asked by the priest of the prison why he robbed a bank, Willie candidly replies: "Because the bank is where the money is". Obviously this is not the answer the priest expected to receive. While the priest was asking why Willie had robbed a bank, *rather than living honestly*, Willy answered the question, why did you rob bank *rather than robbing any other place*? The miscommunication between Willie and the priest occurred

because their implicit contrasts were different. The main tenet of the contrastive approach to explanation is that explanation-seeking questions, scientific or not, have a contrastive form, whether explicit or not (Garfinkel 1981). Notice that the contrastive approach to explanation is neither about what the explainee has in mind when asking for an explanation nor about what the explainer means when giving the explanation, but instead it is about what a given explanation can actually explain (cf. Ylikoski 2001: 19).

In general, to articulate the explanandum in the contrastive form allows picking out the appropriate explanatory factors from the causal history of the phenomenon<sup>45</sup> for which we search an explanation. The point is that the causal history or causal nexus of a phenomenon is typically incredibly complex and not all factors which make up this history are explanatorily relevant.

Following the notation employed by Ylikoski (2001), let us formulate a contrastive explanandum as:  $f [c]$ , where  $f$  is the fact to be explained and  $c$  is the foil, an alternative to it<sup>46</sup>, where  $c$

---

<sup>45</sup>  $f$  and  $c$  can refer both to regularities (every autumn leaves turning their colours), and to singular occurrences (Willie Sutton robbed a bank). Proponents of the contrastive approach are usually concerned with analysing explanations of singular events (singular explanations). Though this makes their terminology and concepts better fit for singular explanations, no major problem arises when applying the framework to explanations of patterns, regularities, laws and the like (theoretical explanations). Since I am here concerned with both singular explanations and theoretical explanations, modifications are sometimes needed. One such modification concerns replacing the notion of causal history, typically used by proponents of the contrastive approach, with that of causal nexus. Talking of the causal history leading to, say, a regularity is somewhat incorrect and this motivates my usage of the broader concept of causal nexus.

<sup>46</sup> Supporters of the contrastive approach disagree on whether the foil has or does not have to be an incompatible occurrence (cf. Lipton, 1990; Ylikoski, 2001)

can be either a single alternative or a list or class of alternatives; in the latter case, we talk about a contrast class. Employing this notation allows us to indicate contrasts in a way which is independent of linguistic vagaries.

But what makes an explanation of a contrast adequate? In order to answer this question we need a way to mark off sensible from non-sensible contrasts. Proponents of the contrastive approach admit that not all contrasts are sensible, but the criteria for sorting them out vary among them. Think of a question such as why spatial agglomeration occurs rather than an increase in the sales of iPod, and compare it with one asking for an explanation of why agglomeration of economic activity occurs rather than spatial dispersion. It is intuitively clear that although the latter is a sensible question, the former is not. We therefore need a way to individuate what makes certain contrasts sensible and others not. Lipton (1990) suggests that in order to construe sensible contrasts,  $f$  and  $c$  should have largely similar causal histories. Lipton's *difference condition* provides the central criterion for individuating relevant causal factors. This says that "to explain why P rather than Q, we must cite a causal difference between P and Q, consisting of the cause of P and the absence of a corresponding event in the history of not-Q" where "a corresponding event is something that would bear the same relation to Q as the cause of P bears to P" (Lipton 1990: 257)<sup>47</sup>. In other words, to construe sensible contrasts we compare phenomena which have similar causal nexuses, and

---

<sup>47</sup> Lipton's difference condition is inspired by John Stuart Mill's Method of Difference (Mill 1843) according to which a cause must lie among the antecedent differences between a case where the effect occurs and a similar case where it does not (Lipton 1991: 43; cf. Garfinkel 1981:40)

conclude that the difference between them is the relevant explanatory factor<sup>48</sup>.

It has been noticed that the contrastive approach to explanation makes sense of scientific experiments as means to acquire causal knowledge (Ylikoski 2001: 22): Scientific experiments work in a contrastive setting and employ manipulation in order to learn how changes in the explanatory factors change the outcome. Theoretical models work quite like scientific experiments in that, by way of theoretical, rather than material, manipulations (cf. Mäki 1992b, 2005; Morgan 2002, 2003) they are used to inquire into how changes in the explanatory factors change the outcome. To do so, quite like scientific experiments, they require isolating away a host of elements for the purposes of focusing on a few causal relations, processes or mechanisms. This makes them unrealistic in the sense of being partial (Mäki 1992a,b). Unlike scientific experiments where this is done by means of material isolations, in theoretical models those isolations are effected by means of

---

<sup>48</sup> Two qualifications are to be made. First, in case of incompatible contrasts the causal history/nexus of  $f$  and  $not-c$  (when we explain why  $f$  rather than  $c$ , it is implied that  $c$  is not the case, hence  $not-c$ ) will be the same. This does not represent a problem for Lipton's difference condition. This does not require the same event to be present in the history of  $f$  and absent in the history of  $not-c$ , but only that the cited cause of  $f$  finds no corresponding event, as defined above, in the history of  $not-c$  (cf. Lipton 1991: 44). Second, there might be a number of differences between the two causal nexuses, but not all of them are explanatory. Ylikoski (2001: 41) proposes to add a mechanism requirement which asks that for  $a$  to be a cause of  $f|c$ , there has to be a mechanism ensuring that  $f$  occurs instead of  $c$  because of  $a$ . Note that the contrastive approach to explanation does not commit us to any particular account of causation. But in applying it, we need to have a way to identify differences which are both causal and explanatory. Not all differences in the causal history are causal, not all causal differences are explanatory, and not all explanations are adequate explanations. Woodward's manipulability theory seems to be quite

omissions and idealizations. The presence of idealizations makes theoretical models unrealistic in a sense that is different from simply being partial since they include elements which are flatly false when interpreted as referring to the real world (Mäki 1992 a,b).

The philosophical conundrum about unrealistic models is how they can succeed in explaining real-world phenomena. Despite the obvious significance of this issue, providing a solution to it is far beyond the scope of this paper. Independently of whether, and how, inferences from the model world to the real world should be made, it is a matter of fact that those inferences are actually made and that many disputes over unrealistic models concern whether a given theoretical model is sufficiently realistic for these inferences to be made.

What contrastivity has to do with the issue of unrealisticness is that the degree to which a model is unrealistic is to be judged by paying attention to the contrast class which the phenomenon to be explained is compared to. If we take explanatory questions to be contrastive, then unrealistic elements such as omissions and idealisations are (at least partly) dependent on the contrasts. Omissions and idealisations will sometimes refer to factors which belong to the shared causal background between the fact and the foil. These unrealistic elements have an ontological justification and, when correctly selected, do not endanger the truth of an explanation of why *f* rather than *c*. Let us identify the function of these assumptions as that of *fixing the causal background*<sup>49</sup>.

---

suitable to accompany the contrastive approach (cf. Ylikoski 2001; Woodward 2000; 2003).

<sup>49</sup> Notice that in some cases the assumptions might refer to factors which in the real world constitute explanatory differences between the fact and the foil. In such cases, some of the assumptions have the effect of turning

Since models typically isolate one or a few causal factors, processes or mechanisms, they might be thought to be at most able to offer a partial or incomplete explanation for the phenomenon. Yet, to provide the complete explanation of a fact, that is, an explanation of the fact itself compared to all its possible alternatives is in most cases impossible. Moreover, the few causal factors, processes or mechanisms isolated in a model might be both necessary and sufficient to account for a given contrastive explanandum. When this is so, then the explanation provided by the model in question is complete. The difference between explanations offered by theoretical models and explanations given by other means is therefore not a matter of kind, but of degree. Let me elaborate.

The focus on one or a few causal mechanisms often depicted at a highly abstract level, usually (though not necessarily!) means that the number of contrastive explananda a model can account for is very limited: the contrast class,  $c$  with respect to a given  $f$  will typically include only one or a few elements. The explanatory potential, assessed in terms of the variety of contrastive questions about a given phenomenon a theory or

---

those differences into background conditions. In so doing, the theorist creates a set up where the only explanatorily relevant difference between the fact and the foil is the one she sets out to investigate. Interestingly, the explanatory factor so isolated is taken to retain its explanatory power outside the model. The explanation of why  $f[c]$  in terms of this isolated causal difference might still be true. Yet, it would be incomplete since in the real world there are further explanatory causes of why  $f[c]$ . The locution ‘fixing the causal background’ is drawn from Ylikoski (2001). Ylikoski employs it to refer to the role of contrasts in discriminating between explanatory and non-explanatory causes. My usage is slightly different as it refers to *assumptions* which fix the causal background: Some of the elements of the assumed causal background are shared by the fact and the foil in the real world (in the real world these are not explanatory differences) while some others are *assumed* to be shared by the fact and the foil by the theorist, even though in the real world they might be explanatory differences.

model is potentially capable of answering, is therefore quite restricted. Let us call this dimension of explanatory potential *explanatory variety*. Explanatory variety should be distinguished from *explanatory depth*, assessed in terms of the number and variety of detailed underlying causal connections, and from *explanatory scope*, assessed in terms of number of kinds of phenomena a theory or model explain (cf. Mäki 2004b for a similar distinction).

It has already been argued that contrastive analyses are of help in the solution of disputes between apparently competing explanations (cf. Garfinkel 1981; Ylikoski 2001; Day 2002). The point that contrastive focus sheds light on the explanatory potential of models and that these have usually restricted contrastive force is made by Morton (1990, 1993). Here, I combine both lines of arguments so as to promote contrastivity as an aid to the solution of debates over unrealistic models. If theoretical models explain by way of showing why  $f[c]$ , then we should not expect either our own models or rival ones to explain a plain phenomenon.

### 5.3 The case of geographical economics

In this section I will examine a few of the unrealistic elements of the GE models which have been contested by economic geographers. Before proceeding, a word of caution is needed on the methodology that I will employ. Although the authors' writings might reveal clues about the intended contrast, the focus of my reconstruction is not meant to sort out the intentions of the authors but rather what a given explanation as it stands can be plausibly supposed to be able to explain. Furthermore, nothing is suggested on the adequacy of the GE explanation. The explanation should thus be intended as

a potential explanation, that is, an explanation about which we do not yet know whether it is correct.

Finally, I discuss a simplified version of the GE approach, which corresponds to the closest extensions and refinements of the core model of GE, namely Krugman (1991a). This simplification does not have any implication for the effectiveness of my arguments. This is, first, because most of the commentaries of economic geographers that I consider here do refer to this core model and its modifications and, second, because my objective, that is, to illustrate the usefulness of the contrastive approach, goes beyond the specifics of this case.

Some of the assumptions that I discuss have been subsequently relaxed and in some cases this might have resulted into an extension of the explanatory potential of the GE models. Though I do not discuss this issue here, the contrastive approach serves well also in a dynamic context: it helps to assess whether and how the explanatory potential of a given theoretical approach changes over time, and hence, to make judgments regarding explanatory progress.

### 5.3.1 The geographical economics framework and its critics

GE seeks to explain the uneven distribution of economic activity in space, and more precisely, the phenomenon of spatial agglomeration taking place at different spatial scales. The concept of agglomeration refers to seemingly very distinct empirical phenomena such as the existence of industry clusters, the emergence of core-periphery patterns, the existence of cities and systems of cities, the pattern of international trade and specialization, the causes of economic growth and development.

The uneven distribution of economic activity repeats itself at different levels of spatial aggregation. To geographical



economists this suggests that these apparently distinct phenomena might be (at least partly) brought about by the same *kind* of economic mechanisms, viz. ‘economic mechanisms yielding agglomeration by relying on the trade-off between various forms of increasing returns and different types of mobility costs’ (Fujita and Thisse 2002: 1). Instead of investigating each of those phenomena separately, using a different theory for each, GE proposes a unified approach. (For a detailed discussion, see Chapter 2).

The framework of GE has two building blocks: the presence of increasing returns at the firm level and transportation/trade costs<sup>50</sup>. Increasing returns at the firm level requires dropping the assumption of perfect competition and replacing it with that of imperfect competition, modelled according to the Dixit-Stiglitz monopolistic competition model<sup>51</sup>. At the aggregate level, increasing returns at the firm level and transportation costs give rise to pecuniary externalities. Pecuniary externalities are transmitted through the market via price effects and, simply

---

<sup>50</sup> Increasing returns to scale are defined as a decrease in the average costs per unit of output for the individual firm as the level of output increases. Transportation costs are generally assumed to be of the “iceberg form” meaning that part of the good “melts” in transit and hence only a fraction of it arrives at destination. Externalities are defined as a decrease in average costs as a result of an increase at the level of output of the whole industry. Pecuniary externalities are externalities that are transmitted through the market via price effects for each firm, which, as a consequence, may decide to change its output decisions.

<sup>51</sup> The Dixit-Stiglitz model provides a way to model imperfect competition. The Dixit-Stiglitz is a general equilibrium model where there is a large number of firms producing differentiated products under increasing returns to scale. Firms take the price-setting behavior of other firms as given and do not take into account the effects of changing their own price onto the price index. The products are symmetric, which means that consumers do not prefer one variety to another. However, they prefer variety for its own sake which means that they always prefer to consume a unit of a new variety than an additional unit of a product they have already consumed.

put, their presence implies that the more firms and workers there are in a locality, the more the locality becomes attractive as a location for further firms and workers. This creates a cumulative process whose end result might be that all economic activity turns out to be concentrated in one locality.

While pecuniary externalities are forces that push towards the concentration of economic activity (agglomerating or centripetal forces), the presence of immobile factors, of congestion and the like push towards dispersion (dispersing or centrifugal forces). The models are characterized by the presence of multiple equilibria: whether or not, and where, agglomeration arises depends on the relative strength of those forces and on initial conditions, that is, on previous locational decisions. The cumulative nature of the process of agglomeration is such that a small advantage of one location due to locational chance events in the past can have snowball effects which turn that location into the centre of economic activity, even though this outcome might not be the optimal one.

Economic geographers who are also concerned with spatial issues have been extremely critical of the GE approach. Criticisms range from blaming GE for its imperialistic tendencies (Johnston, 1992) and for inspiring the feeling of *déjà vu* (Berry 1999; Hoare 1992; Martin 1999) to charges directed at the employment of formal mathematical models based on neo-classical economics that do not allow for capturing the complexity of real-world phenomena (Clark 1998; Dymski 1992; Martin 1999). In particular, economic geographers have identified a series of issues that are poorly treated or even omitted altogether in the GE models. For instance, GE is held to downplay the key role of technological and knowledge spillovers in explaining agglomeration (cf. Olsen 2002) and for paying too little attention to the issue of “spatial contingency”

which results in the failure of acknowledging “spatial interdependence” (Sjöberg and Sjöholm 2002). Finally, economic geographers noted that GE overlooks the possibility that different mechanisms may be at work at different spatial scales (Martin 1999; Sjöberg and Sjöholm 2002).

In what follows I will focus only on a few of the criticisms of unrealisticness made by economic geographers (for a broader discussion of the dispute, see Chapter 3). The following analysis will prove the usefulness of the contrastive approach by showing (i) how some of the unrealistic elements of the GE models can be interpreted as serving the function of fixing the causal background, (ii) how formulating the GE explanandum in contrastive terms helps to delimit precisely its models’ explanatory potential which turns out to be very limited, and (iii) that the criticisms of economic geographers here discussed end up being misguided.

### 5.3.2 Contrastive foci on spatial agglomeration

Take the following (non contrastive) generic explanation-seeking question [Q] about spatial agglomeration.

[Q] Why is economic activity agglomerated in cities, clusters, countries and groups of countries?

Formulating [Q] in contrastive terms gives us at least the following three contrastive questions.

[Q-i] Why is economic activity [*other activities, e.g. political*] agglomerated in cities, clusters, countries and groups of countries?

[Q-ii] Why is economic activity agglomerated [*alternatively arranged in space, e.g. dispersed*] in cities, clusters, countries and groups of countries?

[Q-iii] Why is economic activity agglomerated in (certain) regions countries and groups of countries [*other regions, countries and groups of countries*]<sup>52</sup>?

It is easy to see how adding a contrast to [Q] gives rise to three different explanation-seeking questions whose answers are likely to require three distinct sets of explanatory factors. Questions [Q-i]-[Q-iii] and modifications thereof should be kept in mind for they will guide us in identifying the explanatory potential of the GE models. To do so, I proceed as follows. First, I discuss the GE's selection of explanatory factors and in this way derive a proposal for what might be plausibly taken to be their contrastive explanandum. Second, I consider a selection of the criticisms by economic geographers. These serve to spell out the limitations of the explanatory power of the GE models, even though I show that the criticisms end up failing to hit the target.

### 5.3.3 Spatial agglomeration at all scales and at all times [dispersion]

As a first step towards identifying the explanatory potential of the GE models, it is useful to begin with what geographical economists name the "the folk theorem of spatial economics" (cf. Ottaviano 2003) which helps singling out the main elements of the GE explanation. The folk theorem states that a firm's location decision turns into a non-trivial economic problem when there are [1] transportation costs and [2] increasing

---

<sup>52</sup> Notice that in this formulation reference to cities and clusters has been dropped as cities and clusters are agglomerations.

returns to scale. Under these conditions, the firm faces a trade-off between market proximity and concentration of production (Ottaviano 2003: 666), that is, between reducing transportation costs and reaping advantages from larger scales. However, the presence of both [1] and [2] is incompatible with a perfectly competitive market. As shown by "the spatial impossibility theorem" (Starrett 1978; see Fujita and Thisse 2002 and Ottaviano 2003), if space is neutral, then "total transport costs in the economy must be zero at any spatial competitive equilibrium, and thus regional specialization, cities and trade cannot be equilibrium outcomes" (Fujita and Thisse 2002: 26). As a consequence, the only equilibrium compatible with a competitive market and a neutral space is a situation of local autarkies where each person produces for her own consumption, a situation referred to as "backyard capitalism". Within a competitive framework agglomeration can only arise by virtue of so-called "first-nature factors", that is, differences between locations such as those due to climatic conditions, proximity to means of transportation and to natural resources. First-nature factors, however, cannot be the whole explanation of agglomeration because they cannot account for observed differences in degrees of agglomeration between locations which are fairly similar in terms of these first-nature characteristics (Ottaviano 2003).

To investigate an alternative explanation for the observed uneven spatial distribution of economic activity, GE assumes locations to be identical in all respects, and agglomeration to arise solely as the result of the presence of increasing returns, imperfect competition and externalities. By assuming identical locations, the GE models neutralise the effects of first-nature factors and "isolates second nature as the component of the spatial variation that cannot be explained in term of first

nature” (Ottaviano and Thisse 2004: 3). As geographical economists put it:

One of the attractive features of the core model of geographical economics is the neutrality of space. Since, by construction, no location is preferred initially over other locations, agglomeration (a center-periphery structure) is not imposed but follows endogenously from the model (Brakman et al. 2001: 167)

Focusing on this key isolation allows us to specify the GE explanandum more precisely. In contrastive terms, agglomeration is compared to backyard capitalism. The explanandum is therefore something like: what makes agglomeration occur rather than backyard capitalism? Recall that backyard capitalism is the outcome of a situation characterised by [1] neutral space; [2] transportation costs; [3] constant or decreasing returns to scale and [4] competitive markets. As depicted in Figure 5.1, the only relevant difference in the causal nexuses of the two outcomes (that is, agglomeration and backyard capitalism) is the presence of increasing returns and imperfectly competitive markets, which together with transportation costs can, under certain conditions, produce externalities, strong enough to make agglomeration a stable equilibrium.

[Figure 5.1 about here]

This formulation helps unearthing the rationale of the GE models: in order to explain the actual degree of agglomeration in the real world, increasing returns to scale are necessary. As geographical economists put it, the assumption of increasing returns is essential: “if there is no loss to splitting up production

there is really no location choice to be made" (Baldwin et al., 2003: 60). Backyard capitalism, however, only arises when there are no increasing returns, and hence, it is not a possible outcome in the GE models. The more general explanatory question is thus to be formulated in terms of a contrast between spatial agglomeration and spatial dispersion (where backyard capitalism is to be seen as an extreme case of dispersion where increasing returns are absent), that is, a sub-question to [Q-ii]:

**[Q-ii(a)]** Why is economic activity agglomerated [*dispersed*] in cities, clusters, countries and groups of countries?

The assumption of identical locations can be seen as fixing the causal background shared by both agglomeration and dispersion. In such a set up the only difference between the causal nexuses is the realisation of pecuniary externalities, arising from the interaction between increasing returns and transportation costs, which are strong enough to offset the effects of the centripetal forces.

Though pecuniary externalities are in principle sufficient to discriminate between agglomeration and dispersion, there are other determinants of why economic activity is agglomerated rather than dispersed. These are the presence of first-nature factors and technological externalities. Technological externalities can arise in perfectly competitive markets (Scitovski 1954) as they are transmitted through non-market interactions and they directly affect either the utility function of workers or the production function of firms. This type of externalities is given pride of place as explanantia of agglomeration by economic geographers. GE models, by contrast, virtually ignore their explanatory role, though

geographical economists admit that in the real world technological externalities are relevant (cf. Fujita and Thisse 2002).

Two main reasons for the GE's almost exclusive reliance on pecuniary externalities can be identified. First, GE aims at analysing those mechanisms that can be taken to operate across instances of agglomeration at different spatial scales. Technological externalities, instead, require a high degree of spatial proximity to operate. Thus, they are likely not to be very effective at the international or global scale (Ottaviano and Puga 1998: 708). A similar justification is given by Krugman who argues that technological externalities are not the typical reason for agglomeration, both across different industries and through time (Krugman 1991b: 54). Both statements suggest that GE seeks to explain the phenomenon of agglomeration vis-à-vis spatial dispersion as they occur *at various spatial scales*. To do so, it isolates causal mechanisms that can be thought to operate scale-independently. This focus on scale-independent mechanisms, to which we will return below, points to the unifying thrust of the GE explanation. Explanatory unification means explaining much by little, that is explaining a great number of classes of phenomena by relying on a few basic principles and/or mechanisms. It involves what I termed *explanatory scope*. GE employs the trade-off between increasing returns and transportation costs to explain a host of phenomena at different spatial scales, which were previously attended to by other theories. The pursuit of unification in GE requires the explanantia to be selected among those mechanisms which are thought to be in operation at all levels of aggregation and in various instances of agglomeration (Chapter 2 of this thesis analyses in detail explanatory unification in GE).

The second reason is that technological externalities are difficult to derive from micro-economic considerations. GE sets



out to get inside the black box of agglomeration economies and "derive the self-reinforcing character of spatial concentration from more fundamental considerations" (Fujita et al. 1999:4). As Ottaviano (2003: 666) puts it, the "comparative advantage [of pecuniary externalities] lies in the possibility of relating their emergence to a set of well-defined economic parameters", namely transportation costs, increasing returns to scale and imperfect competition (cf. also Fujita and Thisse 2002: 9). The underlying argument is that since technological externalities cannot be derived from well-defined microeconomic parameters, an explanation citing them is at best a shallow explanation. It would be very much like claiming that agglomerations form because of agglomeration economies, as Fujita et al. (1999: 4) report a physicist to have sarcastically noted.

Before the GE's advent, also pecuniary externalities were largely treated as black boxes. The problem, so the GE argument goes, was that economists lacked the analytical tools to deal with increasing returns and imperfect markets, which as we have seen, are deemed to be necessary for the realization of pecuniary externalities. GE seeks to increase the *explanatory depth* of an explanation of spatial agglomeration relying on pecuniary externalities by showing how these, and hence agglomeration, can be derived from more fundamental economic parameters. This makes sense of the geographical economists' claim that opening up the black box of agglomeration economies, that is producing general equilibrium models in which pecuniary externalities are shown to arise endogenously, is their prime theoretical contribution (cf. Fujita et al. 1999; Brakman et al. 2001).

### 5.3.4 Agglomeration takes place in certain places [other places]

After having identified the contrastive explanandum which can be plausibly ascribed to GE, I now consider a few of the criticisms raised by economic geographers over the unrealisticness of the GE models in order to identify precisely the models' explanatory potential.

We have seen that the assumption of identical locations serves the role of fixing the causal background. Pecuniary externalities arising from the interaction of increasing returns and transportation costs are the sole determinants of agglomeration. Obviously, in the real world locations are not identical. These first-nature characteristics do make a difference<sup>53</sup>, and so do many other specific features of locations. Economic geographers have not failed to notice this highly unrealistic element of the GE models. In particular, they have not so much complained about the absence of first-nature characteristics from the GE models, as about the lack of social, cultural and institutional features of those locations<sup>54</sup>.

---

<sup>53</sup> First-nature factors can explain why and where agglomeration occurs. As we have seen, the GE argument is that they cannot provide the whole explanation of agglomeration because even locations similar in terms of first-nature factors have different degrees of agglomeration. The kind of mechanisms identified by GE is taken to be at work also when locations differ in terms of first-nature factors.

<sup>54</sup> GE explicitly assumes that locations are identical in terms of first-nature factors. Cultural, social and institutional factors instead are omitted without any mention. Assuming that locations are identical in all respects implies that locations do not have the aforementioned characteristics at all. Notice, however, that first-nature factors are present before any agglomeration takes place. In contrast, cultural, social and institutional factors develop together with agglomerations and can be interpreted as

They complain that “the ‘messy’ social, cultural and institutional factors [...] are assumed of secondary importance” in GE (Martin 1999: 75; see also Amin and Thrift 2000, for a similar point regarding economics more generally) and yet “it is precisely the social, institutional and political embeddedness that can play a key role in determining the possibilities for or constraints in development, and thus *why spatial agglomeration of economic activity occurs in particular places and not others*” (Martin 1999: 75, emphasis added). The question the GE models are, according to economic geographers, unable to answer is here already formulated contrastively, namely [Q-iii]. Arguably, [Q-iii] is a question whose answer is likely to require a different set of explanantia than the one employed by GE, even though the former set is likely to encompass the latter. But this does not imply that the GE models are worthless from an explanatory viewpoint. Though GE might well be unable to tell us why agglomeration occurs in particular places and not others [Q-iii], this does not imply that it is also unable to explain why spatial agglomeration rather than spatial dispersion occurs [Q-ii(a)]. This qualification has two important implications, which will recur throughout the discussion: on the one hand, it shows that this criticism is not fatal and, on the other, that GE might lack the resources to provide an answer to [Q-iii].

Another problem which economic geographers take to beset the GE models is that locations can either represent cities, regions, or groups of countries depending on the issue under

---

contributing or counteracting mechanisms. The economic mechanisms of the GE models do sustain, when they do, agglomerations, but they can either be counteracted by the absence of appropriate institutional, cultural or social conditions or reinforced by their presence. Economic geographers, who typically place a strong emphasis on these conditions, variously refer to them with labels such as “untraded interdependencies”, “institutional thickness”, “social capital”.

investigation. Here is one telling exposition from the same critic.

So cavalier is the treatment of space and place that the same model is often used to explain spatial agglomeration [...] at vastly different scales, from the international level, to broad core-periphery patterns within nations, to local urban industrial concentrations, and even intra-urban neighborhoods. Processes are thus assumed to be largely scale-independent [...] It seems likely that different externalities operate at different geographical scales [...] The spatial agglomeration models may well predict, that, under specific assumptions, industrial localisation and specialisation will occur, but they are unable to tell us where it actually occurs, or why in *particular places* and not in others. (Martin 1999: 78, emphasis in the original)

Again, it is claimed here that GE cannot provide an answer to [Q-iii], which, as noted above, is likely to require a set of explanatory factors different from those of GE. Geographical economists justify the conflation between spatial scales by suggesting that the same or similar mechanisms might be at work at different spatial scales, referring once more to the unifying character of their explanation. It seems plausible to think that both scale-dependent and scale-independent mechanisms affect the emergence and persistence of spatial agglomeration—and economic geographers do not seem to deny this. It might well be, as Martin suggests, that scale-dependent mechanisms or processes are necessary explanantia for [Q-iii]. This, however, does not conflict with the possibility that the same or similar mechanisms, responsible for the

occurrence and sustenance of agglomeration vis-à-vis spatial dispersion, are at work at different spatial scales.

In sum, both sets of criticisms discussed here are per se insufficient to undermine the potential of GE to explain [Q-ii(a)]<sup>55</sup>. They do however indicate its limitations in providing an explanation of [Q-iii].

### 5.3.5 Spatial clusters exist [single unilocational firms]?

In a recent article, economic geographers Malmberg and Maskell (2002) notice that explanations of the existence of spatial clustering, that is, the spatial concentration of related firms taking place at a local scale, typically proceed "by identifying and analyzing those permanent advantages that may accrue to firms located close to other similar and related firms, rather than being in isolation" (Ibid.: 432) and argue that thus far these accounts have been unsatisfactory.

They propose what a satisfactory theory of spatial clustering should include:

[...] the theory must provide an explanation for the advantages that many related firms might accrue when collocated but which are not available to a hypothetical

---

<sup>55</sup> This argument involves a simplification. First, GE models cannot provide the complete explanation for why agglomeration rather than dispersion takes place because the models do not include all the explanatorily relevant factors. As seen, first-nature factors and technological externalities are in some cases explanatorily relevant for real-world agglomeration [dispersion]. Second, the GE models could also provide a very partial explanation to the question as to why agglomeration occurs in some places and not others. However, it seems clear that their explanatory potential is much higher if judged against the former question than the latter.

single firm carrying out precisely the same activities, even at the same location, using the same suppliers, customers and workforce (Ibid.: 438).

For Malmberg and Maskell, previous accounts relying on cost reductions cannot explain the existence of clusters because “a superior way to reduce such costs would presumably consist of joining the different activities and placing them under one common ownership, thus eliminating most costs of interaction” (Ibid.: 438). Though geographical economists themselves admit that at the local scale pecuniary advantages are likely to be less important *vis-à-vis* other kinds of externalities (Ottaviano and Puga, 1998), this does not amount to saying that they do not play any role at all. The underlying conception seems to be that the GE explanatory factors are always in operation; what varies is their strength relative to other forces. If this is so, then the critical argument of Malmberg and Maskell applies to the GE models as well: if cost reductions arising from the realisation of pecuniary externalities are the kinds of advantages which are sought for when firms locate close to each other, then a single firm carrying out the same activities, at the same location, using the same suppliers, customers and workforce would be a more superior arrangement to reduce those costs due to distance. The GE models appear not to have the internal resources to provide a response to this puzzle.

Reformulating the claim in contrastive terms, it says that the GE explanantia cannot explain the contrast between spatial clustering and a hypothetical single firm and hence the existence of spatial clustering (for a more thorough comparison, see Chapter 4).

Independently of its accuracy, Malmberg and Maskell's argument serves us here to identify a further foil to agglomeration<sup>56</sup>, which gives us the following explanandum:

**[Q-ii(b)]** Why is economic activity agglomerated in clusters [*unilocational single firm*]?

If the GE explanatory factors cannot account for the contrast between spatial agglomeration and unilocational single firm, then this would further narrow down the GE explanandum, or better reduce its contrastive force. (Analogous reasoning can be provided for other spatial arrangements included in the contrast class to spatial agglomeration when viewed *qua* form of spatial organization). The set of contrasts GE has the resources to account for appears to be limited to just one alternative to agglomeration, namely dispersion. It therefore follows that the GE contrastive explanation seeking question can be reformulated as:

**[Q-ii(a)]** Why is economic activity agglomerated [*disperse d*] in cities, clusters and groups of countries?

---

<sup>56</sup> It can be argued that [*dispersion*] includes [*unilocational single firm*]: if in place of a number of spatially clustered firms, there was a single firm, then the final configuration of the space economy would be one of spreading. Such a conclusion, however, runs counter to the very rationale of the contrastive approach. In order to see why, suppose there are  $n$  firms in an industry and two localities, 1 and 2. While GE compares an outcome in which the  $n$  firms agglomerate in either 1 or 2 (as seen, in the GE models it does not generally matter which location will host agglomeration) to one in which firms spread between 1 and 2, Malmberg and Maskell compare a situation in which there are  $n$  firms located either in 1 (2) to one in which there is a larger single firm in 1 (2). The two questions are indeed different, and it might be the case that they require different answers.

This analysis shows that the range of contrastive questions the GE models seem to be capable of answering is limited, and hence that the degree of explanatory variety is minimal (as mentioned earlier, refinements and extensions of the basic approach are likely to lead to an increase of explanatory variety). Recall that explanatory variety has been distinguished from both explanatory scope and explanatory depth, and that the GE models were said to lack neither (potential) explanatory scope nor (potential) explanatory depth. The GE case seems to suggest that a minimal degree of explanatory variety is compatible with a higher degree of explanatory scope and depth. The GE case shows that the various dimensions on which explanatory potential can be assessed are not related in a straightforward manner. Distinguishing among them, thus, makes judgments of explanatory potential much harder and yet more accurate.

### 5.3.6 A distinct role for unrealistic elements

Obviously not all assumptions fix the causal background. Some assumptions do play other, in some cases connected roles. Some assumptions are to be interpreted as negligibility assumptions, others as applicability assumptions and still others as tractability assumptions (Musgrave 1981; Mäki 2000; Hindriks *forth*). Sometimes, distinguishing these roles might not be so easy in practice, and in other cases, the same assumption may play different roles at the same time (for example, the omission of technological externalities might be



interpreted as a tractability assumption, and yet we have seen that it also has an ontological justification<sup>57</sup>).

To be clear about the contribution of the contrastive approach in the context of the ongoing discussion over unrealistic models in economics, it is important to see how unrealistic assumptions fixing the causal background cannot be reduced to any of the roles identified by the threefold typology mentioned above. I begin with negligibility assumptions. A negligibility assumption takes a given factor  $F$  to have a negligible effect on the phenomenon under investigation given the purposes of the investigation (Mäki 2000: 320). To assert that the factors which are kept fixed belong to the causal background (or are made to belong to it) is not the same as asserting that the factors which are kept fixed have negligible effects on the phenomenon. In GE, differences in underlying characteristics among locations are not taken to have negligible effects on the occurrence of agglomeration. Notice that the definition of a negligibility assumption includes “the purpose of the investigation”. Contrastive analysis could be employed to specify the notion of purpose, and hence, that of negligibility given a purpose. The purpose determines the contrast and the contrast identifies what factors are negligible: some factors might be negligible for some contrasts but not for others<sup>58</sup>.

This nonetheless fails to capture the function of fixing the causal background. For instance, we could reformulate the assumption of identical locations as a negligibility assumption

---

<sup>57</sup> The roles of unrealistic assumptions discussed here are not meant to be exhaustive. For example, Mäki (2000) identifies a further kind, that is, early-step assumptions. Furthermore, the omission of technological externalities implies that the GE models have a better fit when applied to spatial agglomeration vis-à-vis dispersion at larger scales than the local ones. This identifies yet another role for unrealistic elements. Thanks to Uskali Mäki for having suggested this to me.

<sup>58</sup> I owe this observation to Petri Ylikoski

asserting that differences in underlying characteristics have negligible effects given that the purpose of the investigation is to identify what makes agglomeration rather than dispersion occur. But this is not what is claimed in the GE models. Differences in underlying characteristics are not thought to be negligible. They are thought not to be explanatory; the assumption of identical locations fixes them as part of the causal background.

Consider then applicability assumptions. An applicability assumption claims that the model applies only in the domain where the assumption holds true, or where its falsehood is negligible (Mäki 2000: 323-325). Yet, it does not seem to do justice to the GE rationale to take the assumption of identical locations to determine the domain of applicability of the models. If it were so interpreted, the models would be held to be applicable only in those circumstances where locations are identical or very similar indeed. However, as suggested above, the kinds of mechanism identified by GE are supposed to operate independently of whether locations are similar or not.

In a tractability assumption, the effects of a set of factors are neutralised in order to make a problem (more) tractable (Hindriks, forth). Take the assumption of identical locations. This is not made solely for convenience, but it appears to play a role in the search for causal knowledge concerning the contrastive explanandum. Compare this with the assumption that firms take the price indexes to be constant as they solve their profit maximisation problem. Together with the assumption of constant-elasticity demand functions, it implies that all market-size effects work through changes in variety; thereby totally ignoring the procompetitive effects which would work in the same direction (an increase in market size would reduce the price-cost margins and average costs). As geographical economists themselves concede, “[this] is a

dramatic simplification, allowing us to model issues that might otherwise seem quite intractable" (Fujita, et al. 1999: 52), thereby explicitly qualifying it as a tractability assumption.

It follows that contrastivity identifies a distinct role for unrealistic assumptions of theoretical models, that is, *fixing the causal background*, which cannot be reduced to any of the aforementioned roles.

#### 5.4 Concluding remarks

The dispute over the unrealisticness of economics models has a long history. Many of the criticisms over theoretical models both from within and outside the economics discipline centre on the question of those models' degree of realisticness and justifiability.

Contrastivity pinpoints a distinctive role for some of the unrealistic elements in theoretical models, that is, fixing the causal background shared by the fact and the foil. Unrealistic elements of models serving this purpose do not endanger the (possible) adequacy and veracity of the explanation thereby offered. We have seen how some of the unrealistic elements of the GE models can be so interpreted.

The contrastive approach to explanation proves to be a useful aid in the solution of disputes over some unrealistic elements. An analysis of theoretical models in contrastive terms facilitates the precise delimitation of the explanatory potential of a given model and an assessment of the adequacy of its isolations. If theoretical models explain contrasts, then we should not expect either our own models or rival ones to explain a plain phenomenon. I have shown that once the GE explanandum has been precisely delimited, some of the criticisms of unrealisticness against them raised by economic

geographers turn out to be misguided. This said, it still remains an open question whether the GE models are realistic enough for the purposes of making correct inferences from the model world to the real world.

## References

- Amin, A. and Thrift, N. (2000) "What kind of economic theory for what kind of economic geography." *Antipode* 32 (1), 4-9.
- Aronson, J. (1984) *A Realist Philosophy of Science*. London: Macmillan.
- Baldwin, R. E. and Forslid, R. (2000) "The core-periphery model and endogenous growth: stabilizing and destabilizing integration." *Economica* 67, 307-324.
- Baldwin, R. Forslid, R. Martin, P. Ottaviano, G. and Robert-Nicoud F. (2003) *Economic Geography and Public Policy*. Princeton University Press.
- Barnes, E. (1992) "Explanatory unification and the problem of asymmetry." *Philosophy of Science* 59, 558-571.
- Barnes, E. (1994) "Why P rather than Q? The curiosities of fact and foil." *Philosophical Studies* 73, 35-53.
- Barnes, T. J. (2003) "The place of locational analysis: a selective and interpretive history." *Progress in Human Geography* 27 (1), 69-95.
- Bathelt, H., Malmberg, A. and Maskell, P. (2004) "Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation." *Progress in Human Geography* 28, 31-56.
- Beccattini, G. (1990) "The Marshallian industrial district as a socio-economic notion." In F. Pyke (ed) *Industrial Districts and Inter-firm Co-operation in Italy*. Geneva: International Institute for Labour Studies, 37-51.

- Bellandi, M. (1996) "Innovation and change in the Marshallian industrial district." *European Planning Studies* 4, 357-368.
- Berry, B.J.L. (1999) "Déjà vu, Mr Krugman." *Urban Geography* 20(1), 1-2.
- Bhaskar, R. (1978) *A Realist Theory of Science*. Brighton: Harvester, 2<sup>nd</sup> ed.
- Blaug, M. (1996) *Economic Theory in Retrospect*. Cambridge: Cambridge University Press, 5<sup>th</sup> ed.
- Boggs, J. S. and Rantisi, N.M. (2003) "The relational turn in economics." *Journal of Economic Geography* 3, 109-116.
- Brakman, Garretsen, van Marrewijk, (2004), *A second core model*. <http://www.few.eur.nl/few/people/vanmarrewijk/geography/further%20explanation.htm>.
- Brakman, S. and Garretsen, H. (2003) "Rethinking the 'new' geographical economics." *Regional Studies* 37 (6&7), 637-648.
- Brakman, S. Garretsen, H. and van Marrewijk, C. (2001) *An Introduction to Geographical Economics: Trade, Location and Growth*. Cambridge: Cambridge University Press.
- Brakman, S. Garretsen, H. Gigengack, R. van Marrewijk, C. and Wagenvoort, R. (1996) "Negative feedbacks in the economy and industrial location." *Journal of Regional Science* 35, 631-652.
- Camagni, R. (1995) "The concept of 'innovative milieu' and its relevance for public policies in European lagging regions." *Papers in Regional Science* 74, 317-340.
- Caniels, M.C.J. and Romijn, H. A. (2003) "What drives innovativeness in industrial clusters? Transcending the debate", *Cambridge Journal of Economics* 29(4), 497-515.
- Clark, G. L. (1998) "Stylized facts and close dialogue: methodology in economic geography." *Annals of the Association of American Geographers* 88, 73-87.
- Cooke, P. (2002) *Knowledge Economies*. London: Routledge.

- Crang, P. (1997) "Introduction: cultural turns and the (re)constitution of economic geography." In J. Wills and R. Lee (eds) *Geographies of Economies*. London: Arnold, 3-15.
- Day, M. (2002) *Competing Explanations. Exclusion and Importance in Historical Accounts*. Academic Dissertation, Faculty of Philosophy, University of Sheffield.
- Day, M. (2004) "Explanatory exclusion in history and social science." *Philosophy of the Social Sciences* 34(1), 20-37.
- Dixit, A.K. and Stiglitz, J.E. (1977) "Monopolistic competition and optimum product diversity" *American Economic Review* 67, 297-308.
- Duranton, G. and Puga, D. (2000) "Diversity and specialization in cities: why, where and when does it matter?" *Urban Studies* 37 (3), 533-555
- Dymski, G. A. (1996) "On Krugman's model of economic geography" *Geoforum* 27, 439-452.
- Ellison, G. and Glaeser, E.L. (1999) "The determinants of geographic concentration." *American Economic Review* 89 (2), 311-316.
- Feyerabend, P. (1963) "How to be a good empiricist—a plea for tolerance in matters epistemological." In B. Baumrin (ed) *Philosophy of Science, The Delaware Seminar*, 2, 3-39.
- Florida, R. and Kenney, M. (1990) *The Breakthrough Illusion. Corporate America's Failure to Move from Innovation to Mass Production*. New York: Basic Books.
- Friedman, M. (1974) "Explanation and Scientific Understanding." *Journal of Philosophy* 71, 5-19.
- Fujita, M. and Thisse, J-F. (1996) "Economics of agglomeration." *Journal of the Japanese and International Economies* 10, 339-78.
- Fujita, M. and Thisse J-F. (2002) *Economics of Agglomeration. Cities, Industrial Location and Regional Growth*. Cambridge: Cambridge University Press.

- Fujita, M. and Thisse, J-F. (2002) "Agglomeration and market interaction." *CEPR Discussion Paper* no. 3362. London, Centre for Economic Policy Research. <http://www.cepr.org/pubs/dps/DP3362.asp>.
- Fujita, M. Krugman, P. and Mori, T. (1999) "On the evolution of hierarchical urban systems." *European Economic Review* 43: 209-151.
- Fujita, M. Krugman, P. Venables, A. (1999) *The Spatial Economy. Cities, Regions and International Trade*. Cambridge Mass: MIT Press.
- Garfinkel, A. (1981) *Forms of Explanation*. New Haven, Yale University Press
- Giere, R. (1988) *Explaining Science*. Chicago: University of Chicago Press.
- Groenewegen, J. and Vromen, J. (1996) "A case for theoretical pluralism." In J. Groenewegen (ed) *Transaction Costs Economics and Beyond*. Dordrecht: Kluwer Academic Publishers, 365-380.
- Henderson, J. V. (1977) *Economic Theory and the Cities*. New York: Academic Press.
- Hindriks, F. (2005) "Unobservability, tractability, and the battle of assumptions." Forthcoming in *Journal of Economic Methodology*.
- Hirschman, A. (1958) *The Strategy of Economic Development*. New Haven: Yale University Press.
- Hoare, T. (1992) "Review of 'Geography and Trade'" *Regional Studies* 26, 679
- Imrie, R. Pinch, S. and Boyle, M. (1996) "Identities, citizenship and power in the cities." *Urban Studies* 33, 1255-1262.
- Jackson, F. and Pettit, P. (1992a) "In defense of explanatory ecumenism." *Economics and Philosophy* 8, 1-21.
- Jacobs, J. (1961) *The Death and Life of Great American Cities*. New York: Random House.



- Johnston, R.J. (1985) "Introduction: exploring the future of geography." In R.J. Johnston (ed) *The Future of Geography*. London and New York: Methuen & Co, 3-26.
- Johnston, R.J. (1992) "Review of Paul Krugman, 'Geography and Trade'." *Environment and Planning A* 24, 1066.
- Kaldor, N. (1978) "The case for regional policies." In N. Kaldor, *Further Essays on Economic Theory*. London: Duckworth, 139-54.
- Kim, J. (1988) "Explanatory realism, causal realism and explanatory exclusion." *Midwest Studies in Philosophy* 12, 225-39.
- Kim, J. (1989) "Mechanism, purpose, and explanatory exclusion." *Philosophical Perspectives* 3, 77-108.
- Kincaid, H. (1996) *Philosophical Foundations of the Social Sciences. Analyzing Controversies in Social Research*. Cambridge: Cambridge University Press.
- Kitcher, P. (1981) "Explanatory Unification." *Philosophy of Science* 48: 251-81
- Kitcher, P. (1989) "Explanatory unification and causal structure." *Minnesota Studies in the Philosophy of Science* 13, 410-505. Minneapolis: University of Minnesota Press.
- Kitcher, P. (1993) *The Advancement of Science*. Oxford: Oxford University Press.
- Kitcher, P. (1994) "Contrasting conceptions of social epistemology." In F. Schmitt (ed) *Socializing Epistemology: The Social Dimensions of Knowledge*. Lanham, Md: Rowman & Littlefield, 111-34.
- Kitcher, P. (2001) *Science, Truth and Democracy*. New York: Oxford University Press.
- Krugman, P. (1979) "Increasing returns, monopolistic competition, and international trade." *Journal of International Economics* 9, 469-479.

- Krugman, P. (1980) "Scale economies, product differentiation, and the pattern of trade." *American Economic Review* 70, 950-959.
- Krugman, P. (1991b) "Increasing returns and economic geography." *Journal of Political Economy* 99, 183-99.
- Krugman, P. (1991a) *Geography and Trade*. Cambridge, Mass. and London: MIT Press; Louvain: Louvain University Press.
- Krugman, P. (1995) *Development, Geography and Economic Theory*. Cambridge and London: MIT Press.
- Krugman, P. (1996) "How to be a crazy economist." In S. Medema and W.J. Samuels (Eds) *Foundations of Research in Economics: How Do Economists Do Economics?* Cheltenham, U.K. and Lyme N.H: Elgar, 131-141.
- Krugman, P. (1999a) "The role of geography in development." In B. Pleskovic and J. E. Stiglitz (eds) *Annual World Bank Conference on Development Economics, 1998* Washington, D.C. World Bank, 89-107.
- Krugman, P. (1999b) "What economists can learn from evolutionary theorists and vice versa." In J. Groenewegen and J. Vromen (eds) *Institutions and the Evolution of Capitalism*. Cheltenham: Edward Elgar, 17-29.
- Krugman, P. (2000) Where in the world is the 'new economic geography'? In G.L. Clark, M.S. Gertler, M. Feldman, *The Oxford Handbook of Economic Geography*. Oxford University Press.
- Krugman, P. and Venables, A. (1990) "Integration and the competitiveness of peripheral industry." In C. Bliss and J. Braga de Macedo (eds.), *Unity with Diversity in the European Economy*. Cambridge: Cambridge University Press, 56-75.
- Krugman, P. and Venables, A. (1996) "Integration, specialization and adjustment." *European Economic Review* 40, 959-67.

- Ladyman, J. (2003) "A review of Margaret Morrison's Unifying Scientific Theories: Physical Concepts and Mathematical Structures." *Journal of Economic Methodology* 10 (1), 91-97.
- Lewis, D. (1986) "Causal explanation." In D. Lewis, *Philosophical Papers*, Oxford and New York: Oxford University Press, 214-40.
- Lipton, P. (1990) "Contrastive explanations." In D. Knowles (ed) *Explanation and its Limits*. Cambridge University Press. Cambridge, 247-266.
- Lipton, P. (1991) *Inference to the Best Explanation*. Routledge, expanded 2<sup>nd</sup> edition 2004.
- Longino, H. (1990) *Science as Social Knowledge: Values, Objectivity in Scientific Inquiry*. Princeton: Princeton University Press.
- Longino, H. (1991) "Multiplying subjects and the diffusion of power." *Journal of Philosophy*, 666-674.
- Mäki, U. (1990) "Scientific realism and Austrian explanation." *Review of Political Economy* 2, 310-344.
- Mäki, U. (1992a) "Friedman and realism." *Research in the History of Economic Thought and Methodology* 10, 171-195.
- Mäki, U. (1992b) "On the method of isolation in economics" *Poznan Studies in the Philosophy of the Sciences and Humanities* 317-351.
- Mäki, U. (1993) "Isolation, idealization and truth in economics." In B. Hamminga and N. de Marchi (eds.) *Idealization in Economics*, Poznan Studies. Atlanta and Amsterdam: Rodopi, 147-68.
- Mäki, U. (1994) "Reorienting the assumptions issue." In Backhouse, R. (ed) *New directions in Economic Methodology*. London: Routledge.
- Mäki, U. (1997) "The one world and the many theories." In A Salanti and E. Screpanti *Pluralism in Economics*. Aldershot: Edward Elgar, 37-47.

- Mäki, U. (1998a) "Aspects of realism about economics." *Theoria* 13(2), 301-319.
- Mäki, U. (1998b) "Realism." In J. B. Davis, D. W. Hands and U. Mäki (eds) *The Handbook of Economic Methodology*. Edward Elgar, 404-409.
- Mäki, U. (1998c) "Unrealisticness." In J. B. Davis, D. W. Hands and U. Mäki (eds) *The Handbook of Economic Methodology*. Edward Elgar, 409-413.
- Mäki, U. (2000) "Kinds of assumptions and their truth. Shaking an untwisted F-twist." *Kyklos* 53(3), 303-22.
- Mäki, U. (2001a) "Explanatory unification: double and doubtful." *Philosophy of the Social Sciences* 31(4), 488-506.
- Mäki, U. (2001b) "Models." *International Encyclopaedia of the Social and Behavioral Sciences*. Elsevier, 9931-9937.
- Mäki, U. (2004a) "Realism and the nature of theory. A lesson from J.H. von Thünen for economists and geographers." *Environment and Planning A* 36, 1719-1736.
- Mäki, U. (2004b) "Theoretical isolation and explanatory progress: transaction cost economics and the dynamics of dispute." *Cambridge Journal of Economics* 28, 319-346.
- Mäki, U. (2005) "Models are experiments, experiments are models." *Journal of Economic Methodology* 12(2), 303-315
- Mäki, U. and Marchionni, C. (2005a) "Is geographical economics a form of intellectual imperialism?" *Unpublished manuscript*.
- Mäki, U. and Marchionni, C. (2005b) "Inter-theoretic relations in theoretical unification: the case of geographical economics." *Unpublished manuscript*.
- Mäki, U. and Oinas, P. (2004) "The narrow notion of realism in human geography." *Environment and Planning A* 36, 1755-1776
- Malmberg, A. and Maskell, P. (2002) "The elusive concept of localization economies: towards a knowledge-based theory

- of spatial clustering." *Environment and Planning A* 34, 429-449.
- Markusen, A. (1999) "Fuzzy concepts, scanty evidence, policy distance: The case for rigor and policy relevance in critical regional studies." *Regional Studies* 33(9), 869-84.
- Marshall, A. (1890) *Principles of Economics*. London: Macmillan.
- Martin, R. (1999) "The 'new' geographical turn in economics: some critical reflections." *Cambridge Journal of Economics* 23, 65-91.
- Martin, R. (2001) "Geography and public policy: the case of the missing agenda." *Progress in Human Geography* 25, 189-210.
- Martin, R. and Sunley, P. (1996) "Paul Krugman's geographical economics and its implications for regional development theory: a critical assessment" *Economic Geography* 3, 259-92.
- Martin, R. and Sunley, P. (2001) "Rethinking the "economic" in economic geography: broadening our vision or losing our focus?" *Antipode* 33(2), 148-161.
- Martin, R. and Sunley, P. (2003) "Deconstructing clusters: chaotic concept of policy panacea?" *Journal of Economic Geography* 3, 5-35.
- Maskell, P. (2001) "Towards a knowledge-based theory of the geographical cluster." *Industrial and Corporate Change* 10, 921-43.
- Maskell, P. and Lorenzen, M. (2004) "The cluster as market organisation." *Urban Studies* 41(5-6), 991-1009.
- Meardon, S. (2002) "On the new economic geography and the progress of geographical economics." In S. Bohem, C. Gehrke, H. D. Kurz and R. Sturn (eds) *Is There Progress in Economics? Knowledge, Truth and the History of Economic Thought*. Edward Elgar.
- Mill, J.S. (1843) *A system of logic*. Volume VII-VIII of the *Collected Works* Toronto: Toronto University Press.

- Morgan, M. (2002) "Model experiments and models in experiments." In L. Magnani and N. Nersessian (Eds) *Model-Based Reasoning: Science, Technology, Values*. New York: Kluwer Academy/Plenum, 41-58.
- Morgan, M. (2003) "Experiments without material intervention: Model experiments, virtual experiments and virtually experiments" [University of Amsterdam Research Memorandum (2000)] In H. Radder (ed) *The Philosophy of Scientific Experimentation*, University of Pittsburgh Press.
- Morrison, M. (2000) *Unifying Scientific Theories. Physical Concepts and Mathematical Structures*. Cambridge: Cambridge University Press.
- Morton, A. (1990) "Mathematical modeling and contrastive explanation." *Canadian Journal of Philosophy* Suppl. Vol. 16, 251-70.
- Morton, A. (1993) "Mathematical models: questions of trustworthiness." *British Journal for the Philosophy of Science* 44(4), 659-674.
- Muller, F. A. (2001) "Margaret Morrison, Critical discussion of Unifying Scientific Theories. Physical Concepts and Mathematical Structures." *Erkenntnis* 55, 132-143.
- Musgrave, A. (1981) "'Unreal assumptions' in economic theory: the F-twist untwisted." *Kyklos* 34, 377-387.
- Myrdal, G. (1957) *Economic Theory and Underdeveloped Regions*. London: Ruckworth.
- Ohlin, B. (1933) *Interregional and International Trade*. Cambridge, Mass: Harvard University Press.
- Oinas, P. (1995) "Roundtable on AnnaLee Saxenian's 'Regional Advantage'", *Economic Geography* 71 (2), 202-204.
- Oinas, P. (1998) *The embedded firm? Prelude for a revived geography of enterprise*, Helsinki: Acta Universitatis Oeconomicae Helsingiensis, A-143.

- Olsen, J. (2002) "On the units of geographical economics." *Geoforum* 33(2), 153-164.
- Ottaviano, G. (2003) "Regional policy in the global economy: insights from the new economic geography." *Regional Studies* 37 (6-7), 665-673.
- Ottaviano, G. and Puga, D. (1998) "Agglomeration in the Global Economy: A survey of the 'New Economic Geography'." *World Economy* 21(6), 707-731.
- Ottaviano, G. and Thisse, J-F. (2004) "New economic geography: what about the N?" *CORE Discussion Paper* No. 2004/65 <http://ssrn.com/abstract=660124>
- Phelps, N. A. and Ozawa T. (2003) "Contrasts in agglomeration: proto-industrial, industrial and post-industrial forms compared." *Progress in Human Geography* 27(5), 583-604.
- Piore, M. and Sabel, C. (1984) *The second industrial divide*. New York: Basic Books.
- Plummer, P. (2003) "Modelling economic landscapes: a geographical perspective." *Regional Studies* 37(6&7), 685-695.
- Plummer, P. and Sheppard, E. (2001) "Must emancipatory economic geography be qualitative." *Antipode* 33(2), 194-199.
- Porter, M. (1985) *Competitive Advantage: creating and Sustaining Superior Performance*. New York: The Free Press.
- Porter, M. (1990) *The Competitive Advantage of Nations*. London/Basingstoke: Macmillan.
- Porter, M. (1994) "The role of location in competition." *Journal of the Economics of Business* 1, 35-39.
- Porter, M. (1998) *On Competition*. Boston: Harvard Business Review.
- Porter, M. (2000) "Locations, clusters and company strategy." In G. Clark, M. Feldman and M. Gertler (Eds) *The Oxford Handbook of Economic Geography*. Oxford: Oxford University Press, 253-274.
- Psillos, S. (1999) *How Science Tracks Truth*. London: Routledge.

- Psillos, S. (2002) *Causation and Explanation*. Chesham: Acumen.
- Puga, D. and Venables, A. (1997) "The spread of industry: spatial agglomeration in economic development." *Journal of the Japanese and International Economies* 10 440-64.
- Putnam, R. (1993) *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton: Princeton University Press.
- Rodriguez-Pose, A. (2001) "Killing economic geography with a "cultural turn" overdose." *Antipode* 33(2), 176-182.
- Ruben, D-H. (1990) *Explaining Explanation*. London: Routledge.
- Salmon, W. (1984) *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.
- Salmon, W. (1989) *Four Decades of Scientific Explanation*. Minneapolis: University of Minnesota Press.
- Salmon, W. (1998) *Causality and Explanation*. Oxford: Oxford University Press.
- Saxenian, A. (1994) *Regional Advantage. Culture and Competition in Silicon Valley and Route 128*. Cambridge, MA: Harvard University Press.
- Sayer, A. (1985) "Realism and geography." In R.J. Johnston (ed) *The Future of Geography*. London and New York: Methuen & Co, 159-173.
- Sayer, A. (1992) *Method in Social Science. A Realist Approach*. New York: Routledge 2<sup>nd</sup> ed.
- Sayer, A. and Storper, M. (1997) "Ethics unbound: for a normative turn in social theory." *Environment and Planning D* 15, 1-17.
- Scitovsky, T. (1954) "Two Concepts of External Economies." *Journal of Political Economy* 62, 143-51.
- Scott, A. (2000) "Economic geography: the great half-century." *Cambridge Journal of Economics* 24, 483-504.
- Scott, A. (2004) "A perspective of economic geography." *Journal of Economic Geography* 4, 479-499.



- Sheppard, E. (2001) "How 'economists' think: about geography, for example." *Journal of Economic Geography* 1, 131-136.
- Sjöberg, O. and Sjöholm, F. (2002) "Common ground? Prospects for integrating the economic geography of geography and economists." *Environment and Planning A* 34, 467-86.
- Skipper, R.A. (1999) "Selection and and the extent of explanatory unification." *Philosophy of Science* 66 (Proceedings), S196-S209.
- Smith, N. (1987) "Dangers of the empirical turn: some comments on the CURS initiative." *Antipode* 19, 59-68.
- Solomon, M. (1994) "Social Empiricism." *Noûs* 28, 325-43.
- Solomon, M. (2001) *Social Empiricism*. Cambridge, Mass: MIT Press.
- Starrett, D. (1978) "Market allocations of location choice in a model with free mobility." *Journal of Economic Theory* 17, 21-37.
- Storper, M. (1995) "The resurgence of regional economies ten years later: the region as a nexus of untraded interdependencies." *European Urban and Regional Studies* 2, 191-221.
- Storper, M. (1997) *The Regional World: Territorial Development in a Global Economy*. New York: Guilford.
- Sunley, P. (2000) "Urban and regional growth". In E. Sheppard and T.J. Barnes (eds) *A Companion to Economic Geography*. Oxford: Blackwell, 187-201.
- Sunley, P. (2001) "What's behind the models. A review of The Spatial Economy." *Journal of Economic Geography* 1, 136-139.
- Teller, P. (2004) "How we dapple the world." *Philosophy of Science* 71, 425-447.
- The Economist* 1999, "Knowing your place." 13 March, 92.
- Urban, D.M. (2001) "The Spatial Economy: one new economic geographer's view." *Journal of Economic Geography* 1, 146-152.

- Van Bouwel, J. and E. Weber (2002) "Remote causes, bad explanations?" *Journal for the Theory of Social Behaviour* 32(4), 436-448.
- Venables, A. (1996a) "Equilibrium locations of vertically linked industries." *International Economic Review* 37, 341-59.
- Venables A. (1996b) "Localisation of industry and trade performance." *Oxford Economic Policy Review* 12(3), 52-60.
- Waldorf, B. S. (2004) "Path-breaking books in regional science." *Papers in Regional Science* 83(1), 59-89.
- Weber, E. (2002) "Can we dispense with structural explanations of social facts?" *Economics & Philosophy* 18, 259-275.
- Whewell, W. (1847) *Philosophy of the Inductive Sciences*. London: John W. Parker.
- Woodward, J. (2000) "Explanation and invariance in the special sciences." *British Journal for the Philosophy of Science* 51, 197-254.
- Woodward, J. (2003) *Making Things Happen*. New York: Oxford University Press.
- Ylikoski, P. (2001) *Understanding Interests and Causal Explanation*. Academic Dissertation, Department of Moral and Social Philosophy, University of Helsinki.

## Appendix

### The core model of GE and an extension

We first present the core-periphery model of GE from which the conditions for the existence and sustainability of a core-periphery structure in the economy are derived (Fujita, Krugman and Venables, 1999). Second, we report one of the extensions of the core model (Fujita, Krugman and Venables, 1999).

### The core-periphery model

There are two regions, 1 and 2, which are symmetric in all respects. The demand side is modelled as follows.

$$(2.1) \quad U = M^\mu A^{1-\mu}$$

$$(2.2) \quad M = \left[ \int_0^n m(i)^\rho di \right]^{1/\rho}$$

(2.1) represents a Cobb-Douglas utility function over agricultural and manufactured goods for the typical individual, where  $A$  is consumption of the agricultural good and  $M$  is a composite index of the consumption of manufactured good, and  $\mu$  is the expenditure share of manufactured goods. As it is characteristic of the Dixit-Stiglitz model, (2.2) defines  $M$  as a CES function (constant elasticity of substitution), where  $\rho$  represents the intensity of the preference for variety in

manufactured goods and  $0 < \rho < 1$ ;  $m(i)$  is the consumption of each available variety;  $n$  is the range of varieties produced.

The range of manufactures offered is endogenous. This implies that increasing the range of varieties reduce prices and therefore the cost of attaining a given level of utility (love-of-variety effect).

The supply side is modelled as follows. There are two sectors in the economy, agriculture and manufacturing, and two factors of production, each assumed to be specific to one sector. The agricultural sector is perfectly competitive and produces a homogenous good under constant returns to scale. In this model, agricultural workers are immobile.

The manufacturing sector is a Dixit-Stiglitz monopolistic sector, characterized by increasing returns to scale. The technology function, which expresses internal economies of scale, for a manufacturing good  $i$  is given by

$$(2.3) \quad l^M = F + c^M q^M$$

where  $l$  is the cost of producing a given amount,  $q$ , of good  $i$ .  $F$  is a fixed cost and  $c^M q^M$  is the marginal labour input requirement.

There is a large number of firms, each producing a single variety. For all varieties produced at location  $r$ , the pricing rule is as follows:

$$(2.4) \quad p_r^M = c^M w_r^M / \rho$$

where  $w_r^M$  is the wage rate. The zero-profit condition implies that the equilibrium output for an individual firm is:

$$(2.5) \quad q^* \equiv F(\sigma - 1) / c^M$$

where  $\sigma = 1 / (1 - \rho)$ .

It is assumed that transporting agricultural goods is costless. Transport costs for manufactured goods are assumed to be of the "iceberg form" meaning that only a fraction  $1/T_r^M$  of the good arrives at destination, where  $T$  represents the amount of

the good dispatched per unit received. That is, when there are no transportation costs  $T = 1$ .

Workers are assumed to move in response to real wage differentials, that is, they move towards the regions that offer higher real wages. Define the average real wage as  $\bar{\omega} = \sum_r \lambda_r \omega_r$ .

The ad hoc dynamics is then given by

$$(2.6) \quad \dot{\lambda}_r = \gamma(\omega - \bar{\omega})\lambda_r$$

The equilibrium of the model for a two-regions case is given by the following system of equations.

$$(2.7) \quad G_1 = [\lambda w_1^{1-\sigma} + (1-\lambda)(w_2 T)^{1-\sigma}]^{1/(1-\sigma)}$$

$$(2.8) \quad G_2 = [\lambda (w_1 T)^{1-\sigma} + (1-\lambda)w_2^{1-\sigma}]^{1/(1-\sigma)}$$

$$(2.9) \quad w_1 = [Y_1 G_1^{\sigma-1} + Y_2 G_2^{\sigma-1} T^{1-\sigma}]^{1/\sigma}$$

$$(2.10) \quad w_2 = [Y_1 G_1^{\sigma-1} T^{1-\sigma} + Y_2 G_2^{\sigma-1}]^{1/\sigma}$$

$$(2.11) \quad Y_1 = \mu \lambda w_1 + \frac{1-\mu}{2}$$

$$(2.12) \quad Y_2 = \mu(1-\lambda)w_2 + \frac{1-\mu}{2}$$

$$(2.13) \quad \omega_1 = w_1 G_1^{-\mu}$$

$$(2.14) \quad \omega_2 = w_2 G_2^{-\mu}$$

(2.7) and (2.8) give the normalized price index equations, where  $\lambda$  is the share of manufacturing in each region. (2.9) and (2.10) are the normalized wage equations. (2.11) and (2.12) give income in the two regions and finally, (2.13) and (2.14) are the real wage equations. Aided by numerical examples, these eight non-linear equations, which are not very tractable, yield analytical results.

Suppose that all manufacturing is concentrated in one region,  $\lambda = 1$ . To determine whether this constitutes an equilibrium, we ask whether real wage in region 1 is higher

than in region 2. If it is, then the core-periphery pattern is sustainable.

The wage is guessed as being equal to 1 and it turns out to be the case that this is indeed an equilibrium value. In such equilibrium, the expression for the sustain point is:

$$(2.15) \quad \omega_2 = T^{-\mu} \left[ \frac{1+\mu}{2} T^{1-\sigma} + \frac{1-\mu}{2} T^{\sigma-1} \right]^{1/\sigma}$$

When there are no transportation costs, then  $\omega_2 = 1$  and hence location is irrelevant. If transportation costs increase by a small amount, then agglomeration becomes sustainable ( $\omega_2 < 1 = \omega_1$ ). For very high transportation costs, agglomeration is not sustainable. In (2.15)  $T^{-\mu}$  represents the forward linkage: location 2 is relatively expensive and therefore unattractive for manufacturing workers because manufacturing goods have to be imported. The second term of (2.15) represents the nominal wage at which a firm that were to locate in 2 would make zero profits. It says that such a firm would do well in supplying the smaller market (location 2) but badly in supplying the larger one (location 1) and that there is a backward linkage working through demand.

To find the *break point*, we assume a situation of symmetric equilibrium. In order to obtain it, the equilibrium is totally differentiated with respect to  $\lambda$ . Then we check the equilibrium response to  $d(\omega_1 - \omega_2)/d\lambda$  to obtain the change in real wages in response to a movement of workers:

$$(2.16) \quad \frac{d\omega}{d\lambda} = 2ZG^{-\mu} \left( \frac{1-\rho}{\rho} \right) \left[ \frac{\mu(1+\rho) - Z(\mu^2 + \rho)}{1 - \mu Z(1-\rho) - \rho Z^2} \right]$$

where  $Z$  ( $0 \leq Z \leq 1$ ) represents an index of trade cost. If trade has no cost, then  $Z = 0$ ; if trade is not possible, then  $Z = 1$ . If  $d\omega/d\lambda < 0$ , then the symmetric equilibrium is stable, if  $d\omega/d\lambda > 0$ , then the symmetric equilibrium is unstable. For low transportation costs, ( $Z$  close to 0), the symmetric equilibrium is

unstable, and it is stable for sufficiently high levels of transport costs.

### Extension: International Specialization

An extension of the core model derives patterns of international specialization by introducing intermediates for production. In this model locations are intended as countries. Labour is assumed to be immobile between locations, but mobile between sectors. The backward and forward linkages now arise from the use of intermediates. The input for production is assumed to be a composite of labour and intermediates. Assume that the price of intermediates in location  $r$  is  $G_r$  and that the input composite is a Cobb-Douglas function of labor and intermediates with the share of intermediate equal to  $\alpha$ .

The price of the input is defined by  $w_r^{1-\alpha} G_r^\alpha$  and the pricing rule by  $p_r = w_r^{1-\alpha} G_r^\alpha$ . The intermediate is assumed to be a CES function of the available varieties. The price indices for each country in this model become,

$$(2.17) \quad G_1^{1-\sigma} = \lambda_1 w_1^{1-\sigma(1-\alpha)} G_1^{-\alpha\sigma} + \lambda_2 w_2^{1-\sigma(1-\alpha)} G_2^{-\alpha\sigma} T^{1-\sigma}$$

$$(2.18) \quad G_2^{1-\sigma} = \lambda_1 w_1^{1-\sigma(1-\alpha)} G_1^{-\alpha\sigma} T^{1-\sigma} + \lambda_2 w_2^{1-\sigma(1-\alpha)} G_2^{-\alpha\sigma}$$

And the wage equations become,

$$(2.19) \quad \frac{(w_1^{1-\alpha} G_1^\alpha)^\sigma}{1-\alpha} = E_1 G_1^{\sigma-1} + E_2 G_2^{\sigma-1} T^{1-\sigma}$$

$$(2.20) \quad \frac{(w_2^{1-\alpha} G_2^\alpha)^\sigma}{1-\alpha} = E_1 G_1^{\sigma-1} T^{1-\sigma} + E_2 G_2^{\sigma-1}$$

The sustain condition is identical to the one reported for the core-periphery model except for  $\mu$  is replaced by  $\alpha$ .

$$(2.21) \quad w_2^{1-\alpha} = T^{-\alpha} \left[ \frac{1+\alpha}{2} T^{1-\sigma} + \frac{1-\alpha}{2} T^{\sigma-1} \right]^{1/\sigma}$$

Similarly, the break point is given by the following expression:

$$(2.22) \quad \frac{d\omega}{d\lambda} = \frac{-Z}{\mu\Delta} \left[ \frac{\alpha(1+\rho) - Z(\alpha^2 + \rho)}{1-\rho} \right]$$



## Tables and figures

Table 4.1 . Ways in which explanations relate

<i>What is the same</i>	<i>What differs</i>	<i>Explanatory rivalry</i>	<i>Explanatory complementarity</i>
CASE I M	S (e.g., S <sub>1</sub> and S <sub>2</sub> )	-S <sub>1</sub> and S <sub>2</sub> are incompatible	-S <sub>1</sub> and S <sub>2</sub> jointly cause M -S <sub>1</sub> and S <sub>2</sub> are causally related (S <sub>1</sub> causes S <sub>2</sub> and S <sub>2</sub> causes M) -S <sub>1</sub> causes M under conditions C <sub>1</sub> , and S <sub>2</sub> causes M under conditions C <sub>2</sub> -S <sub>1</sub> and S <sub>2</sub> are otherwise related
CASE II S	M (e.g., M <sub>1</sub> and M <sub>2</sub> )	-M <sub>1</sub> and M <sub>2</sub> are incompatible	-S explains M <sub>1</sub> and M <sub>2</sub> -S causes M <sub>1</sub> under conditions C <sub>1</sub> , and S causes M <sub>2</sub> under conditions C <sub>2</sub> -M <sub>1</sub> and M <sub>2</sub> are causally related -M <sub>1</sub> and M <sub>2</sub> are otherwise related
CASE III -	S and M (e.g., S <sub>1</sub> , S <sub>2</sub> , M <sub>1</sub> , M <sub>2</sub> )	-M <sub>1</sub> is incompatible with S <sub>2</sub> -M <sub>2</sub> is incompatible with S <sub>1</sub> -S <sub>1</sub> and S <sub>2</sub> are incompatible -M <sub>1</sub> and M <sub>2</sub> are incompatible	-M <sub>1</sub> and M <sub>2</sub> are causally related -M <sub>1</sub> and M <sub>2</sub> are otherwise related -S <sub>1</sub> and S <sub>2</sub> are causally related -S <sub>1</sub> and S <sub>2</sub> are otherwise related -Both of the above -S <sub>1</sub> and S <sub>2</sub> jointly cause M <sub>1,2</sub>

Table 4.2. Relations of the various cluster explanations

**CASE I**

<i>Explanations of competitiveness</i>	
SAME M	DIFFERENT S
$[CM_{Sax}^G] = [CM_{M\&M}^3]$ Continued <i>[declining]</i> competitiveness of firms in clusters	$[S_{Sax}^G]$ Those cultural features facilitating adaptability $[S_{M\&M}^3]$ Flexible institutional fit
<b>RIVALRY AND COMPLEMENTARITY</b>	
Very similar S (Apparent causal relation: $[S_{Sax}^G] \rightarrow [S_{M\&M}^3]$ )	

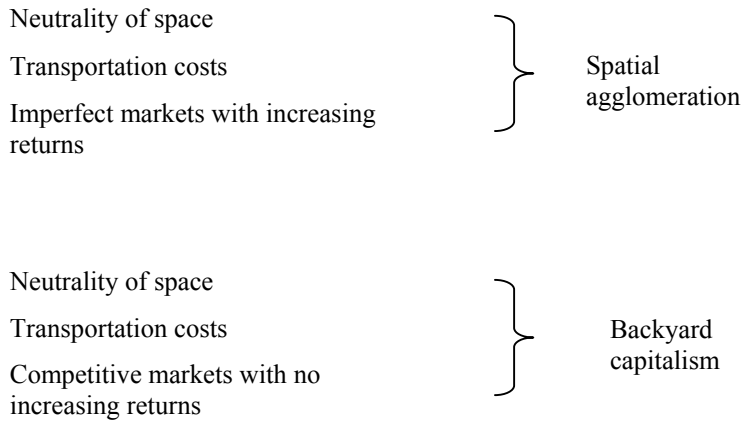
**CASE III**

<i>Explanations of competitiveness</i>	
DIFFERENT M	DIFFERENT S
$[CM_{M\&M}^{2A}]$ The competitiveness of clustered firms <i>[single uni-locational firms]</i>	$[S_{M\&M}^{2a}]$ Efficiencies in knowledge creation among <i>horizontally</i> related firms $[S_{M\&M}^{2d}]$ Shared institutional environment
$[CM_{M\&M}^{2B}]$ The competitiveness of clustered firms <i>[dispersed firms]</i>	----- $[S_{M\&M}^{2b}]$ Efficiencies in knowledge creation among <i>vertically</i> related firms $[S_{M\&M}^{2c}]$ Cost reductions $[S_{M\&M}^{2d}]$ Shared institutional environment
$[CM_p]$ The competitiveness of clustered firms <i>[firms otherwise organized]</i>	$[S_p^1]$ Features of clusters that increase (static) productivity $[S_p^2]$ Features of clusters that enhance firms' capacity to innovate and thus productivity growth $[S_p^3]$ New business formation $[S_p^4]$ Favourable social structure

Table 4.2 (continues from previous page)

RIVALRY AND COMPLEMENTARITY	
(i) $[CM_{M\&M}^{2A}] + [CM_{M\&M}^{2B}] \sim [CM_P]$ (ii) $[S_{M\&M}^{2A}] + [S_{M\&M}^{2B}] \sim [S^1_P] + [S^2_P]$ (iii) $[S_{M\&M}^{2d}] \sim [S^4_P]$ (iv) M&M yield reconstructions of more precise CM'a and their related S'a (v) Partial incompatibility due to the explanatory primacy M&M ascribes to $[CM_{M\&M}^{2A}]$	
CASE III	
DIFFERENT M	DIFFERENT S
<i>Explanations of existence</i>	
$[CM_{GE}]$ Existence of clusters [ <i>spatial dispersion</i> ] at all times and at all scales	$[S_{GE}]$ The agglomerating mechanism of pecuniary externalities
$[CM_{M\&M}^1]$ Existence of local clusters [ <i>other forms of spatial organisation</i> ] in the knowledge-based economy	$[S_{M\&M}^1]$ The competitiveness of clustered firms due to $[S_{M\&M}^{2a}]$ , $[S_{M\&M}^{2b}]$ , $[S_{M\&M}^{2c}]$ , $[S_{M\&M}^{2d}]$
RIVALRY AND COMPLEMENTARITY	
Complementarity: different meanings of 'existence'	

Figure 5.1 Spatial agglomeration versus backyard capitalism



## Samenvatting

Geografische economie is een nieuwe benadering binnen de economie die erop gericht is om een geunificeerd raamwerk te ontwikkelen voor de bestudering van de ruimtelijke concentratie van economische activiteiten. Dit streven naar eenheid of unificatie staat in scherp contrast met de huidige verscheidenheid van benaderingen die erop gericht zijn ruimtelijke agglomeratie te verklaren. De verschillende benaderingen zijn voorgesteld in of op het grensvlak tussen disciplines zoals economie, economische geografie en *business strategy*. Voorstanders van de verschillende theorieën hebben hun kritiek vooral gericht op geografische economie. Dit proefschrift onderzoekt de verscheidenheid aan theorieën en richt zich daarbij met name op de rol van geografische economie. Het bevat vier onafhankelijke essays die eenheid, verscheidenheid en verklaring in relatie tot geografische economie en haar burens betreffen. Deze worden voorafgegaan door een vijfde essay dat de gemeenschappelijke filosofische premissen expliciet maakt, het overkoepelende filosofische raamwerk schetst en de belangrijkste conclusies presenteert.

Het eerste hoofdstuk introduceert een pluralisme binnen een raamwerk van realisme en laat zien hoe een dergelijke benadering inzicht kan geven in het feit dat theorieën slechts deelaspecten belichten en dat ook slechts bij benadering en verheldert de samenhang met een contrastieve benadering van

verklaring. Verder bespreekt dit essay de belangrijkste conclusies van dit proefschrift zowel voor geografische economie (GE) en haar burens als voor de filosofie van de sociale wetenschappen in het algemeen en de filosofie van de economie in het bijzonder.

Het tweede hoofdstuk laat zien dat het streven naar eenheid of unificatie een belangrijke factor heeft gespeeld in de ontwikkeling van GE. GE claimt te hebben laten zien dat ruimtelijke agglomeratie-effecten op verschillende ruimtelijke niveaus voortkomen uit dezelfde basale economische mechanismen. Hiermee zou ze een geunificeerde benadering van deze fenomenen hebben ontwikkeld. De gedetailleerde analyse die hier wordt gepresenteerd verheldert zowel de rol die unificatie recentelijk heeft gespeeld binnen GE als de belangrijkste filosofische benaderingen van unificatie. Ze laat met name zien dat geen van die filosofische benaderingen volledig recht kan doen aan de rol die unificatie binnen GE heeft gespeeld.

Het derde hoofdstuk betreft het meest verhitte theoretische debat in het vakgebied, te weten dat tussen economische geografen en geografische economen. Door mogelijke misverstanden op te helderen maakt dit essay duidelijk waar het debat echt over gaat. Daarmee laat het zien dat de twee benaderingen zowel op methodologische als inhoudelijke gronden complementair zouden kunnen zijn.

Het vierde hoofdstuk analyseert intertheoretische relaties tussen verklaringen van ruimtelijke clustering (dat wil zeggen, de geografische concentratie van bedrijven die met elkaar samenwerken of concurreren). Met dit doel worden de explananda in contrastieve termen geherformuleerd en wordt beargumenteerd dat de verklaringen niet als concurrenten beschouwd hoeven te worden: ze richten zich op het beantwoorden van verschillende vragen over en benadrukken

uiteenlopende aspecten van het fenomeen van ruimtelijke clustering.

Het vijfde hoofdstuk beargumenteert dat de contrastieve benadering goed te gebruiken is om geschillen over onrealistische modellen in de economische wetenschap op te lossen. Het laat met name zien dat de contrastieve benadering van verklaren het nut van bepaalde onrealistische veronderstellingen van theoretische modellen kan verhelderen. Ter illustratie worden de kritieken op GE dat ze onrealistisch zou zijn besproken wat resulteert in een gekwalificeerde verdediging van een aantal van de onrealistische elementen van de relevante modellen. De contrastieve benadering wordt met name gebruikt om de vraag die GE probeert te beantwoorden en daarmee de verklaring die ze nastreeft te specificeren om zo de abstracties of, beter gezegd, de isoleringen die ze maakt te evalueren.

## **Curriculum vitae**

Caterina Marchionni was born in Padova, Italy, on December 17 1974. She received her B.Sc. and MA degrees in Political Science, branch Political Economy, from the University of Padua, Italy, in April 2000. She obtained an M.Phil. degree at the Erasmus Institute for Philosophy and Economics, Erasmus University of Rotterdam, in June 2001. She completed her Ph.D. at the same institute.



