

# Which Duality? Why Dual?<sup>1</sup>

## On the Gains and Losses of Hungarian Sociology's Dual-society Theories

Márk Áron ÉBER

ebermark@tat.k.elte.hu

**ABSTRACT:** Hungarian sociology is abundant in dual-society and dual-structure theories. In most cases the mentioned duality is not the same. In my paper – to answer the question of which duality? – I compile a catalogue of Hungarian sociology's dual-structure and -society theories. As a follow-up step – this time while trying to answer the Why dual? question – I associate Hungarian sociology's dual-society images with Hungarian society's East-Central European semi-peripheral status. I propose that the semi-periphery of the capitalist world-system as a historically formed mode of existence is not evadable in its effects and lays the tendencies out for the cognition of this mode of existence. As a structural constraint it conditions Hungarian sociology to "recognize" the dualities of the social-structure. Finally, I argue for going beyond these duplicating strategies.

Hungarian sociology is abundant in dual-society and dual-structure theories. The image of a "divided", "split in half" society is just as widespread in domestic public-political discourse. If the perception of duality in society and social-structure has proved to be so tenacious, then it deserves a critical revision.

In the case of sociological concepts, the question is likely to arise: wherein *exactly* does the postulated duality lie? This is where difficulties emerge. A more detailed analysis makes it clear that in most cases the mentioned duality is not the same. If we conclude that the alleged dual structure of our society could as well mean – in reality – *several different* dualities, then we may ask: *which duality* is it exactly?

The next question would be whether the idea of a dual society is maintainable if it may mean multiple separate dualities? What relations are revealed or obscured by such distinctions? From where do these originate? Where are the roots? Is it the subject examined that is dual (if so, in what way?) or is the duality in the eye of the contemplator? If in fact it is the latter, then what exactly causes this duality? Is the intention to align with historical conventions (Ferenc Erdei's model)? Or is it some *a priori* tendency of cognition or modelling of society?

<sup>1</sup> I am grateful for the helpful comments and advice of many people on the earlier conference-lecture version of this paper. I would like to specifically thank Róbert Angelusz, Domonkos Sik for their adjoined critiques of the lecture script. I would also like to thank Anna Wessely, Dénes Némédi, Péter Somlai, Róbert Tardos, Erzsébet Szalai, Ottó Gecser, Vera Szabari, Léna Pellandini-Simányi and the two anonymous reviewers of the *Szociológiai Szemle* for their remarks and suggestions, which have helped to deepen my paper's train of thought and strengthen my arguments. Translation by Márton Rakovics, English revision by Borka Richter.

In my paper – to answer the question of *which duality?* – I compile a catalogue of Hungarian sociology’s dual-structure and -society theories. After analysing Ferenc Erdei’s theory on the social structure of the interwar period – perceived (falsely) as a dual-society theory – I present newer theories from both the pre- and post-socialist era, while taking a look at the sociological basis of the popular discourse of the bipolar political division thesis. This detour offers surprising further gains: to some extent it helps our understanding of the prevalence and popularity of Erdei’s concept.

As a follow-up step – this time while trying to answer the *Why dual?* question – I associate Hungarian sociology’s dual-society images with Hungarian society’s East-Central European semi-peripheral status. I propose that the semi-periphery of the capitalist world-system as a historically formed mode of existence is not evadable in its effects and lays the tendencies out for the cognition of this mode of existence. I also claim that the East-Central European semi-peripheral mode of existence as a structural constraint conditions Hungarian sociology to “recognize” the dualities of the social-structure. It forces Hungarian sociology to dissolve the empirical deviations of Hungarian society from the “model-society” with a society-duplicating cognitive strategy. Finally, I argue for going beyond these duplicating strategies.

## Which duality? From Erdei to the second economy

Erdei Ferenc’s work on the Hungarian society of the interwar period is probably the most important prototype and reference for sociological dual-society theories. Despite the strong temptation, it cannot be left out of consideration that the duality of the historico-national and the modern-bourgeois society does not encompass the whole of society. *Nowhere does Erdei actually write about a “dual” structure or society.* What he does mention, however, are the following: “multiple complex structures”, “ensemble” of multiple structures, joint but “discrete structures”, “self-sufficient social formations next to each other”, “labyrinthine social-structure”, “endless line of transient social forms”, “diverse forms”, “different social and cultural forms” (see: Erdei 1976a: 25). Erdei unambiguously implies that *under* (and partly *beside*) the alleged dual social-structure there is a third one, the peasant-society alias historic-folk society.<sup>2</sup> If we only emphasize the upper two structures, a *categorical* distinction and *opposition* is questionable.

2 To quote Erdei: “*The upper most and ruling social-structure is the historico-national society that evolved and modernized as the continuation of the feudal noble society. This assumes the same positions of production and the same roles as it did in the feudal society, only now adjusting to the production mode of capitalism. Underneath lies the social-structure of the somewhat modernized historic-folk society as a continuation of the villeinage of the feudal society and the folk society below all states – all more or less adopting the production mode of capitalism. Beside these historical structures, capitalism’s own structure formed in Hungarian society’s modern-bourgeois society, which is just one component of the era’s Hungarian society in the roles and production positions of industry, the trading of goods and modern city-intellectual life*” (Erdei 1976a: 25 – emphasis added). The 2010/4 issue of *Szociológiai Szemle* publishes the fourth part of Erdei’s work on historic-folk society – found by Károly Halmos (Erdei 2010; Halmos 2010).

The rather prolific Erdei wrote this paper around 1943–44, but he himself never published it. The script that had turned up from the bequest was only published five years after Erdei's death by Tibor Huszár in 1976 (Huszár 1976, 2003). Why Erdei did not publish his paper remains unknown up to this day. There are several interpretations.<sup>3</sup> According to one of the typical opinions, Erdei stood by his concept, but thought the script to be lost (see for example Huszár 1976). According to another standpoint, he no longer took responsibility for the concept (opinions vary as to why). As for the empirical validity of the manuscript, interpreters can be split into two opposing camps.<sup>4</sup> According to the one, the script is an authentic and valid depiction of interwar society. The other view – which has a different opinion also on why the paper remained a manuscript – states that the author's political ambitions and the morale of the early forties left its mark on the 1943–44 paper. Accordingly, Erdei sketched the image of the dual-structure through the genteel–non genteel, gentry–bourgeois, Christian–Jewish dichotomies. These distinctions – though nonetheless very significant in the era – would *not necessarily*, in a professional, scientific study, justify the *structural* duplication of society (Nagy 1993 [1986]; Gyáni 1997, 2001; Bognár 2011 [2003], 2010).

Although György Konrád and Iván Szelényi refer to it in their 1971 paper (Konrád–Szelényi 1974 [1971]), 'Social conflicts of under-urbanization', in post-war domestic social-science the first who explicitly mentions the "dual-nature of social-structure" is Tamás Kolosi (Kolosi 1974: 155). However, the rough distinction of Konrád and Szelényi, later conceptualized by Kolosi, does not resemble Erdei's concept. No wonder: Erdei's idea could not have inspired either Konrád and Szelényi or Kolosi, since at that time Erdei's concept had not yet been discovered. Kolosi's observation concerning the dual-structure has nothing to do with national- and bourgeois-societies. His writings arose out of the reform processes and political-economic measures of the Sixties, in particular the economic and social structural developments resulting from the introduction of the 'New Economic Mechanism'. He points out, that *beside* the redistributive mechanism – disposing over excess goods *produced socially* –, economical reforms created and strengthened another sphere of production, in which goods produced *outside* socially organized production are distributed (Kolosi 1975: 155–163). These phenomena, later labelled "second economy" by István R. Gábor and Péter Galasi<sup>5</sup> was then further developed by Elemér Hankiss in the early eighties – albeit still with quotation and question marks – as "second society". In the late eighties, from the pen

3 In the interpretation of Erdei's dual society concept, I rely mainly on the following works: Huszár 1976, 2003; Némédi 1978; Nagy 1993 [1986]; Kovács–Melegh 1997; Kövér 2001 [1998]: 22–24; Gyáni 2001; Romsics 2006; Bognár 2011 [2003], 2010; Halmos 2010.

4 This disagreement was partially the intellectual stake of the dispute between Gábor Gyáni and Viktor Karády in *Budapesti Könyvszemle (Bukasz)* (Gyáni 1997, 1998; Karády 1998).

5 Gábor–Galasi 1981; cf. Kolosi (2003). For the overview of the vast literature on second economy, see Sik 1996. The division of economy to primary and secondary (shadow) was later attached to the formal–informal distinction, which came from the fallacy that – in János Kornai's (1983) terms – the bureaucratic coordinated primary economy is formal and the quasi-market coordinated secondary economy is informal. For a detailed critique of this view, see József Böröcz's works (Böröcz 1990, 1993, 2000; Böröcz–Southworth 1998).

of Szelényi and his associates, as well as Kolosi, come the double-triangle and L-models – canonized in sociology curricula for more than two decades – that were to describe the dual social-structure of late state socialism (Szelényi–Manchin 1987; Szelényi 1986–87, Szelényi 1988; Kolosi 1987: 89–120; cf. Andorka 2006: 185–186).

Although Tamás Kolosi and Iván Szelényi refer to and depend on Erdei’s model in the eighties, *the parallels* in the social structure of the interwar period and that of the eighties the differences are *much more apparent*. The first economy’s redistributive and the second economy’s quasi-market sectors can scarcely be deduced from the national- and bourgeois-societies. With his thesis of ‘interrupted embourgeoisement’ Szelényi does not state that (in the individual history and genealogy) Erdei’s bourgeois society reappears in the form of the socialist entrepreneur; he rather mentions the embourgeoisement of Erdei’s peasant-society (Szelényi 1988). Be that as it, the alleged dualities of the interwar society and that of the eighties are nowhere near the same duality.

The research of the seventies and eighties errs not only by accepting Erdei’s analysis with little reflection on it<sup>6</sup>, but it also misunderstands Erdei by consequently perceiving his depiction of society as the theory of dual-structure (Kolosi 1987: 90–92; Szelényi 1988: 67–70; cf.: Andorka 2006: 176).<sup>7</sup> Iván Szelényi interprets the developments of the seventies and eighties as a ‘return’ to the ‘normal’ state of the interwar period. As he puts it:

...the most striking development is the *resurgence* of the ‘second hierarchy’ [...]. Social structure seems to be returning to its ‘normal’ state: the second, market-based, burgher hierarchy which was *temporarily* forced into the dominant rank order *regains* its relative autonomy, although it remains more subordinated than it was before 1945 (Szelényi 1988: 71 – emphasis added).

While, among others, Tamás Kolosi and Iván Szelényi date the formation of Hungarian society’s dual-structure and the strengthening of the second economy from the sixties, István Kemény, representing the older researcher generation, sees a *completely opposite* tendency. According to him, “only with the economic reforms of sixty-three did the *merger* of the two structures start, because the rigid partition wall between the two Hungaries – the communist and the other – *fell*” (Kemény 1992a: 291 – emphasis added). István Kemény states that “...neither Rákosi’s terror, nor the Kádár-system’s retaliations could break the antecedent social-structure. [...] They only built their structures on it” (Kemény 1992a: 291). In his observation “...there are two structures beside each other: a purely communist

6 Tamás Kolosi’s interpretation of Erdei is the least critical (Kolosi 1987: 90–92). Iván Szelényi has a more subtle view on Erdei’s oeuvre (Szelényi 1988: 58–63), but is just as accepting of his conclusions as Kolosi.

7 See footnote 2.

structure that is logically built or rather copied, and a more complex and subtle one that developed over centuries. And these two societies coexisted” (Kemény 1992a: 291).<sup>8</sup> While according to others, society duplicated because of the economic reforms of the sixties, Kemény claims that the strengthening of the second economy is what makes a merger possible. *Erdei, Kemény, Szelényi (and Kolosi) all talk of a dual-society – all in a different sense.*

## Duplicating views of society after the system change

The question arises, whether the duality of Hungarian social-structure in the distinction of redistribution and (quasi)market is valid not only before the system change or also after it? In other words: is this kind of social-structure only possible in the case of socialist economies or also of in capitalist formations?

In a thought experiment, Iván Szelényi claimed that *capitalist* systems that are linked primarily to market-integrated welfare state can also create dual economic structures if beside the capital-based mode of production they also establish a state-dominated one (Szelényi 1981, 1990: 449–469., cf. Szelényi 1978; Szelényi–Manchin 1987). It is important to stress that Szelényi never explicitly states that a dual-society can be seen in capitalist systems, unlike Tamás Kolosi, who claims that there is a ‘high probability’ for the dual-structure scheme to be true for the post-system-change social-structure.<sup>9</sup> It is worth quoting his explanation precisely:

...in Hungary – and to a lesser extent in other Eastern European post-socialist countries – a distinctive duality of a redistributive and a market type social structure already came about in the socialist era, becoming part of common knowledge in social sciences under the name of the L-model and the ‘dual pyramid theory’. It is highly plausible that the same duality describes modern capitalism, *because*, in opposition to the classical capitalism of the nineteenth century, in modern capitalism redistribution has an indisputable structure forming role (Kolosi–Sági 1996: 159, Kolosi 2000: 33 – emphasis added).

According to him then, the fact that social structure is generally dual in modern capitalism and specifically post-system-change Hungarian society is *the result of the effects of two* structuring mechanisms (market and redistribution).<sup>10</sup> In Tamás

8 István Kemény takes his view partly from Gyula Tellér. He thinks that it was Tellér in the seventies, who – together with Ferenc Donáth, Zoltán Zsille and Pál Juhász – can be considered the first to formulate the concept of dual economy and society (Kemény 1992b, 2010: 130–132, 154–155).

9 To quote Kolosi, from a 1991 interview: “My – maybe a bit opinionated – assumption is that on the level of large theoretical models, what I have written on social structure in Hungary will be true for a long time. [...] Nothing has changed in the fundamental structural processes in Hungary; the dual social-structure model – with the particular mixture of redistributive and market mechanisms – is absolutely virulent today, meaning that it is still applicable” (Kolosi 1991: 111).

10 If we build on Karl Polányi’s economic-integration mechanisms (Polányi 1957), then why is the third scheme, reciprocity, consistently being disregarded? – asks József Böröcz (1990, 2000). The simplifying opposition of redistribution versus market, bureaucratic versus market coordination generously ignores the varying forms of reciprocity, like a clan and client-system, favoritism, nepotism, fixing, swindling, corruption and other relations based on mutual assistance (see for example: Czákó–Sik

Kolosi's view the L-model is continues to be valid – adding that the ratio of the two structuring mechanisms has shifted: since 1990 the role of state redistribution has become secondary to market integration (Kolosi-Sági 1996: 159–160; Kolosi 2000: 33–48). The fact that the analysis of the post-system-change social structure is done with the conceptual tool of market elite versus redistributive elite is very reflective of the idea (Kolosi-Sági 1996; Kolosi 2000).

However, other dualities appear in the sociological interpretation of the effects of the system change: analysts talk of sliders and risers in mobility and of “winners” and “losers” of the change (cf.: ex. Ferge 1996; Kolosi-Sági 1996: 155–157, 185–192; Kolosi 2000: 165–169; Habich-Spéder 1998). The ethical and political value judgement behind the “duplicating views of society” and the adequate nature of this approach has sparked a debate in the profession (cf.: Zentai 1999). Júlia Szalai's work on the social-political institutional structure also fits this discourse (Szalai 2007). In her book, titled *Aren't there two countries...? (Nincs két ország...?)* she foreshadows an affirmative answer. According to her, the rupture between the two societies runs between the poor (more precisely the “indigent”) and non-poor in the struggle over redistributive sources. The parallel enforcement of the indigence- and contribution-principal in state redistribution creates and supports a dual structure. Whoever finds work in the labour-market – according to the contribution-principal – is granted a broader access of redistributive sources, contrary to those who are excluded from the labour-market and can only access (scarce) benefits on the grounds of indigence.

Erzsébet Szalai sees a duality in post-system-change in Hungary that is unlike anything else (Szalai 2001). As she states: “A Western, market-connected, foreign-dominated, concentrated ownership and institutional structure is up against a domestic-market-dependent, diffuse structure with a feudal and paternalistic production-management model and a lifestyle based on it” (Szalai 2001: 240, see also: Szalai-Krausz-Szigeti 2002). This opposition is referred to, in brief, as multi-national (or Western) and domestic (feudal) spheres. Erzsébet Szalai proposes her dual-society concept by the sweeping generalization of the foreign- and domestic-owned “dual economy” concept (cf.: Bartha 2003).

An approach intersecting all others is that of Béla Pokol. He does not set forth his concept by the opposition in Hungarian or East-Central European societies' dual structure, but by the opposition of two types of social-structure theories. Just as Ralf Dahrendorf differentiates between integration- and coercion-theories (Dahrendorf 1959 [1957]: 157–165), or rather how Gerhard Lenski talks of conflict-theory based and functionalist (consensus-theory) stratification-theories (Lenski 1984 [1966]: 14–23, 441–443), for Béla Pokol (2004b), the “dual structure of society” means that we have to make a distinction between the views that concentrate on the functional differentiation of society or institutional structure

1987; Sik 2001, 2002, 2010). The colorful expressions indicate the many forms of reciprocal networks to be found. According to the above analysis, these could form a 'third' society.

and those that focus on the structure of social rule or the hierarchically organized social macro groups' struggle for power (Pokol 2004a: 138–150, 390–391, 404–418, 2004b).

In many ways, Zsuzsa Ferge uses similar conceptual distinctions in her newest book (Ferge 2010). In connection with the concept of David Lockwood's system integration and social integration (Lockwood 1956, 1964) and with Jürgen Habermas' system and lifeworld (1987 [1981]) concepts, Ferge analyses the change in Hungarian society's system structure and social structure after 1989. It is clear that Zsuzsa Ferge does not talk of two distinct structures, instead she simply uses two analytical standpoints, while – just like Béla Pokol – she does not try to avoid the term “dual structure” (see for example Ferge 2010: 22).

*While all authors discussed here talk of “the” society's dual structure, they all seem to discover a different duality.* Before we resume our interpretation, let us see if the widespread thesis of society's political divide can be empirically validated.

## Political division after the system change

At the end of the eighties, the developing political pluralism and multi-party system brought to the surface a variant of the folk–urban opposition (Heller–Rényi 1995). This opposition now takes shape in the bloc-formation of the party-structure: polarization can be observed in both the right–left and the conservative–liberal dimensions, in addition, the coupling of the two parameters – what Blau terms consolidation – had intensified (Angelusz–Tardos 2005: 65–93; Fábián 2005: 219–230; cf. Angelusz–Tardos 2003). The press oriented towards public life and politics has in many ways undergone a parallel differentiation. Of the daily papers, preferring *Magyar Hírlap* or *Magyar Nemzet* over *Népszabadság* and *Népszava* – despite their differences – is a definite ideological stance, just like – out of the weekly papers – favouring *Heti Válasz*, *Demokrata*, *Magyar Fórum* or *Barikád* instead of *Magyar Narancs*, *Élet és Irodalom*, *168 óra* or *HVG*. This peculiar separation is similarly way prevalent in the online media, even if we leave internet forums and blogs out of consideration. Just as the political commitment of *Echo* and *Hír TV* is obvious, there is also no doubt about the standpoint of *ATV* or that of *Nap TV* (until 2009). The political logic of the smaller *Lánchíd Rádió* and opposing *Klub Rádió*<sup>11</sup> led to the replacement of *Danubius* and *Sláger Rádió* by *Class FM* and *Neo FM*, two radios that fit better into the political scheme. What does this suggest?

If we accept that in everyday thinking the information obtainable on socio-political reality is pre-selected and pre-interpreted for the individual, then the source of the information becomes very significant. Furthermore, if it is also true that the individual's view of the socio-political reality is not only built on the mass media

11 Against this backdrop, it is not surprising, that the parliament elected in 2010 passed a new law (2010/CLXXXV) on ‘media-services and mass-media’ that bears the same division in terms of which newspapers and media do and do not protest against the statutory media regulation.

but also on the information obtained through the network of personal relationships, then the role of informal relationships, and especially that of opinion-leaders, has a great effect on political and public life identity. If we also suppose that in Hungary the most influential factor in personal-informal relationships is public-political identity, party-bloc affiliation and political party preference – confirmed by the empirical research of Róbert Angelusz and Róbert Tardos (Angelusz–Tardos 2005: 118–120) – then this tendency toward political homophily must have severe consequences for society’s (political) differentiation. As Róbert Angelusz and Róbert Tardos state: “we talk of a bloc-like crystallization of informal contacts of the voters, which contributes to the forming of large blocs of voters and a pregnant political rupture” (Angelusz–Tardos 2005: 145).

The political reorganization of the past few years has somewhat redefined the system of political ruptures, but the ideological-political polarization and the tendency of political homophily – in other words those with a *similar* political mindset seek each other’s company, while interaction density between people from different political-world-view blocs decreases – is stronger than ever according to their later studies (Angelusz–Tardos 2010).

With the partial media coverage of the ever-polarizing public life, the strong bonds (family, relatives and friends) supported by institutional networks (for example, church-religious, civic and trade-union affiliations) are so intensely determined by the bloc-affiliation, that it is not unreasonable to talk of a ‘political divide’. Even if this expression has become devalued because of its political usage and even if it means a three-way instead of a two-way division (because of the 30–35% of politically inactive voters), and even if it can be shown that the traditional variables of social structure (access to economic and cultural resources) cannot adequately explain the political ruptures and their explicative power is decreasing, we can still talk of a political divide (Angelusz–Tardos 2005: 93; Fábíán 2005: 208, 229).

The dual theories of social science are not too common in every-day, public-political discourse. In contrast, the discourse of political divide is quite strong, which has a huge effect – as Róbert Angelusz and Róbert Tardos show – on the development and cultivation of social connections and the character of interactions. As the Thomas theorem states: “if men define situations as real” – for example, the fundamental division of political and public life – then “they are real in their consequences” (Thomas 1966 [1931]: 301, cited by: Némedi 2005: 463, 477).

All this sheds light on the reception of Erdei. Erdei’s “dual society” concept is so popular in social science circles possibly because both sides are able to define their own position and that of the opposing camp based on it. The right-wing conservative and the left-liberal poles can both project their identities and the picture they have constructed of their opposition on to the interwar period, and also legitimize it with Erdei’s authority. In a superficial reading, the duality of the histori-



co-national and modern-bourgeois societies seemingly justifies the last two-three decades' story of political division, and embeds it into history. Even the emergence of new sources that are ruining the dual structure theory cannot worsen the possibility of Erdei's 'ever-actualization' (Erdei 2010; Halmos 2010). More so, if the apathy of the apolitical silent masses under the political division can be (mis)interpreted as our age's folk society. The substitution lends itself, because the masses that turn away from politics are there. András Lányi's warning is relevant: "Do not forget, it is not the country that is divided – it is no longer in that state – only those who talk on behalf of others: the public actors and their audience. The rest is silence" (Lányi 2009: 113).

## Why dual?

The causes of the sociological and every-day popularity and hardness of these dual society views can only be – in my opinion – revealed, if we indeed look at the historical tendencies of Hungarian society's formation from a macro-sociological perspective. First of all, we have to take into account that traditionally it is a luxury of social structure research only in classically closed state-socialist societies and in capitalist core countries to look for the main structuring mechanisms solely from *within* their own nation-state societies. In the case of a small, open capitalist economy (and a society built on it) – in a semi-periphery status like Hungary's – the position it has in world-economy cannot be disregarded.<sup>12</sup> Only those *stratification-researchers* can settle for the container model of society and the underlying methodological nationalism (Wimmer–Glick Schiller 2002), who – based on Tamás Kolosi's distinction (1987: 27–33) – undertake only on the description of the given *population's lay-out* (stratification) and not the *factors that shape the social structure*.

Nation-states on the semi-periphery and periphery of the world-economy develop unevenly (see for example, Szigeti 2010). The social forms represented as dual society on this basis become the ideal-typical opposition of 'developed' and 'underdeveloped' segments. Erdei's modern-bourgeois society, Iván Szelényi's small agricultural commodity producers and socialist entrepreneurs, Kolosi's quasi-market actors, Kemény's non-communist structure and Erzsébet Szalai's multinational sector are similar – despite their several differences – in the sense that these social formations are closer to the capitalist world-system's (Western) core in their organization and lifestyle. A good demonstration of this is the counter-pole, as it is interpreted to be a feudal, rank-ordered (late, backward, underdeveloped) social organization by all of them.<sup>13</sup>

12 According to Immanuel Wallerstein's world-system-theory, the capitalist world-system came into being in the long XVI. century in Western Europe. The geographic division of labor within the system causes differentiation and the world-system becomes hierarchic. The leading economic activities, in a socio-economic sense, concentrate in the core(s) of the system, which – because of the unequal exchanges – is (are) surrounded by 'less developed' peripheral areas. The semi-periphery, like East-Central Europe with Hungary in it, occupies an intermediate position – with all its circumstances, that cannot be addressed here (Wallerstein 2004; Arrighi–Drangel 1986; Arrighi 1990; Böröcz 1992; Melegh 2009; Radice 2009; Szigeti 2010).

13 The theoretical typologies of Béla Pokol and Zsuzsa Ferge – who are interpreting a theoretical duality as a real structural one – do not, while Júlia Szalai's distinction only partly fits into this explanatory scheme.

In addition, the *judgment* on Western, more developed, modern structure (in short: core) and the feudal, underdeveloped, lagging society (in short: periphery) is diametrically opposed in several concepts. While the core seems to be ranked higher and more desirable by most, the underdeveloped periphery is in turn labelled lower-ranked and obsolete. This evaluation clearly coincides with the capitalist world-system's cultural-ideological mapping: the opposition of the concepts of "civilization" and "culture" (Elias 1998 [1939]; Wessely 1996) and the discursive order of the East–West slope (Melegh 2006, 2009). Despite the fact that Hungarian society's place in the capitalist world-system was not taken into account in the dual society concepts, the cognitive scheme based on these did appear in their interpretation of reality.

A good demonstration of this duplicating cognition method is Tibor Kuczzi's analysis. In his view, the researchers from the societies of the world-economy's (semi)periphery are trying to define and explain why their own society is not model-like, using concepts developed for the Western, core capitalist countries' idealized circumstances. (They rely on the scientific literature by Anglo-Saxon, French and German structure- and stratification-researchers.) What could they possibly do, if the model-society reflected in the cognitive methods is dissolving into the analyzed Central-European reality, and won't describe it properly? Tibor Kuczzi claims that when this happens they resort to the cognitive strategy of duplication:

Some of the sociologists of the societies-to-be tried to overcome the problem by dividing the human world they lived in into a traditional, pre-social world and one in which procedures and human behaviour can be understood with the rules of the model-society. They postulated that they are dealing with a dual society. This metaphor can cause a lot of problems in the interpretation, because there are no two societies divided by some border, rather it is the application of two different theoretical concepts to the same subject. The concept of dual society is misleading, because it applies an abstract and a concrete spatial metaphor at the same time. Modern society is not alongside the traditional one, since the former has no spatial reference (Kuczzi 1998: 50).

When the reality of societies of the (semi)periphery does not fit smoothly into the models, theoretic schemes and approaches developed for the interpretation and explanation of the relations of the core-capital countries, then researchers often resort to the method of *duplication*. One of Erdei's sentences is a good illustration of this: "So these are capitalist class societies too, though not real and pure bourgeois social structures, but feudal historical or unhistorical bourgeois formations" (Erdei 1976a: 24). The result is the seemingly empirically validated idea of the dual society.

## On the gains and losses of Hungarian sociology's dual-society theories

In my paper, I am *not* saying that Hungarian society's structure is unified and that no ruptures, brakes, superposed structures or differing institution- and relation-organizing principles can be observed. On the contrary: in my opinion, the approach of Hungarian society's *multiple dualities* is just as convincing as the consistent denial of "the" duality. It is all so complex that it cannot be condensed into a single duality. Out of the entire complex society a duality thought to be fundamental can always be emphasized – this being useful for Hungarian sociology to draw attention to a supposed rupture. Altogether, the view of "the" duality can only originate from the *overgeneralization* and unjustified *overemphasis* of a partial duality and finally the duplication of society. Of course, there could be other methods to correct the maladjustment of the normative view and the 'examined subject' – falling short of expectations – with, on the one hand maintaining the ideal of the model society and the *orientalization* of their own society, on the other hand, the rejection of the model society and – at the same time – the rejection of Western sciences altogether. This is altogether a *regression* into national science (thought to be authentic) and scientific autarchy.

As a structural constraint, the East-Central European semi-periphery position conditions Hungarian sociology to "recognize" this duality. The structure of Hungarian society can only be seen (be made to seen) as dual, because both the real processes (the semi-periphery position) and the cognitive apparatus make the examiner susceptible to it. Beside the *orientalising* and the *regressive* cognitive strategies, the duplicating method is the most defensible and acceptable one – but we do not have to be content with it. East-Central European sociology's cognitive chances are given: the chance to go beyond this *duplicating* strategy is still before us.

Hungarian social science provides several examples to follow. Iván Szelényi and his colleagues drew a rather convincing picture of the structure of Central European society of the mid-nineties (and of Hungary within this) in their book, *Making Capitalism without Capitalists*. They maintain the perspective of sociological understanding without giving up the possibility of a critical interpretation of society (Eyal-Szelényi-Townsley 1998). They give an example of how to construct an image of the organization of social-space, the formation of elite groups and classes and the struggle for limited resources while avoiding the society duplicating strategy. Róbert Tardos's work can also serve as an important example, which – as a continuation of the pre-system-change cultural-interactional stratification model started by Róbert Angelusz and Róbert Tardos – builds a structure model by combining social network analysis, profession distribution survey and milieu-theory (Angelusz-Tardos 1991; Tardos 2008). Both depictions of social structure prove that the strategy of society duplication can be avoided even if we try to find

a way to describe social relations focusing on power and even if we try to disclose the organizational patterns of social relationships.

## References

- Andorka, Rudolf (2006): Társadalmi szerkezet és rétegződés. In Andorka, Rudolf: *Bevezetés a szociológiába*. Budapest: Osiris Kiadó, 155–202.
- Angelusz, Róbert – Tardos, Róbert (1991): *Hálózatok, stílusok, struktúrák*. Budapest: ELTE Szociológiai Intézet – Magyar Közvéleménykutató Intézet.
- Angelusz, Róbert – Tardos, Róbert (2003): A konszolidáció/interszekció fogalompar perspektívái a szociológiában. *Szociológiai Szemle*, 13(4): 3–19.
- Angelusz, Róbert – Tardos, Róbert (2005): Választói tömbök rejtett hálózata. In Angelusz, Róbert – Tardos, Róbert (eds.): *Törések, hálók, hidak. Választói magatartás és politikai tagolódás Magyarországon*. Budapest: Demokrácia Kutatások Magyar Központja Alapítvány, 65–159.
- Angelusz, Róbert – Tardos, Róbert (2010): A törésvonalak átrendeződése, avagy egy divergencia-spirál megjelenése? *Szociológiai Szemle*, 20(1): 4–25.
- Arrighi, Giovanni (1990): The Developmentalist Illusion. A Reconceptualization of the Semiperiphery. In Martin, William G. (ed.): *Semiperipheral States in the World-Economy*. New York: Greenwood Press, 11–42.
- Arrighi, Giovanni – Drangel, Jessica (1986): The Stratification of the World-Economy: An Exploration of the Semiperipheral Zone. *Review*, 10(1): 9–74.
- Bartha, Eszter (2003): Újkapitalizmus, elitek és a másként gondolkodás esélyei és lehetőségei Magyarországon. Szalai Erzsébet két könyvéről. *Eszmélet*, 15(61).
- Bognár, Bulcsu (2011 [2003]): „Csak egy rettenetes nagy söprögetés segít”. Erdei Ferenc értelmezése a magyar társadalomfejlődésről a kézirat szövegben. *Valóság*, 2003. 46(8): 57–79. Reprinted: in Bognár, Bulcsu: *Népi szociográfia és társadalomtudomány. Írások Erdei Ferenc szociográfiáinak (1931–1944) társadalomszemléletéről*. Budapest: Loisir Kiadó, 2011. 248–281.
- Bognár, Bulcsu (2010): A szociológus végső kudarca – a kézirat munka (1944). In Bognár, Bulcsu: *Erdei Ferenc szociológiája*. Budapest: Loisir Kiadó, 104–114.
- Böröcz, József (1990): A kádárizmustól a parlagi kapitalizmusig. A fejlett informális építésének időszerű kérdései. *Mozgó Világ*, 16(8): 61–66.
- Böröcz, József (1992): Dual Dependency and Property Vacuum: Social Change on the State Socialist Semiperiphery. *Theory & Society*, 21(1): 77–104.
- Böröcz, József (1993): Simulating the Great Transformation: Property Change under Prolonged Informality in Hungary. *Archives européennes de sociologie / Europäisches Archiv für Soziologie / European Journal of Sociology*, 34(1): 81–107.
- Böröcz, József (2000): Informality Rules. *East European Politics and Societies*, 14(2): 348–380.
- Böröcz, József – Southworth, Caleb (1998): “Who You Know?": Earnings Effects

- of Formal and Informal Social Network Resources under Late State Socialism, Hungary, 1986–87. *Journal of Socio-Economics*, 27(3): 403–427.
- Czakó, Ágnes – Sik, Endre (1987): Managers' Reciprocal Transactions. In Lengyel, György (ed.): *Education, Mobility and Network of Leaders in a Planned Economy*. Budapest: Department of Sociology, Karl Marx University of Economic Sciences, 141–171.
- Dahrendorf, Ralf (1959 [1957]): *Class and Class Conflict in Industrial Society*. Stanford, California: Stanford University Press.
- Elias, Norbert (1998 [1939]): A "kultúra" és a "civilizáció" ellentétének szociogenezise Németországban. In Wessely Anna (ed.): *A kultúra szociológiája*. Budapest: Osiris Kiadó – Láthatatlan Kollégium, 41–51.
- Erdei, Ferenc (1976a): A magyar társadalom a két háború között I. *Valóság*, 19(4): 22–53.
- Erdei, Ferenc (1976b): A magyar társadalom a két háború között II. *Valóság*, 19(5): 36–58.
- Erdei, Ferenc (2010): A történelmi népi társadalom (IV. A paraszttársadalom). *Szociológiai Szemle*, 20(4): 12–42.
- Eyal, Gil – Szelényi, Iván – Townsley, Eleonor (1998): *Making Capitalism without Capitalists. Class Formations and Elite Struggles in Post-Communist Central Europe*. London–New York: Verso.
- Fábián, Zoltán (2005): Törésvonalak és a politikai-ideológiai azonosulás szerepe a pártszimpátiák magyarázatában. In Angelusz, Róbert – Tardos, Róbert (eds.): *Törések, hálók, hidak. Választói magatartás és politikai tagolódás Magyarországon*. Budapest: Demokrácia Kutatások Magyar Központja Alapítvány, 207–242.
- Ferge, Zsuzsa (1996): A rendszerváltás nyertesei és vesztesei. In Andorka, Rudolf – Kolosi, Tamás – Vukovich, György (eds.): *Társadalmi riport 1996*. Budapest: TÁRKI – Századvég Kiadó, 414–443.
- Ferge, Zsuzsa (2010): *Társadalmi áramlatok és egyéni szerepek. Adalékok a rendszerstruktúra és a társadalmi struktúra átalakulásának dinamikájához*. Budapest: Napvilág Kiadó.
- Gábor, R. István – Galasi, Péter (1981): *A "második" gazdaság. Tények és hipotézisek*. Budapest: Közgazdasági és Jogi Könyvkiadó.
- Gyáni, Gábor (1997): Polgárosodás mint zsidó identitás. *Buksz*, 9(3): 266–278.
- Gyáni, Gábor (1998): Forráskritika és bizonyítás. Viszontválasz Karády Viktornak. *Buksz*, 10(1): 20–28.
- Gyáni, Gábor (2001): Érvek a kettős struktúra elmélete ellen. *Korall*, 2(3–4): 221–233.
- Habermas, Jürgen (1987 [1981]): Intermediate Reflections: System and Lifeworld. In Habermas, Jürgen: *The Theory of Communicative Action. Vol. 2. Lifeworld and System: A Critique of Functionalist Reason*. Cambridge: Polity Press, 113–197.

- Habich, Roland – Spéder, Zsolt (1998): Vesztesek és nyertesek: a társadalmi változás következményei három országban. In Kolosi Tamás – Tóth István György – Vukovich György (eds.): *Társadalmi riport 1998*. Budapest: TÁRKI, 117–139.
- Halmos, Károly (2010): A történelmi népi társadalom. A magyar társadalom a két háború között című tanulmány negyedik része. *Szociológiai Szemle*, 20(4): 4–11.
- Hankiss, Elemér (1984): „Második társadalom”? *Valóság*, 27(11): 25–44.
- Heller, Mária – Rényi, Ágnes (1995): Diszkurzív stratégiák az új magyar nyilvánosságban. A népi–urbánus vitától a magyar–zsidó szembenállásig. In Kapitány, Ágnes – Kapitány, Gábor (eds.): *Jelbeszéd az életünk. A szimbolizáció története és kutatásának módszerei*. Budapest: Osiris Kiadó – Századvég Kiadó, 432–448.
- Huszár, Tibor [H. T.] (1976): [Bevezető Erdei Ferenc A magyar társadalom a két háború között c. kéziratához.] *Valóság*, 19(4): 22.
- Huszár, Tibor (2003): Egy kézirat történetéhez. In *Harminc év. Mannheim-előadások*. Budapest: ELTE Szociológiai és Szociálpolitikai Intézet, 89–97.
- Karácsony, Gergely (2005): A történelem fogságában. Generációk, életutak, politikai preferenciák Magyarországon. In Angelusz Róbert – Tardos Róbert (eds.): *Törések, hálók, hidak. Választói magatartás és politikai tagolódás Magyarországon*. Budapest: Demokrácia Kutatások Magyar Központja Alapítvány, 161–205.
- Karády, Viktor (1998): Zsidó és nem zsidó polgárosodás? Válasz Gyáni Gábornak. *Buksz*, 10(1): 9–20.
- Kemény, István (1992a): Utószó helyett. [Kemény Istvánnal az interjút Gábor Kálmán készítette.] In Kemény, István: *Szociológiai írások*. Szeged–Budapest: Replika Kör, 289–296.
- Kemény, István (1992b): [Előszó a 'Kompromisszumok egyezség nélkül' (1978) című tanulmányához.] In Kemény, István: *Szociológiai írások*. Szeged–Budapest: Replika Kör, 201.
- Kemény, István (2010): *Előadások a magyar társadalomfejlődés szociológiájáról*. Szeged: Belvedere Meridionale.
- Kolosi, Tamás (1974): *Társadalmi struktúra és szocializmus*. Budapest: Kossuth Könyvkiadó.
- Kolosi, Tamás (1987): *Tagolt társadalom. Struktúra, rétegződés, egyenlőtlenség Magyarországon*. Budapest: Gondolat Könyvkiadó.
- Kolosi, Tamás (1991): Interjú Kolosi Tamással. (Készítette: Gábor Kálmán.) *Replika*, 2(2–3): 105–113.
- Kolosi, Tamás (2000): *A terhes babapiskóta. A rendszerváltás társadalomszerkezete*. [Sági Matild és Róbert Péter közreműködésével.] Budapest: Osiris Kiadó.
- Kolosi, Tamás (2003): Szubjektív megjegyzések a struktúra-kutatás harminc évéhez. In *Harminc év. Mannheim-előadások*. Budapest: ELTE Szociológiai és Szociálpolitikai Intézet, 99–107.

- Kolosi, Tamás – Sági, Matild (1996): Rendszerváltás és társadalomszerkezet. In Andorka, Rudolf – Kolosi, Tamás – Vukovich, György (eds.): *Társadalmi Riport 1996*. Budapest: TÁRKI – Századvég Kiadó, 149–197.
- Konrád, György – Szelényi, Iván (1974 [1971]): Social Conflicts of Under-urbanization. In Brown, A. A. – Licari, J. A. – Neuberger, E. (eds.): *Urban and Social Economics in Market and Planned Economies. Vol. 1*. New York: Praeger, 1974. 206–226.
- Kornai, János (1983): Bürokratikus és piaci koordináció. *Közgazdasági Szemle*, 30(9): 1025–1038.
- Kovács, Éva – Melegh, Attila (1997): Tükör által homályosan. Kísérlet Erdei Ferenc A magyar társadalom a két világháború között című tanulmányának rekonstrukciójára. In Szabó, Márton (ed.): *Szövegvalóság. Írások a szimbolikus és diszkurzív politikáról*. Budapest: Scientia Humana Kiadó, 133–156.
- Kövér, György (2001 [1998]): Értelmezési keretek. In Gyáni, Gábor – Kövér, György: *Magyarország társadalomtörténete a reformkortól a második világháborúig*. Budapest: Osiris Kiadó, 11–37.
- Kuczi, Tibor (1998): Társadalomsakk. *Replika*, 9(30): 41–50.
- Lányi, András (2009): A tizenöt éves kapitány. In Lányi, András: *Porcelán az elefántboltban. Az ökológiai politika kezdetei Magyarországon*. Budapest: Heti Válasz Kiadó, 73–113.
- Lenski, Gerhard E. (1984 [1966]): *Power and Privilege: A Theory of Social Stratification*. Chapel Hill: University of North Carolina Press.
- Lockwood, David (1956): Some Remarks on “The Social System”. *The British Journal of Sociology*, 7(2): 134–146.
- Lockwood, David (1964): Social Integration and System Integration. In Zollschan, George K. – Hirsch, Walter (eds.): *Explorations in Social Change*. Boston: Houghton Mifflin, 244–257.
- Melegh, Attila (2006): *On the East–West slope. Globalization, Nationalism, Racism and Discourses on Central and Eastern Europe*. Budapest – New York: CEU Press.
- Melegh, Attila (2009): A globalizáció és Magyarország. In Koppány, Judit – Udvarhelyi, Éva Tessa (eds.): *A világ az iskolában. Tanulmányok a globalizációról tanároknak*. Budapest: Artemisszió Alapítvány, 16–23.
- Nagy, J. Endre (1993 [1986]): Harcban az angyallal. Erdei politikai világnézetének és struktúrarajzának összefüggései az 1940-es évek elején. In Nagy, J. Endre: *Eszme és valóság. Magyar szociológiatörténeti tanulmányok*. Szombathely: Pesti Szalon – Savaria University Press, 103–114.
- Némédi, Dénes (1978): A szintézis keresése. Erdei Ferenc szociológiája a 40-es évek első felében. *Szociológia*, 8(4): 511–525.
- Némédi, Dénes (2005): *Klasszikus szociológia 1890–1945*. Budapest: Napvilág Kiadó.
- Pokol, Béla (2004a): *Szociológiaelmélet. Társadalomtudományi trilógia I*. Budapest: Századvég Kiadó.

- Pokol, Béla (2004b): A társadalom kettős szerkezete. *Szociológiai Szemle*, 14(3): 36–51.
- Polányi, Karl (1957): The Economy as Instituted Process. In Granovetter, Mark – Swedberg, Richard (eds.): *The Sociology of Economic Life*. Boulder, CO: Westview Press.
- Radice, Hugo (2009): *Halfway to Paradise? Making Sense of the Semiperiphery*. (Working Paper CSGP 09/3.) Peterborough, Ontario, Canada: Trent University, 1–18.
- Romsics, Ignác (2005): Magyarország(ok) a két világháború között. In Gombár, Csaba – Volosin, Hédi (eds.): *Két Magyarország?* Budapest: Osiris – Korridor Kiadó, 49–101.
- Sik, Endre (1996): Egy ló-öszvér a lovokról és a szamarokról. Adalék a második gazdaság hazai eszmetörténetéhez. *Közgazdasági Szemle*, 43(7–8): 704–725.
- Sik, Endre (2001): Aczélhálóban. Adalék a kapcsolati tőke működésének megismeréséhez. *Szociológiai Szemle*, 11(3): 64–77.
- Sik, Endre (2002): Mindennapi korrupcióink. *Szociológiai Figyelő*, 6(1–2): 131–145.
- Sik, Endre (2010): Network Capital Dependent Path-Dependency. *Corvinus*, 1(1): 77–102.
- Szalai, Erzsébet (2001): Kettős gazdaság – kettős társadalom. In Szalai, Erzsébet: *Gazdasági elit és társadalom a magyarországi újkapitalizmusban*. Budapest: Aula Kiadó, 240–247.
- Szalai, Erzsébet – Krausz, Tamás – Szigeti, Péter (2002): Gazdasági elit és társadalom a magyarországi újkapitalizmusban. Beszélgetés Szalai Erzsébet azonos című könyvéről (Aula Kiadó, Budapest, 2001). *Eszmélet*, 14(54): 83–106.
- Szalai, Júlia (2007): *Nincs két ország...? Társadalmi küzdelmek az állami (túl)elosztásért a rendszerváltás utáni Magyarországon*. Budapest: Osiris Kiadó.
- Szelényi, Iván – Manchin, Róbert (1987): Social Policy under State Socialism. In Esping-Anderson, G. – Rainwater, L. – Rein, M. (eds.): *Stagnation and Renewal in Social Policy*. White Plains, N.Y.: Sharpe, 102–139.
- Szelényi, Iván (1978): Social Inequalities in State Socialist Redistributive Economies: Dilemma for Social Policy in Contemporary Socialist Societies of Eastern Europe. *International Journal of Comparative Sociology*, (1–2): 63–87.
- Szelényi, Iván (1981): The Relative Autonomy of the State or the State Mode of Production. In Dear M. J. – Scott, A. J. (eds.): *Urbanization and Urban Planning in Capitalist Societies*. London: Methuen, 565–591.
- Szelényi, Iván (1986–87): The Prospects and Limits of the East European New Class Project: An Auto-critical Reflection on The Intellectuals on the Road to Class Power. *Politics and Society*, 15(2): 103–144.
- Szelényi, Iván (1988): *Socialist Entrepreneurs. Embourgeoisement in Rural Hungary*. [In collaboration with Manchin, Róbert; Juhász, Pál; Magyar, Bálint and Martin, Bill.] Madison: Wisconsin University Press.



- Szelényi, Iván (1990): Utószó: Jegyzetek egy szellemi önéletrajzhoz. In uő.: *Új osztály, állam, politika*. Budapest: Gondolat Könyvkiadó, 433–489.
- Szigeti, Péter (2010): A félperiféria vita jelentősége. *Eszmélet*, 22(85): 175–185.
- Tardos, Róbert (2008): Foglalkozás, miliő, kapcsolathálózatok: külön világok? Egy tipológiai kísérlet körvonalai. *Századvég*, 13(50): 5–50.
- Thomas, William Isaac (1966 [1931]): The Relation of Research to the Social Process. In Janowitz, Morris (ed.): *W. I. Thomas on Social Organisation and Social Personality*. Chicago: The University of Chicago Press, 289–305.
- Wallerstein, Immanuel (2004): *World-Systems Analysis: An Introduction*. Durham, North Carolina: Duke University Press.
- Wimmer, Andreas – Glick Schiller, Nina (2002): Methodological Nationalism and Beyond: Nation-state Building, Migration and the social sciences. *Global Networks*, 2(4): 301–334.
- Wessely, Anna (1996): The Cognitive Chance of Central European Sociology. In Hadas, Miklós – Vörös, Miklós (eds.): *Colonization or Partnership? Eastern Europe and Western Social Sciences. (Replika Special Issue 1.)* Budapest: Replika, 11–19.
- Zentai, Violetta (1999): Győztesekből tettesek, vesztesekből áldozatok? Kétosztatú társadalomképek a rendszerváltás magyarországi kritikájában. *2000*, 11(9): 10–20.

REVIEW OF SOCIOLOGY, 2011/4: 21–47.

# Topography and Dynamics of Co-author Networks<sup>1</sup>

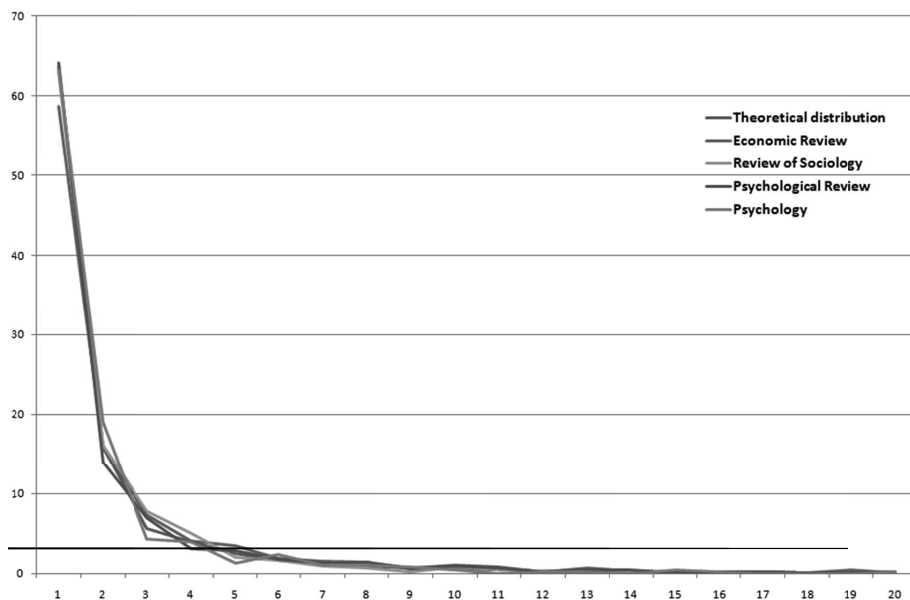
István MICSINAI

istvan.micsinai@gmail.com

**ABSTRACT:** Co-author networks have become the center of the attention of both scientometrics and network researches during the last decade. In this article I put more emphasis on the scientometrics side, I compare the actual and the international results related to my topic.

The outcome of the research is that the rate of co-authorship of the last fifty years in Hungary has no significant change. The number of the largest component of all publishing scientists in co-authorship covers only twenty-eight, thirty-three percent. The international co-operation shows surprisingly low number of articles and the distribution is uneven. The growth of networks - contrary to the 'mature' Western science - can be modeled easily by the logistic growth function.

Keywords: networks, co-authorship, scientometrics, logistic growth



<sup>1</sup> English revision by Evelin Szarka.

## Introduction

After the emergence and success of the concept of scale-free law in the beginning of the new millennium many researches were conducted in the field of co-author networks. The growing scale of available databases provides an excellent background for its analyzation from a network and from a scientific historical perspective. Many articles were written in several areas (in physics: [Newman 2000, 2001a, 2001b]; about the Colombian science: Bernal-Llanos [2005]; psychology and philosophy in Cronin et al. [2003]; evolutionary computation: Merel-Cotta [2006]; sociology: Moody [2004]; Mapping the network of co-authors of the journal *Scientometrics* was carried out by [Hou et al 2006]. Hungarian scientometric analysis (distribution of numbers of publications, citation networks) was made in some journals. Researches about the *Economic Review* Such-Tóth (1989), the *Space and Society* Reisenger-Tóth (2007) and the *Political Review* Cselényi (2009) have been conducted.

Analyzing co-author networks raises some methodological questions: what is the significance of co-authorship of scientific operation? Can we use co-authorship for the analyzation of the functioning of science at all? If yes, in what sense? Has it always the same meaning for each connection?

Right from the beginning of the study I would like to state that besides the authorship of scientific publications there are other additional levels of scientific analysis. One of them is the well-known citation data (since 1960) and the citation networks. The following analysis of a potential level of involvement lies in the keywords: keywords indicate the directions of science, which are the most interesting for scientists. The presented aspects are interesting separately, without no doubt, however, their combination offers the greatest potential (Mutschke Quan-Haase 2001).

Before I proceed with the presentation of the causes of authorship, I go back to scale-free law for a while. The concept mentioned is widely used within a short time, great number of studies analyzed its validity but it should be noted that its history is not without precedent. In the field of scientometrics Alfred Lotka has carried out a research in the twenties about what proportion of articles could be found at  $n$  number of authors. Under the formula of Lotka's law ( $1/n^2$  the ratio of  $n$  publications of researchers) the distribution is the same as in the Albert-Laszlo Barabási idea: there are many who have little and few have much.

I clarify the necessity of returning to the reasons of co-authorship: co-operation has both structural and emotional reasons. The availability of laboratory equipment has increased the level of co-operation in natural sciences; however, there is no need of laboratory application equipment in social sciences but in many cases large-scale data collection needs teamwork. The terms of payment systems will also promote cooperation in scientific work: its impact can be experienced in natural sciences.

To make it simple I consider the relations among co-authors symmetrical in this article. Yoshikane distinguishes (2006) two roles: leader and follower. Usually the first person of the authors is the leader of the research, who in most cases is also the main organizer of it. A very different approach can be found by Patel (quoted by Glänzel-Schubert 2004): the analyzation of scientific articles found that the publisher often indicates those who write the article indirectly: e. g. they have participated only with their ideas and critics. The term sub-authorship was introduced on this basis of this concept. Cronin (2001), however, found that an article has fifty or sixty or even more than hundred authors in certain areas of science, for example high-energy physics, so he introduced the concept of hyper-authorship in this context.

In the case of sociology, during the examination of the American sociology James Moody (2004) finds that within various science topics large differences can be observed. It goes without saying that theoretical issues can be explored independently, while empirical researches are increasingly common. According to the research, co-author articles are more likely to appear in health care and education issues than the already mentioned theoretical ones. The fact of co-authorship is under the influence of a researcher's career as well. The more one has been moving for a long time in science, it is more likely that they cooperate. Moody has shown that every single publication of the 1.28 percent increase the likelihood of participation in co-authorship. Although, in areas with low co-author articles rates compared to earlier periods in subsequent periods, this ratio was even lower. It can be concluded that specialization has become more important over time. In the field of gender studies the probability of co-author articles has an average value at the beginning of the examined period but later only half of them were written in co-operation. The same process can be observed in reverse: in the case of methodology the rate of co-author articles was also the average for the early period, in the future, however, a 1.21-fold increase can be seen in the likelihood. The number of writers for quantitative works the likelihood of co-author articles is five times greater. In Moody's view, works of co-authors become normative once they appear in an area, they become common.

The co-authorship has become normative and we know its importance from Brian Uzzi's writings (Uzzi et al. 2007). The ISI Web of Science database has been analyzed from the mid-fifties (the starting year of the data vary with different fields) until the millennium. During work it has been found that the authorship of all scientific disciplines (natural and engineering sciences, social sciences, humanities and patents are available from 1975) has increased significantly. This is also valid for mathematics (the level of authorship grew 19 percent to 57 percent in the period under review), which many have been seen as the field of solitary geniuses. These results are not new to those already mentioned but Uzzi and his team has taken another important step. During their study they compared the

received citation of the self-made and that of co-author articles. The surprising result is that co-author articles have much more citations than self-writings. In addition, the number of received citations grows at the same time: while in the fifties in natural sciences there were 1.7 times as many references in co-author articles as in those of single authors; this share was 2.1-fold in the 2000s. (The advantage remained even when self-citations were eliminated.)

The relationship between recognition and the authorship of the work also means that good work and creativity are linked closely. Subsequently Uzzi investigated this issue further and during the success and failure analysis of New York's Broadway musicals he found interesting relations (Uzzi 2005). Those cuts have become a success where teams were both heterogeneous and homogeneous. Homogeneous in the sense that staff members have previously worked together, heterogeneous meant that new members came up with new ideas.

The emergence of creativity in terms of structural analysis is indeed an interesting task. Finally, I mention the work of Randall Collins (Collins 1999) who examined the history of philosophy through the birth of the conditions for innovative ideas. Collins' theory of the innovative ideas was born in the heart of scientific networks.

## Evolution of networks

Besides the topographic analysis of networks, the analysis of the growth dynamics of networks is also an important issue under the Barabási model (Barabási 2003) – “who is to be given” – that is the principle of preferential attachment: who has a large number of connections, new relationships can be obtained by them more easily.

According to Barabási, all points can acquire new ones on the basis of the already existing proportion of relations. According to this research the relations of the points will increase proportionally with the square root of time. If we consider the points in various co-author networks we can see that they have different lifetimes, and it reveals a more nuanced picture. Jose Mendez (2000) says points lose their ability to attract after a while and they cannot create new connections. Amaral (2000), Mendez take forward the idea, which is the logical consequence of this: if central points cannot create new connection points after a time (either outside networks), then centers will be of limited size.

The Barabási model – in order to explain the growth of dynamics – refines the existing model by the analysis of the spectacular success of Google and introduces the concept of fitness. Now we can imagine two kinds of networks: a scale-free network – the rich will be richer – but it also displays a newer kind of organization where the winner takes it all virtually (e. g. Microsoft). This, however, is no longer a scale-free network, the majority of the links points towards a central operator.

Barabási et al (2002) examined mathematical neuroscience and the development of networks of co-authors in the period of 1991–1998. The tested network proved to be scale-free. The average degree has increased over time; the initial value of around five in 1998 grew up to twelve points. Surprisingly, the diameter of the network declined. Two reasons explain this decline: one explanation is the increasing number of network connections within the reduced distance the other says that the diameter reduction is the result of the limited lifetime of the database. (The network density decreased slowly in time.) The study analyzed, inter alia, that what proportion of authors has the largest network. In the case of neuroscience from the initial fifty-five percent in 1991 it grew up to ninety percent until the end of the test period. The largest network of co-authors in Maths reached seventy percent from zero between 1991 and 1998, but has not reached saturation yet. The process of time, the average connection number is growing. In neuroscience the average number of connections is from five to twelve, in mathematical sciences it increased from two to four. Barabási recognizes that the number of authors of publications was ignored in their model, and due to a short-term view they also did not take in consideration the dynamics of players leaving the network.

## II. Analysis

Co-author networks of three Hungarian social scientific journals – an economical, a psychological and a sociological one are presented in this study. The analysis focuses on the leading papers of these sciences: the *Economic Review*, the *Psychological Review* and the *Review of Sociology*. In the field of sociology there is no other paper that would have been published for a longer time. Although in the field of economics there are more reviews their profiles differ significantly from the leading one. Psychology constitutes the only exception: another paper appeared beside the official journal even before the change of regime.

I analyzed the co-author networks in each case from the very first issue to the last one in 2009. The *Economic Review* started in 1954, the *Psychological Review* in 1960 and the *Psychology* in 1981. The *Review of Sociology* started only in 1991, although its predecessor, the *Sociology* was published from 1972 to 1990. The *Economic Review* is published eleven times a year (with a double issue in summer). The *Review of Sociology* and the *Psychological Review* appear four times a year. It is important to note that the *Psychological Review* published six issues per year between 1975 and 1995 but there were certain years, in 1987–88 and in 1992–93, when only three double issues appeared.

During the collection of the names of authors I have faced the problem that the same person could also published under various names. In case of women after marriage their husband's name was partly or totally included and used in publications. Besides the different names of the same person, it also caused difficulty

when I might have considered two researchers the same but accidentally just their name was identical. Besides the previous mentioned problems, approximately one percentage of the data seems uncertain.

The *Economics Review* was published 608, the *Psychological Review* 200, the *Review of Sociology* 132, the *Psychology* 116 times during their history. The analysis includes book reviews, too. The number of book reviews refers to their role in scientific communication. Book reviews cannot be found in science journals, the speed of the development of knowledge does not allow researchers to publish their results in a book. On the contrary, in social sciences and humanities books constitute an important part of the scientific communication like journals.

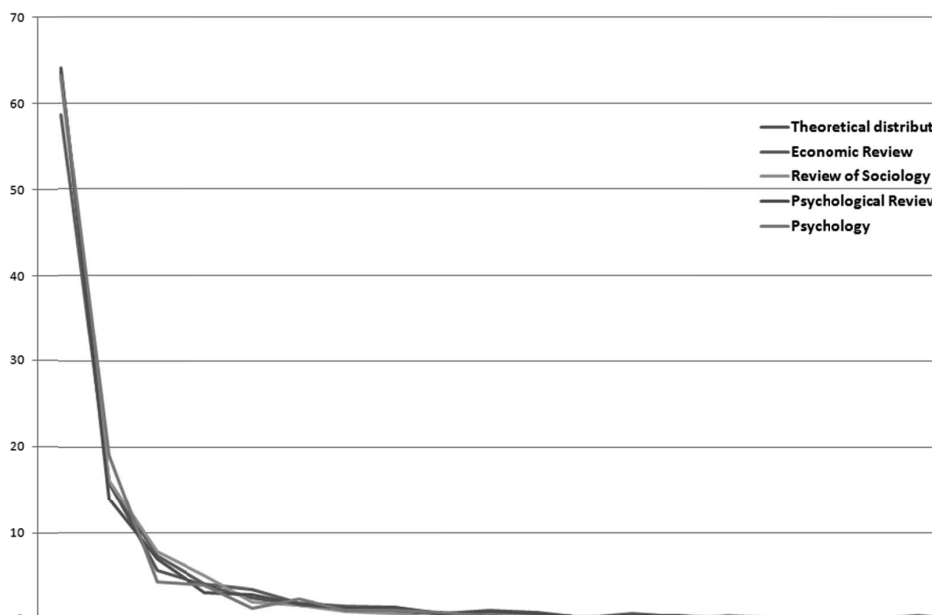
In Canada (Laviri re et al. 2004) book reviews represent 27 percent in social science journal articles and 57 percent in those of humanities. In case of Hungarian journals the ratio of book reviews are lower: it is seventeen percent in the *Psychological Review*, it is fifteen percent in the case of the *Economic Review* and it is also seventeen percent in the *Review of Sociology*. (Some co-authored articles can be found among them but their number is insignificant.) Differences between the ratios of book reviews can be found when we operate with time. In the case of the *Economic Review* it decreased by half (from nineteen to nine percent) from the initial period of 1956–1975 to the period of 1991–2009. In the initial phase of the *Psychological Review* (between 1960 and 1975) there is no book review: between 1976 and 1990 twenty, between 1991 and 2009 twenty-four percent of the articles were published as reviews. Book reviews represented 20 percent of the articles in the *Review of Sociology* in the first twenty years (1972–1990) while this number declined to seventeen percent in the period of 1991–2009.

## Publication Activity

Scientometrics research started in the early twentieth century with the works of Alfred Lotka. Although Lotka was a chemist, he analyzed the distribution of the publications' number in scientific journals and found that all number of authors who have written  $n$  articles average with  $1/n^2$ . Lotka's law is, therefore, that during a specified period of 100 publications with one author there are 25 articles with two and 11 articles with three authors. I have examined the data until twenty publications and I present the theoretical distributions in this light. In percentage terms this means that articles which represented the 63.56 percent have authors with one publication, another 15.89 percent has authors with two and finally, 5.72 has authors with three publications. We can group authors in three different ways: all members of a co-authored article can be handled the same way (single authorship), or co-authors get the proportion of authorship according to the number of authors (fractional authorship), and finally, only the first author is considered. In this study all authors were considered equal. The percentages of authors of the

studied journals are shown in the *first figure*. The *Economic Review* and the *Psychological Review* differ from the theoretical distribution. In the case of the former it is five percent lower (it is fifty-eight percent instead of sixty-three), in the case of the latter one, the number of authors with two articles is higher, while those with three publications represent a lower percentage than the expected theoretical value. The overall results can be said to correspond with the expected theoretical distribution.

Figure 1: Distribution of authors according to the number of articles and Lotka's law



Researchers having the most publications in different journals can be seen in the first table. The first number in the parenthesis is the number of self-authored publications the second one is that of the co-authored articles. Except for Kolosi Tamás, Angelusz Róbert, Tardos Róbert in the field of sociology, Molnár Márk and Csépe Valéria in that of psychology, there are no more researchers among the top ten most publishing authors who co-authored more than wrote independently. Six or more co-authored articles were written by only fourteen researchers from the forty, therefore the emphasis is clearly on independent publications.



Table 1: Researchers with the highest number of publications

Economic Review		Review of Sociology	
Name	Number of publications	Name	
Bródy András	64 (60+4)	Bertalan László	
Kornai János	53 (43+10)	Némedi Dénes	
Erdős Tibor	39 (38+1)	Andorka Rudolf	
Pócs Ervin	35 (33+2)	Kulcsár Kálmán	
Csikós-Nagy Béla	33 (30+3)	Tamás Pál	
Simonovits András	33 (26+7)	Kolosi Tamás	
Török Ádám	33 (31+2)	Sik Endre	
Csaba László	31 (30+1)	Szántó Zoltán	
Falusné Szikra Katalin	31 (29+1)	Tardos Róbert	
Mátyás Antal	31 (31+0)	Angelusz Róbert	
Psychological Review		Psychology	
Name	Number of publications	Name	
Pléh Csaba	81 (56+25)	Marton Magda	
Buda Béla	51 (48+3)	Halász László	
Klein Sándor	47 (40+7)	László János	
Hegedűs T András	36 (33+3)	Cziegler István	
Kardos Lajos	25 (15+10)	Pléh Csaba	
Kozeki Béla	24 (22+2)	Csépe Valéria	
Forray Katalin	23 (22+1)	Farkas András	
Szegál Borisz	23 (18+5)	Vargha András	
Salamon Jenő	22 (21+1)	Pataki Ferenc	
Feuer Mária	21 (20+1)	Molnár Márk	

(The first number in parentheses is the number of self-authored articles; the second indicates the number of co-author publications.)

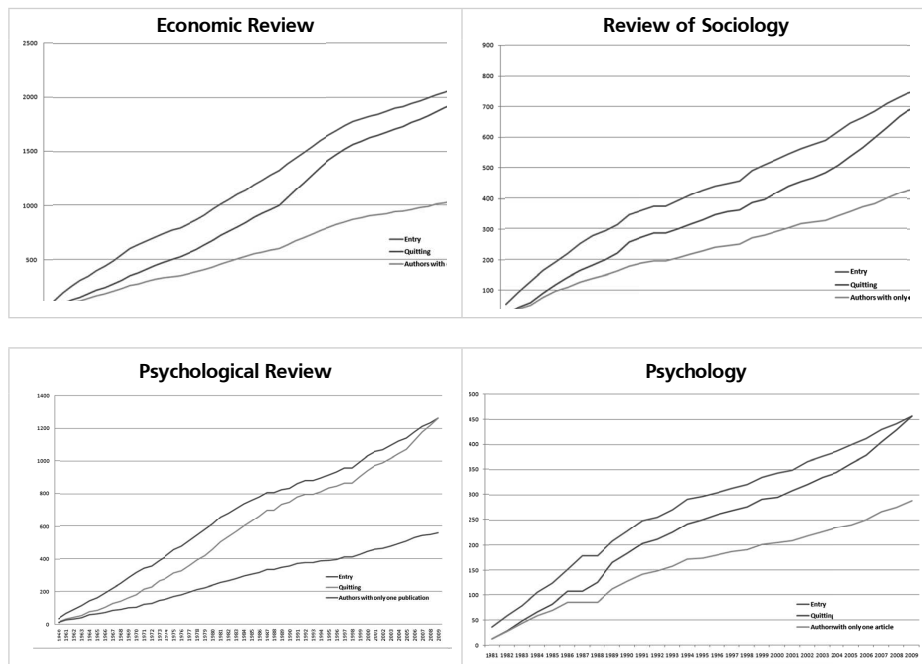
## Measurement of entering and quitting

In table two the dynamics of the emergence of new authors can be seen. New authors were considered as those who had not previously published in the journals studied. Similar to the newcomers I aggregated the numbers of authors who published their last article in that year. As I have previously demonstrated, the highest proportion is of the one-article authors for whom the entry and the exit date are the same, so I marked this group for all three journals. (The authors create a group in which everyone has two articles published in the same year and later did not have any. Despite the entry and the exit number is the same in their case; I considered them a separate group.) The proportion of authors with one article is 44 percent in the *Psychological Review*; which is 51 in the *Economic Review* and 59 percent in the *Review of Sociology*.

In all of the four journals the entry and the exit dynamics follow each other. The starting growth of dynamics in psychology and sociology diminishes and stabiliz-

es in eight to ten years after its first issue. In case of the *Economic Review* a decline can be observed in the dynamics of entry before 1965 and around 1990. In the *Psychological Review* the above-mentioned difficult period between 1987 and 1998 reduced the number of new entrants but then the previous growth rate returned.

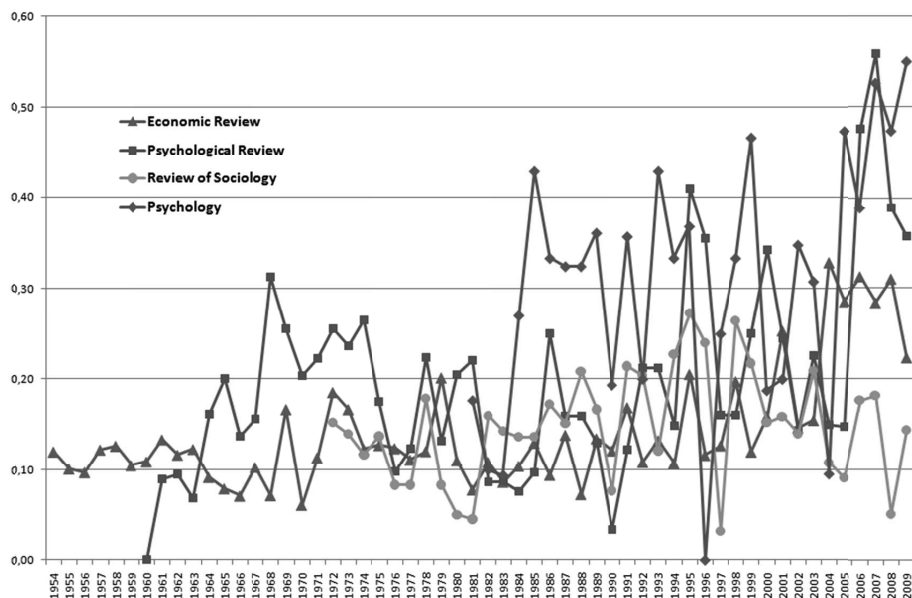
Figure 2: Measurement of entering and quitting



## The proportion of co-authors

The third table shows the annual rate of co-author articles. The proportion of co-author articles was around ten percent in the *Economic Review* from the start until the nineties (with smaller fluctuations), then it slowly decreased to twenty until 2000, and now this ratio is around thirty percent.

Figure 3: *The proportion of co-author articles*



In the *Psychological Review* thirty percent of the articles were written in co-operation as far back as in the sixties but in the mid-eighties and early nineties it fell with ten percent. A new growth started in the mid-nineties. In the last year co-authored articles represented more than fifty percent. (The increase is partly explained by the fact that thematic issues have been published from 2006 and more research results made in co-operation were accepted.)

The proportion of articles co-authored has no significant change from the start in the case of the *Review of Sociology*. The values range between five and twenty-five percent. In 2004 and 2005 only ten percent of the articles were co-author ones.

Besides co-operation rate the average number of authors is also useful information. Only in the case of the *Psychological Review* can changes be observed in the last few years. It reached a value above three many times during its history, what's more, it was 3.5 in the last year studied. Among psychologists, it can be stated that in recent years both the rate of co-authorship and the average number of authors has increased.

## The biggest co-author networks

The most interesting part of the research is the analysis of the networks of co-authors but it also raises several methodological questions. Two journals have more than fifty and one of them nearly forty years' history. If a chart shows the total connections – those who had published in the seventies but after died or those who have finished academic activity – it can be misleading as these researchers could not contact today's publishing authors because of their age. The investigation of the whole period is indispensable regarding the growth of networks. Therefore the following pages present the networks from the very beginning. I indicate the last decades with the data when the last article was published.

Table 2: Size of components

	Economic Review	Review of Sociology	Psychological Review	Psychology
<b>Biggest component</b>	<b>280 (27%)</b>	<b>74 (24%)</b>	<b>240 (34%)</b>	<b>69</b>
<b>2ndbiggest component</b>	<b>31</b>	<b>10</b>	<b>19</b>	<b>61</b>
<b>2 persons</b>	161	61	69	23
<b>3 persons</b>	41	13	16	11
<b>4 persons</b>	15	6	15	5
<b>5 persons</b>	9	7	10	5
<b>6–10 persons</b>	15	4	20	6
<b>11+ persons</b>	5	0	1	0

Two researchers can publish together more times as well. In the *Review of Sociology* 93 percent of the relations, in the *Economic Review* and the *Psychological Review* 89, in the *Psychology* 85 percent are simple-weighted relations (they have only one common article). In the field of sociology the biggest weight of a relation is seven (between Angelusz Róbert and Tardos Róbert), in the field of economics it is six (between Galasi Péter – Kertesi Gábor; Köllő János – Kertesi Gábor). In the *Psychological Review* and the *Psychology* the same pair (Márton Magda – Szirtes József) have published the most articles (eight in the former and seven in the latter.)

## Biggest networks

In the *Economic Review* the size of the biggest component is 280, which is 240 in the *Psychological Review* and 74 in the *Review of Sociology*. Two bigger components can be found in the *Psychology*: a network with 69 persons and a little smaller one with 61 people. The economist Balázs Sándor has the most connections within different fields (twenty-two), he is followed by the psychologist Pléh Csaba (thirty-four) and the sociologist Kolosi Tamás (thirteen).

Figure 4: The biggest network of the *Economic Review*

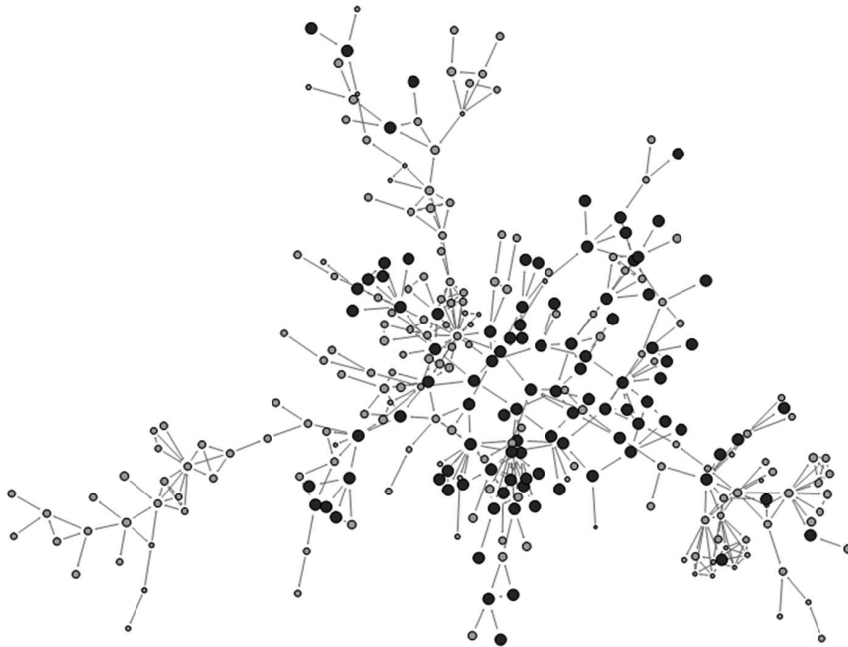


Figure 5: *The biggest network of the Psychological Review*

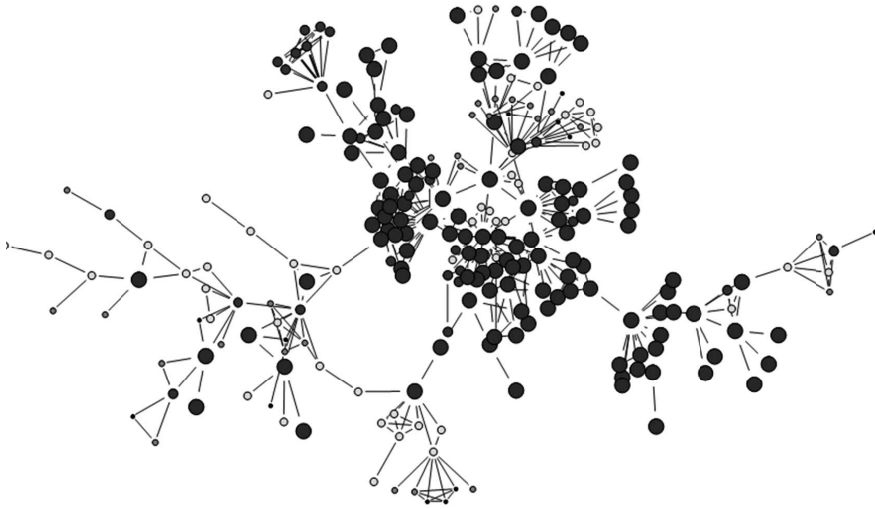
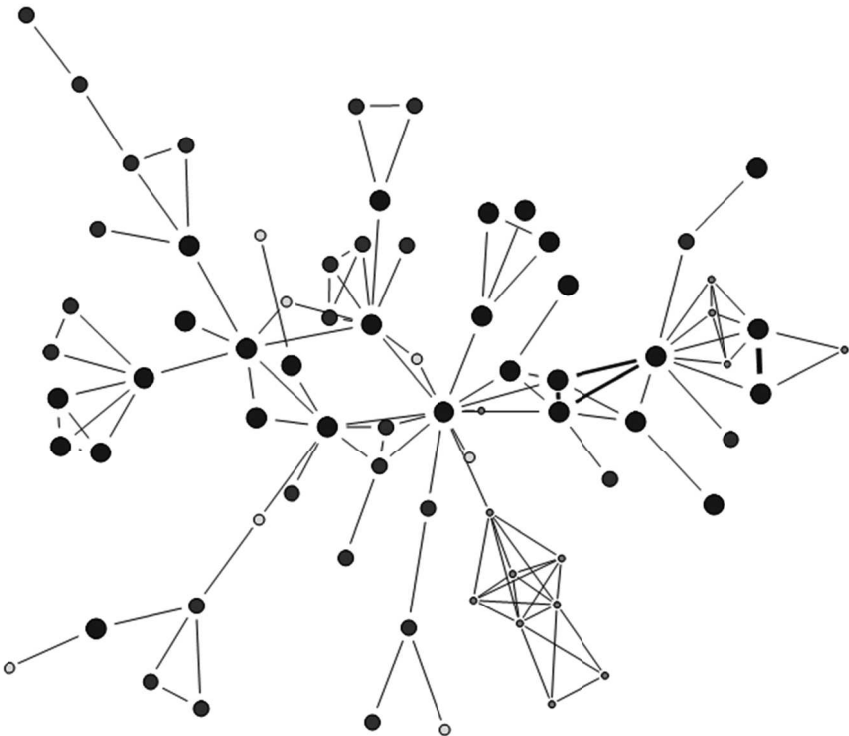


Figure 6: *The biggest network of the Sociology and the Review of Sociology*

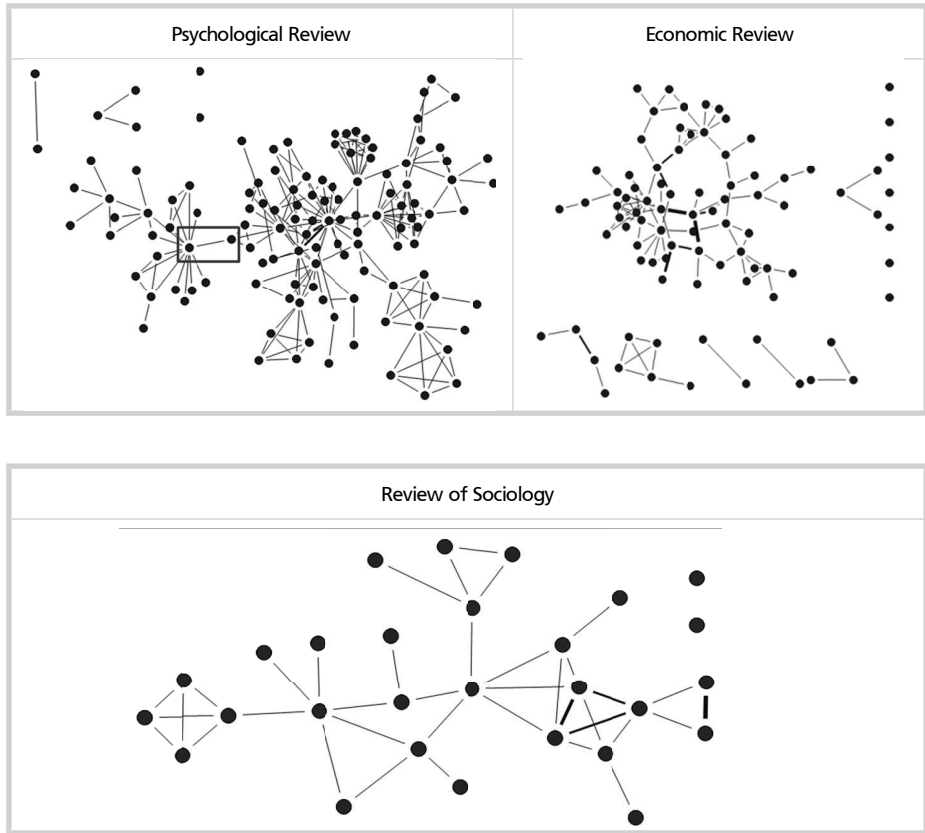


I mentioned that the networks presented here a forty- or fifty year span. This implies that authors who had published in an earlier period could not write together with researchers who entered scientific life later. The *following figures* show the decade of the last (single or co-authored) publications, their various sizes are marked according to the decade. In the case of all three journals it can be seen that publishing authors within the same decade are closer to each other. In order to ensure that these data are comparable, I collected them in *table 3* in various decades according to the date of the last publications. Networks based on the publications from the last ten years (2000–2009) can be viewed separately in the *seventh figure*.

Table 3: According to the last publications

	Economic Review	The Review of Sociology	The Review of Psychology
1950–1959	12 (4,3%)	–	–
1960–1969	20 (7,2%)	–	12 (5%)
1970–1979	30 (10,7%)	13 (18%)	28 (12%)
1980–1989	83 (29,6%)	7 (9%)	48 (20%)
1990–1999	35 (12,5%)	26 (35%)	28 (12%)
2000–2009	100 (35,7%)	28 (38%)	124 (51%)
<b>Sum</b>	<b>280</b>	<b>74</b>	<b>240</b>

Figure 7: The network of authors publishing between 2000–2009



It turns out from the networks made from publications over the past ten years that the largest network of co-authors can be found in the *Psychological Review*. The second is in the *Economic Review* and only the third is in the *Review of Sociology*. The filtered sub-networks based on the last publications (the last decade) show that while there are only six people in the *Psychological Review* who are not involved in the largest component, in the case of the *Economic Review* twenty-seven people were missed. A spectacular case of cluster can be observed, a connection (cutting point, rectangle marked in the figure) can be seen, if this was removed, the network would split in two parts. The left cluster is a good example of how a particular subject might be important. The key player here is Dull Andrea who is a recognized researcher in the area of environmental psychology and other researchers working on the same topic can be found around her.

Since from the area of psychology two journals can be found in the analysis, I examined their shared network, too. The common networks of the two journals



from 240 grew to 424. 176 authors from the *Psychology* (36.5%) can be found among the authors of the *Psychological Review*. A lot of interesting questions arise by merging the two networks. Both networks has a number of authors of ninety-five but only 59 people are members of the largest component of the *Psychological Review*, so only thirty-six persons become part of the largest component, if authors of both journals were taken into account (57 people are authors who just wrote in the Review and they were not previously members of the largest network). In the case of authors of both journals the average numbers of articles are evidently much higher than in the case of the other members of the network (average number of articles is 9.54 and 1.84).

## Analyzing the networks

During the descriptive analysis of networks basic centrality measures are used. The first indicator is density which is the rate of possible and completed connections as a percentage. (The statistical analysis of the networks was made by the program UCINET). The data consists of the strength of the relationship which slightly changes the method of preparation of metrics. Regarding density, this means that not the number of connections but the weight of the amount is allocated by the program, divided with number of possible connections. The average degree number shows the number of connections per capita, while the diameter indicates the largest distance that can be found between the members of the network. The geodesic distance shows the average distance within the network. This is calculated by connecting every node with each other and then dividing by the number of roads. The fifth indicator is the betweenness centrality value which provides the number of trips crossing the nodes of the network.

Table 4: Centrality measures

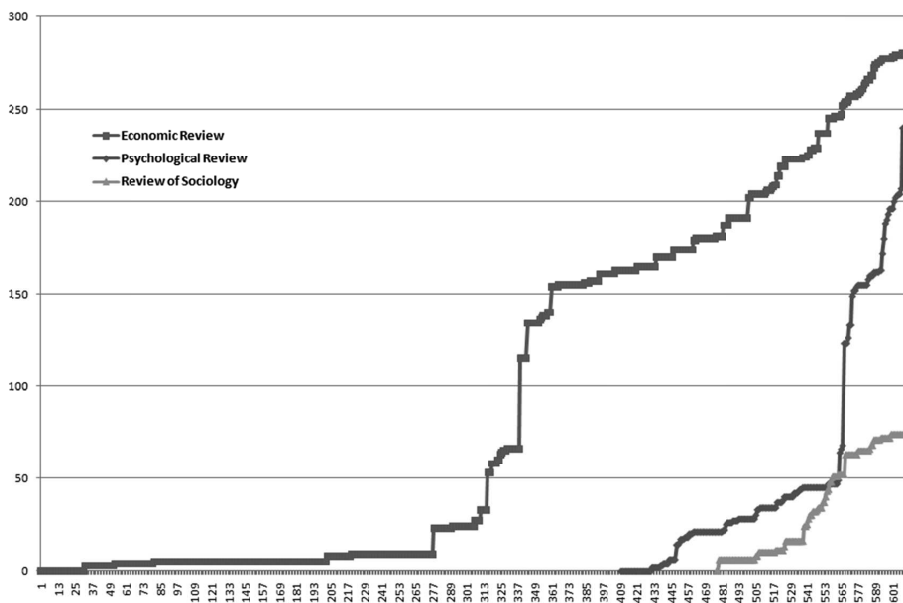
	N	Density	Average degree	Diameter	Geodesic distance	Betweenness centrality
<i>Economic Review</i> 1956–2009	280	0,011	3,193	23	8,136	0,026
<i>Hungarian Psychological Review</i> 1960–2009	240	0,016	3,767	18	5,94	0,021
<i>Review of Sociology</i> 1972–2009	74	0,044	3,180	10	4,38	0,047
<i>Economic Review</i> 2000–2009	68	0,046	3,08	10	4,8	0,058
<i>Hungarian Psychological Review</i> 2000–2009	117	0,035	4,05	10	4,439	0,030
<i>Review of Sociology</i> 2000–2009	26	0,117	2,93	7	3,3	0,103

## Evolution of networks

Network analysis can focus both on topography and on dynamics. The 8<sup>th</sup> figure shows the growth of dynamics of the largest networks. During the interpretation of the data, it should be noted that the growth of networks were adjusted to the latest issues released. As the three journals appeared first in different years and they are published in different months of the year, I did not want to sacrifice the details. So I chose it instead of the aggregation of growth data.

The network size can be read firstly from the diagram. This information is already known so I do not return to it again. The interesting part of the diagram is that the sizes of the network of the *Economic Review* and the *Psychological Review* show a sudden increase at the same size after a slow initial growth phase. The previous one grows from sixty-six to 114 and the latter grows suddenly from sixty-seven nodes to 120. In the case of the *Review of Sociology* a similar dynamic growth cannot be observed.

Figure 8: Evolution of the biggest components



A question arises during the analysis of the growing of networks: what kind of dynamics do networks follow? Social and natural sciences often describe the dynamics of networks with logistic growth (Fokasz 1999). In social sciences it can be used for analyzing demographic data, how widespread technological innova-

tions are (e. g. mobile phones). The canonization of Saints in the middle ages, the number of rockets sent to the Moon during the Cold War shows similar dynamics, more recently the dynamics of topics in the media is in the focus from this aspect (Fokasz 2004). The *Review of Sociology* also released a more detailed specification about this type of growth (Fokasz 2006), so I do not want to deal with it in more detail.

The use of logistic growth is based on the fact that different waves can be observed already from the raw data, while the entry data revealed that the appearance of new researchers is essentially linear. At the same time, however, I may not say that the model suits other scientific fields. In “mature” science, where a high degree of co-authorship and a higher average number of authors can be observed, other approaches may be used, as in the period under review.

Figure 9: The dynamics of growth of the *Review of Sociology* (left), the result of Fisher transformation (right)

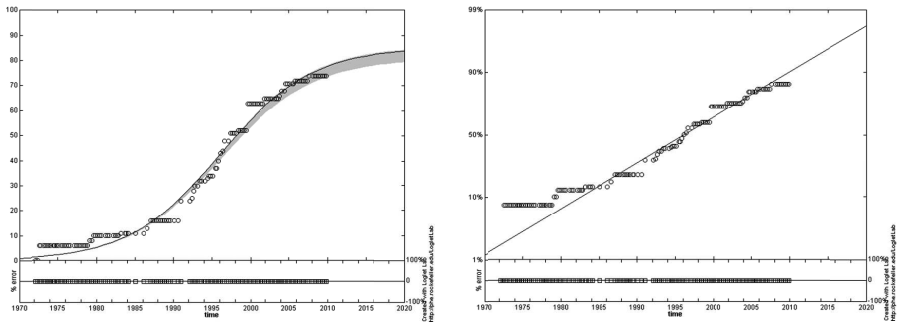


Figure 10: The dynamics of growth of the *Psychological Review* (left), the result of Fisher transformation (right)

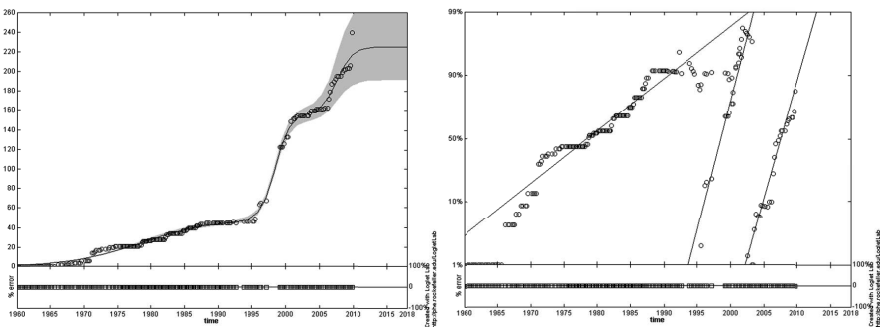
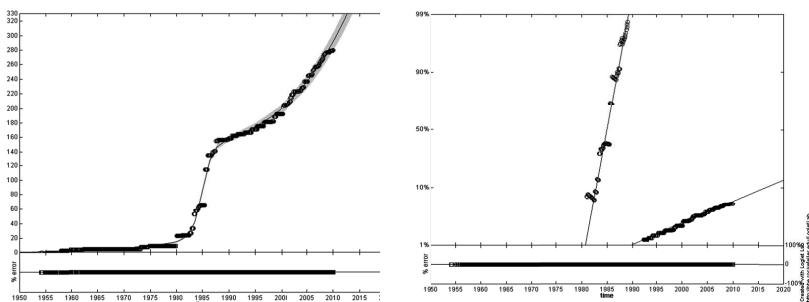


Figure 11: The dynamics of growth of the Psychological Review (left), the result of Fisher transformation (right)



Figures 9–10–11 show the dynamics of growth of the largest networks analyzed with the help of the logistic fit (this was made by the program Loglet Lab<sup>2</sup>). The figures on the left include the fittings, while the Fisher–Pry transformation results can be seen on the right side. (The essence of the transformation is that values of the logistic function are shown on a logarithmic scale which result a straight line. A relationship of successive waves of growth can be seen after data conversion.)

The fittings verified my conjectures; the dynamics of co-author networks grow logistically. More series of successive waves of growth: a bi- or tri-logistic growth is concerned in two cases. It can be assumed from the data in the 8th figure that if the network has a logistic growth, it has more series of successive waves of growth.

One growing wave can be fit on the largest network of the Sociological Review which is 90 percent full. Taking into account the data of recent years, it seems that the growth process is of a slow pace and even a few years are expected to pass until it reaches complete saturation. Considering that the number of co-author articles is low and there is no bigger-size network beside the largest component, we cannot expect major changes in the near future.

In the case of the Economic Review the fitting of two curves was necessary. On the basis of the current data it can be also shown that the increase of the largest network arrives at saturation point in the near future, however, this transitional period can take almost three to five years. The logistic growth is represented by consecutive waves; the slope of the second wave of growth is smaller than the first one. While interpreting the growth of the first wave, we should note that a sudden increase in size is due to the merger of parallelly growing smaller networks. But this does not appear during the analysis of the dynamics of growth; it seems that it was actually a single process.

The Psychological Review’s network analysis needed three logistic fittings. In this case the logistic processes do not follow each other sequentially but we can

2 The program can be found on the following site: <http://phe.rockefeller.edu/LogletLab/>

observe two logistic processes peaking at the same time and a start-up of a new wave can be seen, too.

Let's compare the growth of the largest network of the Economic Review and the Psychological Review. We can see that in the initial period, during which the growth of networks was due to similar-sized networks connecting to each other and substantially increasing in size, this happened in twenty-five or thirty years. In case of the Economic Review this dynamic growth began in 1985. I have already quoted the Such-Tóth article in the introduction – they stated that a significant change had happened in the mid-eighties when the Economic Review had become independent from politics. The first logistic growth saturated this time. Of course, I can't prove the relation between the two facts; a possible explanation could be policy revision with the professional aspect coming to the fore, which resulted in the fact that researchers who had had less publishing possibilities could benefit from the situation. However, further analysis should be needed to test the hypothesis. The initial growth period was approximately thirty years in the Psychological Review. The transitional period has not already finished and a dynamic new growth began in the meantime.

The similar length of the initial period is surprising particularly because the Economic Review is published eleven times a year while the Psychological Review only six times in its initial period. This points to the fact that not the number of the issues but the calendar time influences more the dynamics of growth of networks. I consider it important to mention that during this important period of evolving networks approximately 600 authors published in the Economic Review and 450 in the Psychological Review. The difference is significant so I think that besides time the growth of networks are determined by the number of potentially available authors besides time. The growth of the network of the Review of Sociology shows that it is still in its early days.

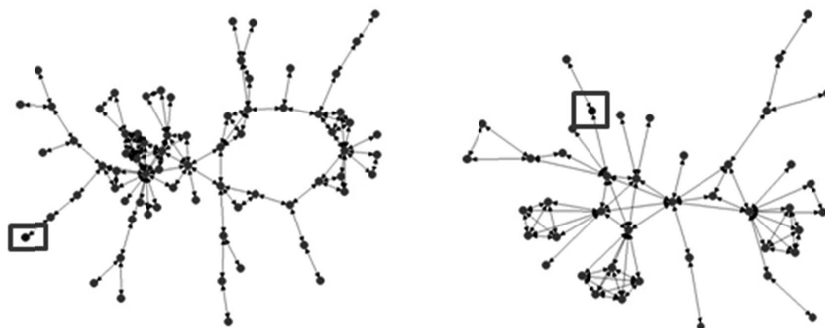
I return to the growth of the largest network of the Psychological Review for a while. The overall resolution of the initial phase passed through three locked wave of growth. Jumping a level in the analysis and considering different waves as the basic unit, it is possible to compare the duration of waves and their size (number of members of the network). I devote specific attention to the period of transformation. The network size of the Psychological Review has grown increasingly during the series of waves of growth. The first wave has increased with 18, the second with 26 and the third with 111 people. (The latter is debatable as it is not a continuous growth but another network is involved in the merger process.) The time of growth of the first two waves is fifteen to seventeen years and the transition period was approximately ten years. The Economic Review started with low initial numbers and suddenly there was an increase to one hundred and seventy. In the second phase (which is not complete yet) it continues to grow with approximately hundreds of people. The first and second stages are/were fifteen years.

## The structure of the networks before the transition

The points marked with a square connect the two networks. In case of the *Economic Review* the relations of certain authors merge the two networks: Gábor R. István's in the larger network and Kővári György's in the smaller network. Regarding the *Psychological Review* the cooperation of Csibra Gergely and Csepeli György connects the two networks. It is true for both journals – except for the member Csepeli György – that the collaboration of these networks is established by the cooperation of authors with few connections on the periphery (Gábor R. István and Csibra Gergely have one, Kővári György has two).

At this time Cziegler István can be found in the center of the larger network of the *Psychological Review*. B. Kakas Gizella, Kardos Lajos, Barkóczy Ilona or Klein Sándor have more connections but they are on the edge of the network. Popper Péter stands in the heart of the smaller network. His central role in this network is indisputable. Csepeli György and Füredi János have more connections but this does not change the fact that basically the network has a chain structure.

Figure 12: The bigger (left, N=66) and the smaller (right, N=48) part of the network of the *Economic Review* before the merging in 1985\*



\* Networks merge after the formation of the relationships between the points marked by squares

Figure 13: The bigger (left, N=67) and the smaller (right, N=53) part of the network of the *Economic Review* before the merging in 1999\*



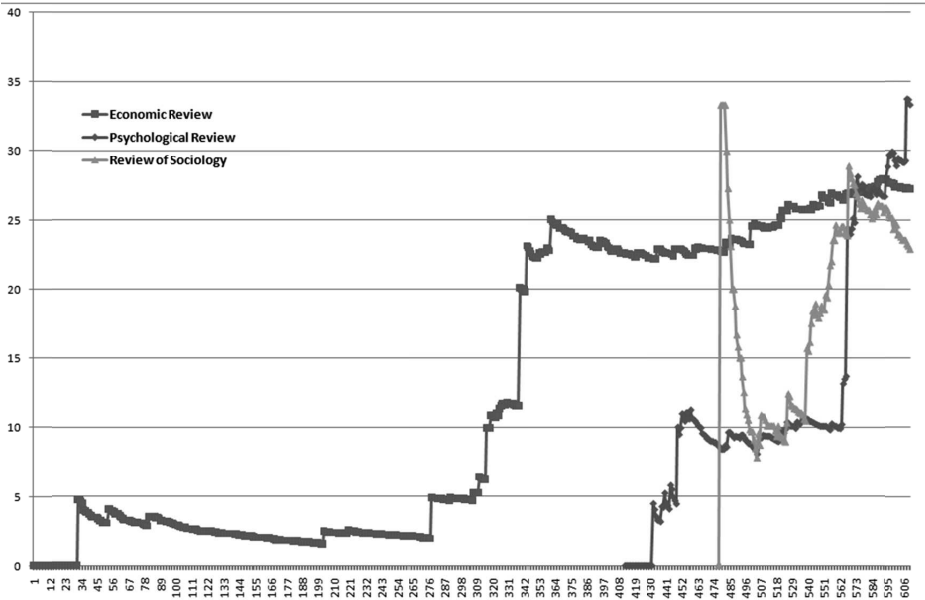
\* Networks merge after the formation of the relationships between the points marked by squares

## The place of the biggest components within the whole scientific community

The 14<sup>th</sup> figure shows the proportion of researchers who published their articles in co-authorship within the largest networks. The sizes of the three networks are significantly different from each other thus it is surprising that about twenty-five to twenty-eight percent of the co-authors are integrated in all three cases.

In the case of the *Economic Review* and the *Psychological Review* it can also be observed that they show not only a similar dynamics of growth, but also a similar dynamics of the number of authors integrated in their largest networks. Initially the first two networks include five percent of the authors, which doubled to ten and then it stagnates around twenty-five percent.

Figure 14: The proportion of the largest networks compared to the number of writers of co-author articles



## Inter-journal connections

These journals focus on the fields of social sciences but is there a connection between them? Only Csepeli György, Garai László and Varga Károly published their work in all of the four papers. The number of the common publishers of the *Review of Sociology* and the two psychological journals is eleven; this number is seven regarding the *Economic Review* and the two psychological journals.

The number of pairwise connections can be found in table five. The strongest link can be observed between economic and sociological journals, and between psychological ones. (The numbers are somewhat biased upwards, several common names can be found in the overlapping relationships among which it is conceivable that different persons are behind the names.)



Table 5: Number of researchers who have published in two journals

	Economics Review	Review of Sociology	Psychological Review	Psychology
Economics Review	–	150	25	9
Review of Sociology		–	26	7
Psychological Review			–	146
Psychology				–

In the case of the biggest networks except for the *Economic Review* and the *Psychological Review* there is a connection. Between sociology and economics the brokers are: Kertesi Gábor (two in the *Economic Review* and five in the *Review of Sociology*), Kiss István (five and two), Kovách Imre (seven and four), Kuczsi Tibor (four and three), Sik Endre (two and seven). The *Review of Sociology* and the *Psychological Review* have also five common members: Csepeli György (six in the sociological and eight in the psychological one), Kóczán György (seven and seven), Neményi Mária (four and two), Rudas Tamás (two and one), Váriné Szilagyfi Ibolya (one in both journals).

The brokerage role of sociology between the different fields of social sciences is consistent with the results of Moody (Moody 2006). In his citation network analysis of social science journals he found that economics, law, political science and psychology have the most stable boundaries and they are seated on the edge of the network, while sociology journals refer less to themselves so they can be found in the middle of the network. (The American Sociological Review is the most central journal and also seven from the ten most central reviews belong to sociology.) Thus I was led to the conclusion that sociology can convey ideas between different disciplines and it can easily borrow new ideas as well.

## Summary

I focused on two aspects of co-author networks in this article: topography and dynamics. The results showed that the level of cooperation in domestic social science is very low in comparison with the American which is twice as much. This raises several interesting questions: what can influence the low cooperation rate while it is clear that scientific work can benefit from collaboration? Is it due to the individualism of society, or other structural causes stand behind this?

The question remains unanswered in this article as the methodology used here is inappropriate to find this out.

The analysis revealed that logistic curves can describe the growth of the largest components of co-author networks effectively; what's more, it is possible to fit more curves on the networks. In the case of two or three waves we can talk about bi-logistic or tri-logistic growth. Regarding the *Review of Sociology*, however, only one wave can be identified. If we extrapolate this wave, it can be seen that a significant change cannot be expected in the coming years in the level of cooperation. It is a major and serious recognition for sociology that it has a linkage role between psychology and economics, so in this sense it has the same role as in American science in this sense.

The dynamics of networks was approached several ways but there are still more possibilities for further analysis. The centrality measures presented here can be applied not only for a moment, but also for all moments of the life of networks. This becomes particularly interesting if both entering and leaving the network become part of the analysis. How does dynamic look like? Does network have a state of balance? What processes can be observed when an important person leaves a network? However, these questions could be interesting for scientific networks and for network studies, too. As for co-author networks there are further possibilities of analysis if additional data are available. These mean for example: connecting the network results with citation- and topic networks, but we can also focus on the situation<sup>3</sup> of women and men within science, too.

The analysis of co-author networks shows the skeleton of the working process of science. This is one of the different levels of cooperation which connects researchers, where new ideas are born and various thoughts spread. I believe that creativity is a social phenomenon which will provide several new results during the research of innovation.

## References

- Amaral, L. A. N. – Scala, A. – Barthelemy, M. – Stanley, H. M. (2000): Classes of Small-world Networks. *Proc. Natl. Acad. Sci. USA* 97: 11149–11152.
- Barabási, A. L. – Jeong, H. – Ravasz, E. – Neda, Z. – Schuberts, A. – Vicsek, T. (2002): Evolution of the Social Network of Scientific Collaborations. *Physica, A* 311, 590–614.
- Barabási, A. L. (2003): *Behálózva*. Budapest: Magyar Könyvklub.
- Beaver, D. deB (2001): Reflections on Scientific Collaborations (and its study): Past, Present, and Future. *Scientometrics*, 52(3): 365–377.
- Bernal, A. – Llanos, E. (n.d.): *Redes de coautoría en la ciencia colombiana: 1999–2004*. [http://www.ocyt.org.co/esocite/Ponencias\\_ESOCITEPDF/2COL030.pdf](http://www.ocyt.org.co/esocite/Ponencias_ESOCITEPDF/2COL030.pdf) (last update: 2011. January).
- Cotta, C. – Merelo, J.-J. (2005): The Complex Network of Evolutionary Computation

<sup>3</sup> I wrote about the network positions of women and men in these three journals in a previous article. (Micsinai 2007).

- Authors: An Initial Study. <http://www.arxiv.org/abs/physics/0507196v2> (last update: 2011. January).
- Cronin, B. (2001): Hyperauthorship: A Post-modern Perversion or Evidence of a Structural Shift in Scholarly Communication Practices? *Journal of the American Society for Information Science and Technology*, 52 (7): 558–569.
- Cronin, B. – Shaw, D. – Barre, K. L. (2003): A Cast of Thousands: Coauthorship and Subauthorship Collaboration in the 20th Century as Manifested in the Scholarly Journal Literature of Psychology and Philosophy. *Journal of the American Society for Information Science and Technology*, 54(9): 855–71.
- Cselényi, E. (2009): Publikációs és idézettségi aktivitás a Politikatudományi Szemle hasábjain. *Politikatudományi Szemle*, 3.
- Dorogovtsev, S. N. – Mendes, J. F. F. (2000): Evolution of Reference Network with Aging. *Physical Review*, E 62, 1842.
- Fokasz, N. (2006): Növekedési függvények, társadalmi diffúzió, társadalmi változás. *Szociológiai Szemle*, 3.
- Glänzel, W. – Schubert, A. (2004): Analyzing Scientific Networks Through Co-authorship. In: Moed, Henk F. et al. (eds.): *Handbook of Quantitative Science and Technology Report*. Kluwer Academic Publishers, 257–276.
- Hou, H. – Kretschmer, H. – Liu, Z. (2006): The Structure of Scientific Collaboration Networks in Scientometrics. In: *Proceedings International Workshop on Webometrics, Informetrics and Scientometrics & Seventh COLLNET Meeting*, Nancy (France).
- Kürtösi, Zsófia (2004): Nők a felsővezetésben: hozhat-e újat a kapcsolathálózati megközelítés? *Szociológiai Szemle*, 2.
- Laviriére, V. – Gingras, Y. – Archambault, E. (2004): Comparative Analysis of Networks of Collaboration of Canadian Researchers in the Natural Sciences, Social Sciences and the Humanities. *Scientometrics*, 68(3): 519–533.
- Lehmann, B. – Lautrup, B. – Jackson, A. D. (2006): Citation Networks in High Energy Physics. *arXiv:physics 02110210*.
- Micsinai, I. (2007): Nők és férfiak tudományos hálózatokban. *Magyar Tudomány*, 5.
- Moody, J. (2004): The Structure of a Social Science Collaboration Network: Disciplinary Cohesion from 1963 to 1999. *American Sociological Review*, 69: 213–238.
- Moody, J. – Light, R. (2006): A View from Above: The Evolving Sociological Landscape. *American Sociologist*, 2: 67–86.
- Mutschke, P. – Quan Haase, A. (2001): Collaboration and Cognitive Structures in Social Science Research Fields. *Scientometrics*, 52(3): 487–502.
- Newman, M. E. J. (2000): The Structure of Scientific Collaboration Networks. *Proc. Natl. Acad. Sci. USA* 98, 404–409.
- Newman, M. E. J. (2001a): Scientific Collaboration Networks. I. Network Construction and Fundamental Results. *Physical Review*, E 64, 016131.

- Newman, M. E. J. (2001b): Scientific Collaboration Networks. II. Shortest Paths, Weighted Networks, and Centrality. *Physical Review*, E 64, 016132.
- Reisinger, A. – Tóth P. (2007): Citációs hálózatok a Tér és Társadalom folyóiratban. *Kistáska*, 40–41.
- Such, Gy. – Tóth, I. J. (1989): A magyar közgazdaság-tudomány a Közgazdasági Szemle tudományometriai vizsgálatának a tükrében. *Közgazdasági Szemle*, 10.
- Uzzi, B. – Spirro, J. (2005): Collaboration and Creativity: The Small World Problem. *American Journal of Sociology*, Volume 111, Number 2: 447–504.
- Yoshikane, F. – Nozawa, T. – Tsuji, K. (2006): Comparative Analysis of Co-authorship Networks Considering Authors' roles in Collaboration: Differences between the Theoretical and Application Areas. *Scientometrics*, 81(2): 499–511.
- Wuchty, S. – Jones, B. F. – Uzzi, B. (2007): The Increasing Dominance of Teams in Production of Knowledge. *Science*, 316(5827): 1036–1039

# Instability and Constraint<sup>1</sup>

## Interdisciplinary Essay on the Origin of Social Innovations

Nikos FOKAS

fokata.bt@chello.hu

**ABSTRACT:** Within the framework of the present paper we are interested in the circumstances of the origin of social innovations. Presenting and analysing the interdisciplinary literature we may conclude that there really exists a general model in the circumstances of the origin of significant innovations impacting on the whole of some level of hierarchic system. We found that innovations may emerge amid the dissolution, disintegration anarchy, crisis, anomaly, or revolution. In order to remove the political connotations of the above mentioned terms I propose the term instability as a collective for these conditions. It is noteworthy that complex systems may be permanently in state of instability, which may serve as a fundamental source of endogenous creativity in the system.

### Introduction

Only the alteration of something leads to change (Hegel 1979). This *alteration* may be the birth of something new or it may be the adoption of an earlier innovation at a specific moment and its diffusion, in this way. Birth or diffusion? Concrete categorization is often merely a question of viewpoint. The present essay postulates that the emergence and diffusion of social innovations are two closely related constituents of social change, often hard to separate even by analysis, which nevertheless must be distinguished conceptually.

The investigation of innovation diffusion as an autonomous research program began with the study of the diffusion of *technological* innovations. The empirical works (Griliches 1988 [1957]), which for a long time determined the direction, nature, and methodology of later research orientation, were published in the fifties and the first comprehensive works of academic standard followed only a decade later (Rogers 1962; Coleman et al. [1966]). After a half-century long accumulation, the study of the diffusion of technological innovations can be said to have come to constitute an autonomous, self-contained, as well as both theoretically and meth-

<sup>1</sup> The study is based on research carried out at the Peripato Research Group founded within the framework of OTKA research no. K73033.

odologically fully developed field of research (Geroski 2000).<sup>2</sup>

At the same time, social innovation may obviously be not only a technique embodied in some object, but also any new idea, individual behavior, or collective action. It is noteworthy that the standpoints and technical procedures worked out for dealing with the diffusion of technological innovations were extended, with unusual effectiveness considering the heterogeneity of the new fields involved, to the analysis of the most diverse individual and community learning and modeling processes. As a result, today, concrete investigations of the diffusion processes of social innovations encompass unrelated and very diverse areas ranging from demography to media dynamics (Marchetti 1986, 1997; Marchetti et al. 1996; Mod-elski–Gardner 2002; Konstandopoulos–Modis 2003).

The present essay intends to focus on the other, much cruder aspect of social change, namely, the origin, rather than diffusion, of innovations. I know of no accepted methodology or theory that may be applied here, and I have none to offer. I merely aim to bring forward a general scheme of the circumstances that make possible the birth of certain – and not just social – innovations. For this reason, although my primary aim concerns the understanding of the origin of *social* innovations in order to convincingly substantiate the existence of such a general model with the necessary number of examples and, perhaps unacquaintedly, harmonizing standpoints taken from the most diverse fields of research, I must, by necessity, take an interdisciplinary approach.<sup>3</sup>

In my opinion, the “*creative anarchy*” metaphor gives a very plastic description of the circumstances enabling the birth of the most diverse social, organizational, scientific, technical, and even biological innovations, while the *instability* and *constraint* metaphor pair offers a rhetorically less impressive, but somewhat more exact picture.

The choice of the essay form is an indication mainly of the fact that in two senses I still cannot tell precisely where the limits of the scope of validity of the foregoing statement lie. I am convinced that the range of effect of the model far exceeds the province of the social sciences, on the one hand, and its validity for every social innovation is far from certain, on the other. I do not have, and, left to myself, perhaps I will never have, a more precise answer to the problem. The perspective of the essay as well as my answers to the countless questions and problems arising in the course of writing it may also be arguable; hopefully, also in the sense that they prove worthy of argument. This is why I have decided to bring the study to the critical attention of the professional public.

2 Researches in economics are particularly advanced. Already in 1930 Simon Kuznets recognized that technological change advances in the form of an S curve, and the Rostovian take-off model also suggests this type of growth pattern. Joseph Schumpeter to date has a following in the field of innovation research. Next to him William Baumol is already regarded a living classic. However, I have deliberately avoided references to the literature on economics that would fill a library. Instead I propose to show that this problem complex – though perhaps with a different perspective and terminology than common to economists – is also present in the literature of broadly understood social science which is not directly concerned with the question of innovations.

3 Naturally, it is by no means self-evident that the examples from different disciplines could not have negligible common characteristics.

## Creative tension

In investigating the circumstances of the origin of social innovations, we start out from the microsociological position, according to which, there exists in social reality a special level described as the meeting of the individual and the social. This is the province wherein “social influence reaches and compels the individual to modeling”, who in turn, “by following or altering or perhaps rejecting the model, by accepting, modifying or averting the influence, reacts upon the model itself and everyone else party or witness to the event” (Méri 1996: 10).

I am convinced that ultimately this *tension*<sup>4</sup> – inherent in the meeting of the social and the individual, in the inseparably connected union of modeling and alteration – is the source of all social innovations and, even though unintentionally, makes this microenvironment the primary, permanent source of social innovations.

The overwhelming majority of the innovations arising there are but tiny and, even from the point of view of the direct participants in the event, no more than insignificant adjustments to the world. Other innovations may be more important, but even the most commonplace situations offer numerous examples of how much it depends on the recipient medium<sup>5</sup> whether they will cross the boundaries of the microenvironment at all.

This also means that the delimitation of the origin and diffusion of some innovation is mostly a question of perspective. Viewed from the outside, an innovation crossing the boundaries of a microenvironment appears as newly emerging, but this had to be preceded by diffusion within this environment first. As mentioned above, innovation diffusion is not the subject of the present paper; moreover, I have already discussed the question previously (Fokasz 2006). Therefore, I shall present only a few obvious but all the more descriptive examples to illustrate the relationship between the recipient medium and diffusion.

In the early seventies, I was present when in a number-theory seminar at the Budapest university, ELTE, a first-year mathematics student presented a new and extraordinarily elegant proof of some well-known theorem. A few days later, the professor already included this proof in his lecture, naming the student, of course.

Mihaly Csikszentmihalyi also mentions a case, similar in nature but greater in volume. A few years ago, a PhD student in physics in Munich came forth with a new idea and formula, news of which spread among the physicists at German universities within a week, and aroused the interest of colleagues on the western shores of the US by the end of the second week.

The above examples are noteworthy particularly because Csikszentmihalyi

4 In the wake of Ferenc Mérei, I use the term *tension* in the simplest possible everyday sense. At the same time, one of the main assertions of the essay appears latently here for the first time. Clearly, some kind of a tension – by now as a parameter of the description of the state of a dynamic system – may contribute to the creation of the unstable dynamic state, which the present essay will show to be the precondition for the origin of innovations.

5 Clearly, the nature of the recipient medium fundamentally influences the diffusion of an innovation. It is also well-known that in the case of technological innovations the literature on this subject matter is vast. Its analysis, however, falls outside the scope of this study.

adds that this could never have happened in his own chosen field of psychology. "If a student were to stand up in a psychology seminar in any school in the world and set forth the most profound thoughts ever, it would not cause a stir outside the walls of the classroom" (Csikszentmihalyi 2008: 49).

István Hajnal described Australia's indigenous society as a medium resembling psychology from the point of view of innovation diffusion, where "whatever individual thought arises, whatever individual skill creates, it gets lost, as it were, in the sand drift-like loose and inflexible population organization" (Hajnal 1988: 19). Fernand Braudel gives a similar description of the desert societies of the Sahara that seem from afar as merely a "handful of human dust blowing in the wind" (Braudel 1996: 176). I cannot attest to the objective exactness of these statements, but that is not the issue here. What is of interest is the obvious presence of this perspective in scientific areas far removed from the specialized fields of innovation research.

Thus, it stands to reason that in this special province of the meeting of the particular and the general, where "the general and the particular are present as expectation and emotional tension, as model and impression, as norm and choice, as the imperative and need" (Mérei 1996: 10), not only this "exuberant, bustling chaos" (Koestler 1967: 191) of flashing innovations, but also their dying belong among the most commonplace phenomena of social existence. For instance, historians of economics and science know well that the innovative work focusing on minor details, routinely carried out at technological or scientific workshops hidden from the outside world, is the prerequisite for making a truly new product or scientific discovery.

All this closely resembles the way in which the biological equivalents of creative social acts in evolution, namely, the random changes in the gene struggle for survival. "Atomic changes occur continuously" (Koestler 1967: 191). At the same time, "the overwhelming majority of the ensuing changes are temporary ... leave no trace on the operation of the whole" (Koestler 1967: 191). Relatively few mutations, also influencing heredity, reach the higher levels of the hierarchy. "Between the ensuing chemical changes and the appearance of the final result on the stage of evolution ... lies the full hierarchy of internal processes, which exercises tight control over the province of possible mutations" (Koestler 1967: 175).

Although all this is more or less self-evident, the practical reason for the special emphasis is that henceforth not all innovations will be of interest to us. Offhand, hoping to be more specific later on, let me just say that within the framework of the present essay we are interested in the circumstances of the origin of innovations which successfully pass the immediate environmental filters and bring about not only significant changes, but even big leaps. If such things exist. This is precisely what we shall try to find out in the following.



## Evolutionary big leaps?

The possibility of big leaps in history is well illustrated by Braudel's remark according to which, "the past has been richer in catastrophes and brutal revolutions than in slow evolution" (Braudel 1996: 88). Although, obviously, a statement like this should not be taken literally, it definitely indicates how naturally the possibility of historical big leaps is present in historiography. World War I will serve as a recent example, which "changed everything in the life of Europe: borders, regimes, mental attitudes, and even morals. It threw the most glorious of modern civilizations into such deep disarray that everything changed in its wake" (Furet 2000: 40).

However, "wars, the breakdown of encompassing systems of morality, political revolutions" (Feyerabend 1993: 16) transform not only behavior patterns, but the more important pattern of argumentation as well. No wonder it left its mark on the history of science. It is common knowledge that Thomas Kuhn conceived the development of science through scientific revolutions, so-called paradigm shifts, in the course of which the whole world view of scientific communities undergoes radical change. In the course of these science obviously develops by way of big leaps. In Paul Feyerabend's interpretation: "the history of science will be as complex, chaotic<sup>6</sup>, full of mistakes, and entertaining as the ideas it contains, and these ideas in turn will be as complex, chaotic, full of mistakes, and entertaining as are the minds of those who invented them" (Feyerabend 1993: 11).

Although the term *evolution* is mostly used as a synonym for gradual change, it is well-known that "there occur in biological evolution periods of crisis and transition when there is a rapid, almost explosive, branching out in all directions, often resulting in a radical change in the dominant trend of development" (Koestler 1964: 226). Accordingly, evolution based on natural selection also knows the possibility of big leaps, and the "Paleocene growth of the human brain is an excellent example of what we call the evolutionary explosion"<sup>7</sup> (Koestler 1967: 354).

John Maynard Smith and Eörs Szathmáry went even further when, instead of listing the specific big leaps, they distinguished two steps, also distinct in point of principle (Maynard Smith–Szathmáry 1997). In the first case, the information itself – stored in the genes – is modified via almost daily mutations, while in the course of the far less common *major transitions* "the modes of information storage and transmission also change". The emergence of multicellular organisms,

6 I confess, I pounced upon this passage because of the word *chaos*. The situation with this word is the same as it was in the case of *tension*, and will be in the case of *instability*. I use all three words mostly in their everyday sense, but in the theory of dynamic systems they may be attributed strict and mutually related meanings. However, I have been unable to explore within the framework of the present study the consequences of this fact with regard to the circumstances of the origin of social innovations. Therefore, let me just note that though the use of these terms may, for the time being, seem to be mere rhetoric, there is much more to it.

7 "According to the evidence of fossils, the growth of the brain of hominids ... started on an unparalleled acceleration course about half a million years ago; the tempo far exceeded the speed of all observable anatomical changes in lower animals." (In: W. E. Le Gros Clark: *The Advancement of Science*. London, 1961, cited by Koestler 2000: 354.)

the appearance of animal colonies, or most recently the developments of human language are examples of major transitions in evolution. What is common to them is that the entities previously capable of autonomous replication thereafter could replicate only as part of some larger entity.<sup>8</sup>

It is noteworthy that according to Szathmáry and Maynard Smith, there is nothing inevitable in the occurrence of these major transitions. Moreover, they are exceedingly rare. The authors believe that in the several billion-year history of life on earth there were only eight instances. At the same time, they play a determinative role in the emergence of the astonishing forms of adaptation of living organisms. In point of fact “the theory of evolution by natural selection does not predict that organisms will get more complex... Empirically, many and perhaps most lineages change little for many millions of years” (Szathmáry–Maynard Smith 1999: 15). Nor did the major transitions take place in order to enable the evolution of complexity. Yet, the ensuing new coding procedures are ultimately responsible for the appearance of increasingly complex organisms and the rise of what biologists reverently term the “wonderfully adapted organisms” (Szathmáry–Maynard Smith 1999: 1).

As I have already mentioned, instead of the thoroughly explored question of the diffusion of innovations, I would like to discuss the much more complex and much less known problem of their origin. However, I cannot, already at the start, specify with a definition-like accuracy those innovations whose origin I would like to explore. As I have indicated, I have no intention of dealing with each of the minute modifications, alterations that keep on bubbling up and almost as frequently bursting in the social microenvironment. I prefer to study those with greater influence, significance, though not necessarily of the magnitude of Maynard Smith’s and Szathmáry’s major transitions.

First I would like to indicate, rather than define, what I have in mind. These are innovations of the type and nature which lead to new paradigms in science, give rise to new species in biology, and result in new technologies in the economy.<sup>9</sup>

If we wish to be more specific as to the kind of changes, that is, the type of innovations, whose origin we are interested in, then we can build on the fact that most social, biological, physical systems, as well as man-made symbolic systems are comprised of interrelated subsystems. These systems are generally known as complex hierarchic systems (Simon 1982). Below I seek to answer what the preconditions may be for the realization of innovations leading to changes at the

8 Social scientists, too, may find it informative to learn the answer evolutionary biologists offer concerning the question of why the emerging more complex entities are not taken apart by lower-level interests. Its investigation, however, falls outside the scope of the present paper.

9 Let me note that *macro-invention* is a well-known designation in the history of technology. Although the usage is strongly reminiscent of the big leaps I have studied, they are not really the same. Macro-invention designates discoveries where “origin cannot be precisely determined, there is a clear break compared to previous techniques” (Mokyr 2004: 400). The designation refers primarily to antecedents rather than consequences. In this sense the airship is a macro-invention, but the railway, which transformed our world, is not. Compared to these, World War I, the “event colossus” (Nora 1974) of the twentieth century is more likely to hold our attention, because it was not only “bigger than its causes”, but may also be considered as the “pristine state of a new epoch” (Furet 2000: 57).

system's level – or involving those touching on any of the subsystems – at the level of the system – or corresponding subsystem – impacted upon by the change itself.<sup>10</sup>

Thus we have come to the central problem of the essay. We have seen above that big, that is, dramatic changes affecting the whole of a system (subsystem) – to keep the somewhat more accurate terminology – do exist. The question is: How are they possible?<sup>11</sup> The fact being that their existence is not so obvious in spite of the reality of the big leaps, since a complex system, a hierarchy with built-in self-control safety mechanisms is quite stable; it cannot be simply let out here and tightened a little there (Koestler 1967: 178). On the whole, it may be said that in complex systems – due to resistance of other parts of the system – dramatic, sudden changes are unlikely because of the necessity to maintain the compatibility of the constituent parts (Mokyr 1990: 407).

Except in the event that the complex system itself collapses. This happened in the case of the Western Roman Empire, opening the way to the emergence of the western-type of organization of society, one of the truly original innovations in the history of civilization.

## On the question of the nature of the West as innovation

I have neither the wish nor the ability to discover even a single new historical fact in a field already subjected to very close scrutiny by historians, economists, and sociologists. Nor do I propose to rearrange known facts. In the following two chapters I shall only endeavor to put in a new light the well-known doctrines of an existing well-known school.

In order to detect the original characteristics of the West as an innovation, we start out from Jenő Szűcs' epochal study (Szűcs 2006). The reason for this is the author's repeated declaration with definitive clarity that "what sharply delimits the medieval West from other civilizations is the development of the autonomous concept of society" (Szűcs 2006: 45); more concretely, "the separation of 'society' and 'state', in point of fact a structural change ... or to be more precise, a series of structural changes, in which this duality was present" (Szűcs 2006: 36). Moreover, as Szűcs emphasizes, this "separation is not really an endogenous feature of human history" (Szűcs 2006: 36). Although "every state is built on some society, it is by the gravitation inherent in the high cultures of five thousand years that the emerging state finds its own legitimization 'outside' of society, and consequently develops a device and operational mechanism, in which the society appears as a derivative of the state, and not the other way around. The autonomous existence of any of the sectors of society separate from (and at the same time functionally

10 With this our usage obviously becomes more accurate, compared with the phrase *big leap* used so far. But the change in usage has important consequences. Namely, use of this phrase as well as a part of the cases elaborated in the paper may lead the reader to believe – as I believed when I began to write this – that we are concerned exclusively with the logic of macro-social changes. However, we do not wish to distinguish among the various levels of the hierarchic systems. Consequently, the preconditions of the emergence of innovations inducing changes on any level of the hierarchy may claim our interest.

11 The fundamental question in ethics is: Good people exist – how are they possible?

connected to) the state is the rare exception" (Szűcs 2006: 36). Exceptions were such luxury products of history as the Greek polis, the original model of the autonomous society.

In view of the foregoing, Szűcs obviously considered the separation of state and society to be the most original innovation of the western-type of organization of society. Existence of the autonomous society implies a certain limitation of the state, that is, of political power. Two further characteristics of the West, closely related to the separation of state and society, indicate the same limitation, namely the separation of secular and spiritual powers, and of political and civilizational integrations.

The former stems from the period following the fall of the Western Roman Empire, when "during the political chaos and vacuum the Western Church escaped from the subordinate position it naturally occupied in late antiquity ... since Constantine (337), and which was subsequently speedily reproduced by the Byzantine Justinianus (532). Separation of the spiritual and the secular, the ideological and the political, spheres is one of the uniquely fertile separations in the West, without which the future 'freedoms', the abstract emancipation of 'society', the nation states to be, the Renaissance, and the Reformation are all inconceivable" (Szűcs 2006: 39–40).

As it is well-known, there never existed a state or imperial formation which would have set the western civilization in a uniform state framework. Charlemagne's empire was perhaps the only western attempt to resolve the synthesis of antique and barbarian elements in the "usual" manner, that is, by correlating the concept of "civilization" with an "imperial", or political, integration. The empire exhausting its last reserves and, with it, this attempt were destroyed by the same element "with which Charlemagne tried to balance the statics shaky from the very first, namely, vassalage. This brought to conclusion the final separation of 'civilization' and 'political framework' in the West" (Szűcs 2006: 40).

Vassalage, appearing for the first time in our analysis, represents the fourth constituent of European society. In itself there is nothing special about the personal dependence characterizing vassalage. "The form is known to every prefeudal society, even the nomads, its network is the binding material of every feudalism. ... What first of all distinguished western vassalage from kindred structures was that it almost fully absorbed into the system all social elements left free-floating after the social erosion, on the one hand, and, instead of taking a position beside or below the state, it virtually replaced the state, on the other; consequently, it virtually substituted the 'state' formula with 'social' relations" (Szűcs 2006: 41).

Its "particular 'contractual' character" is another specific feature of western vassalage. This bond is the "relationship of unequals in the name of contractual reciprocity with bilateral obligations: this fundamental endogenous feature of western vassalage was perhaps fictive in given cases, but it was fertile fiction

acting with the force of a value norm" (Szűcs 2006: 41). This led to another important characteristic of the West, to "the human dignity motive preserved even in subordination" (Szűcs 2006: 41).

"Over and beyond this, the territorial consequence of vassalage, the many small provinces governed by their own customs represented at the given level of intercommunication far more suitable grounds for the development of direct lawfulness and the scrupulous assertion of the law as 'custom', than the roughly articulated from top down, sketchily broad and uniform political-governmental framework" (Szűcs 2006: 42).

In the final analysis we have before us the "the original western model" (Szűcs 2006: 35), the social innovation that István Bibó Hungarian political scientist and sociologist characterized as a structure where "customary, personal, mutually guaranteed obligations and rights, and the small spheres of freedom hold one another in balance" (Szűcs 2006: 35), which "prevent the concentration of power, and represent a force of resistance against the 'brutally expedient' methods of unilateral subordination" (Szűcs 2006: 35).

## The question of the origin of the West as innovation

It is obvious from the above that for Szűcs – as well as many others of – the West was a truly important innovation. Therefore, in accordance with the aim of the essay, the circumstances of its origin would claim our interest in any event.

Let us start out with Hajnal's standpoint according to which, "intensive work confined to *narrow, local development* lay the foundations" for European expansion during the age of geographical discoveries, proving the strength of medieval European social development (Hajnal 1988: 7). In the sequence of examples illustrating as well as interpreting this academic observation he notes that "antique and Arabian shipping conquers distance by the multiplication of towing power, that is, by slavery; the mode of European society building does not allow this, so the mariner is left to his own devices" (Hajnal 1988: 7). Thus, *narrow, local development* acts as a constraint for Hajnal, and László Lakatos also emphasized this feature when he said that after the fall of the Roman Empire "life was confined within narrow, local frames, and became paltry, even wretched; small local communities struggled for mere survival" (Lakatos 1998: 59).

Accordingly, the circumstances of the origin of the West may be described as the constraint of circumstances. But these constraints acted under very peculiar circumstances inasmuch as "the dynamics and integration of the West after the turn of the millennium was conditional on the disintegration process of the preceding period, which is the condition for the separation of state and society" (Szűcs 2006: 38). That is to say, the West "started out by *breaking up and crumbling* beyond recognition both state formations in just three centuries (6th–8<sup>th</sup> centuries). ...

The public authority of the Germans... *dissolved* as did the institutional system of the Empire. ... But not only did the 'state' sphere break up, both original 'social' frameworks also became radically *divided*. Germanic population formations *disintegrated*, as did the society, based on law, of the remainder of the Roman populace. With the *expiry* of public authority, political sovereignty itself became illusory, and with the *decomposition* of society, every conventional cohesive force *expired*. For the time being, the private ownership of land as the main crystallizing force also assisted primarily the political-social disintegration" (Szűcs 2006: 39). "Actually, a kind of integration, a fusion of antique and barbarian heritage also took place in the West ... but the fusion of the elements was so extensive as to nearly extinguish one another. ... Today, an increasing number of details clearly indicate that this *breaking down* of everything is to a considerable degree the condition of the special dynamics which reversed the signs of contrast in relation to Islam and the Byzantine Empire, the two other successors to Rome" (Szűcs 2006: 37).

*Broke up, broke down, crumbled, dissolved, divided, expired* – even this offhand selection well illustrates how Szűcs' rich rhetoric served to demonstrate what kind of *constraints caused by general state, political, and social disintegration* characterized the circumstances of the origin of the West. The development of vassalage gives a condensed picture, since "instead of some glorious motif, its development was precipitated by wretched *exigency*, where, given the *dissolution* of public authority, protection could only be found in some dependent position under private law, on the one hand, and further power prestige could be acquired, on the other" (Szűcs 2006: 40). In conclusion, according to Szűcs, "the first half-millennium of the West represents a wholly unusual 'take off' in the line of the birth of civilizations: disintegration instead of integration, and what is more, amid definite 'civilizational' decline, re-agrarianization, and prolonged political anarchy" (Szűcs 2006: 37).

As concerns the depth, scale, and duration of disintegration, the circumstances of the birth of the West as innovation must certainly have been unusual. However, the model according to which, the birth of a momentous social innovation is conditional on the breaking up, expiry or dividing of previous organizational forms of society, seems to be a general one. At this point it will perhaps suffice to point out that in Asia as well as in America "great cultures emerged in the narrow sections of the continent ... in the narrow areas of the congregation of peoples. Congregations like these *break up* the social organizations founded on natural, day-to-day living, *put an end* to the simple hunter-gatherer way of life, confine small units to an area and *compel* them to set up artificial ... permanent organizations" (Hajnal 1988: 21) [*italics added, N. F.*]. In the foregoing Hajnal speaks of concrete cases, yet, discernibly, he sets forth a universal historical developmental model in showing that social organizations confined to an area and land use also came into being amid *constraints caused by disintegration*. Accordingly, the fact that the

circumstances of the emergence of important historical-social innovations may be described basically by some kind of a decay, breaking up, that is, disintegration applies not only to the birth of the West, but may be considered as a general rule.

However, I would go further. In the following I shall use examples from diverse fields to corroborate that the validity of the model points well beyond the perspective of history, and that fundamental change affecting the whole of the operation of a more or less self-regulating system may come about only through the dissolution, breaking up, or, one might say, crisis of the system.

## Complex unstable conditions

Let us first take a look at a formation of society which is deeply opposed to change due specifically to its most characteristic features. It is well-known that the most conspicuous features of the model of the bureaucratic organizational system described by Michel Crozier is the system of impersonal regulations covering nearly everything, which "define in detail the various positions and prescribe the mode of behavior that those filling the position must in most cases observe" (Crozier 1981: 286). At the same time, it is clear that in order to "keep relations impersonal, the decisions not falling under the impersonal regulations must unavoidably be made on a level, where those responsible for the decisions are protected from being pressured by those, whom the decisions concern... If the pressure exerted in the direction of impersonality is strong, then the tendency of centralization is unstoppable" (*ibid.*: 289).

Since in the bureaucratic organizational system "the impersonality of decisions and the centralization of decisions" become "an integral part of the internal balance of the organization" (Crozier 1981: 285), the "only instrument of action ... is to frame new regulations, thereby increasing centralization even more" (Crozier 1981: 294). However, the "organizational system, which is characterized mainly by rigidity, naturally, cannot easily adjust to change and tries to resist every transformation" (Crozier 1981: 296). For this reason, a "bureaucratic organization is an organization that cannot correct its behavior by learning from its errors, and in which dysfunctions constitute the decisive elements of balance" (Crozier 1981: 296).

Actually, "no organization can avoid the necessary transformation; it is constantly compelled to adjust to changes in its environment, as well as to the less conspicuous, though just as profound, changes in its members" (Crozier 1981: 283). Consequently, "resistance to change in a bureaucratic organization is only one aspect of the problem" (Crozier 1981: 297). Sooner or later even the most rigid organization is compelled to change, giving rise to another, perhaps a still more important, aspect, "the special mode of adjustment of a bureaucratic organization" (Crozier 1981: 297).

The logic of Crozier's model implies that "in a bureaucratic organization change must follow a top-down course, must be universal, extending to the whole organization. This change cannot be gradual, occurring by stages and in parts. It is a prerequisite of change that the dysfunction of the organization be severe enough to threaten its survival" (Crozier 1981: 298), in short, that the system undergo *crisis*. Crisis is the only "way whereby the necessary corrections can be carried out, therefore, it also plays an important role in the development of the system – which is possible only through crisis" (Crozier 1981: 298).

In a bureaucratic organization change can take place only through crisis permeating the whole organization. Is this example too extreme? Have we chosen an excessively rigid case of social organization, so the mode of its change cannot be typical? I rather think that in our case this really extreme example serves to emphasize rather than distort. It makes obvious what is also valid but less readily perceived in other areas, for instance, in science.

The complex relationship between change and rigidity, novelty and orthodoxy in science is well-known. I have nothing new to add, but I would like to refer to this fact, too, with the help of some select tendentious quotations. For instance, Isaiah Berlin described cultural history as a "changing system of ideals that first liberate, but later inevitably become a straitjacket" (Berlin 1999: 159). Arthur Koestler gives a similar view of the uncertainties of science "between relatively brief eruptions which lead to the conquest of new frontiers, and long periods of consolidation. ...The new territory opened up by the impetuous advance of a few geniuses acting as a spearhead ... and soon the revolution turn into a new orthodoxy, with its unavoidable symptoms of one-sidedness overspecialization, loss of contact with other provinces on knowledge, and ultimately, estrangement from reality" (Koestler 1964: 225).

Kuhn's theory is more systematic and comprehensive than Koestler's, yet he says something similar. According to his standpoint, if Koestler's new orthodoxy, or scientific paradigm in Kuhn's terminology, comes into being, then the "normal science does not aim at novelties of fact or theory" (Kuhn 1970: 52). So, the question concerns how such changes may nevertheless come about and how can new "theories ... arise from normal science" (Kuhn 1970: 66). According to Kuhn, "one can appropriately describe the fields affected – by the change NF – as in state of growing crisis" (Kuhn 1970: 67). Just as the "Aristotelian fortress collapsed" or, as John Donne lamented, "tis all pieces, all coherence gone" (Koestler 1964: 229) in the early 17<sup>th</sup> century, thus numerous other examples verify that a "novel theory emerged only after a pronounced failure in the normal problem-solving activity" (Kuhn 1970: 74). The emergence of new theories is preceded by strong professional uncertainty, an overgrowth of the variants of a previously dominant theory and according to Kuhn "that proliferation of versions of a theory is a very usual symptom of crisis" (Kuhn 1970: 70), and "by proliferating versions of the



paradigm, crisis loosens the rules of normal puzzle-solving in ways that ultimately permit a new paradigm to emerge" (Kuhn 1970: 80).

Since "there are certain analogies between the characteristic stages in the history of an individual discovery, and the historical development of a branch of science as a whole" (Koestler 1964: 224), the above findings may be carried over to the field of individual discovery with relative ease. Clearly, "when life presents us with a problem it will be attacked in accordance with the code of rules which enables us to deal with similar problems in the past" (Koestler 1964: 118). "But novelty can be carried out to a point – by life or in the laboratory – where the situation still resembles *in some respects* other situations encountered in the past, yet contains new features or complexities which make it impossible to solve the problem by the same rules of the game which were applied to those past situations" (Koestler 1964: 119). This is why Feyerabend could say that "the success of 'science' cannot be used as an argument for treating as yet unsolved problems in standardized way" (Feyerabend 1993: 2).

The unsolvability of some problem produces a block in the individual's thinking. It is quite possible that in the resulting stress situation "when all hopeful attempts at solving the problem by traditional methods have been exhausted ... organized purposeful behavior itself seems to go to pieces" (Koestler 1964: 119). Koestler has "coined the term 'bisociation' in order to make distinction between the routine skills of thinking on a single 'plane' as it were, and the creative act, which ... always operates on more than one plane. The former may be called single-minded, the latter a double-minded, transitory state of unstable equilibrium where the balance of both emotion and thought is disturbed" (Koestler 1964: 35).

Although Koestler enriched our vocabulary with a new term when describing the creative activity of the mind<sup>12</sup> as an unstable state, he actually depicted the same situation as the above citations. Thus the road to the firm 'a-ha' experience manifested in the moment of insight of scientists also follows the hitherto distinguished model of the birth of innovations.

The above examples from diverse fields of learning use different words – *disintegration* or *anarchy* in the case of the birth of the West, *crisis* in the course of the change of the bureaucratic organization, *anomaly*, *crisis*, *revolution* in the case of new scientific theories, and *bisociation* in that of individual discovery – but they mean the same. Accordingly, an innovation may come into being only amid some

12 Bisociation describes the state of the creative mind; therefore, at this point the question arises whether special personality traits distinguish people successful in problem-solving from others. According to Csikszentmihalyi, most personality traits attributed to creative people show strong dependence on culture. Nevertheless, if he had to say in one word what makes creative people different from others, Csikszentmihalyi would use the attributive *complexity*, because creative people often "fuse thoughts and behavior in themselves that are separate in others" (Csikszentmihalyi 2008: 66). It is interesting that in connection with creative activity Koestler too speaks of a transitory, that is complex unstable state, moreover, Feyerabend also uses similar term in framing the requirements of scientific method, underlining that "a complex medium, which is surprising and holds in store unforeseeable developments, requires complex methods" (Feyerabend 2002: 39). I do not think this identical usage is accidental. It may be that in a seemingly insoluble blocked situation a complex personality, imbued with extremes and subjected to the pressure of this internal tension, would more readily surrender the rules preset on the basis of previous experience, are customary but useless in the given situation, and in this respect may become a part of the complex methods leading to the solution. All this raises the possibility of connecting our study to the nowadays revived research program on complex systems, however, we are yet unprepared.

kind of a disintegration, anarchy, crisis, or revolution<sup>13</sup>. The question remains: How does this dissolution of many names come about?

## From blind-alley development to creative anarchy

The answer is simple if – in the spirit of the principle, when “the senators sit there without legislating... Once the barbarians are here, they’ll do the legislating” (Cavafis 1992) – we attribute dissolution to the impact of an external force. As an endogenous process, however, the answer depends basically on its antecedent, that is, what we have gotten into immediately before the onset of disintegration. Mostly a blind alley, it seems. This holds for normal science which “repeatedly goes astray” (Kuhn 1970: 6), and explains why “the great breakthroughs in science, the arts, or philosophy are all successful escapes from some impasse” (Koestler 1967: 231). It also holds for biology, where – although the complex adaptation forms of living organisms fill us all with wonder – “the theory of evolution by natural selection does not predict that organism will get more complex. ... many and perhaps most lineages change little for many millions of years” (Szathmáry–Maynard Smith 1999: 15), and in this respect it has come to its end. We find a similar blind-alley quality in the field of social development, concerning which – beside Bibó’s often cited metaphor – we again refer to Hajnal. He said that the “possibilities and methods of human communication trigger the creative forces, and the organizations, institutions thus established shape the community and the individual; as soon as the social organization becomes mechanized ... as soon as it surmounts the society instead of actively mediating, encouraging man’s material-mental activity, its prolific capacity for development ceases: it can no longer induce people to produce something new. It may become an increasingly sophisticated culture but without the renewed advance, elaboration of ancient deep forces” (Hajnal 1988: 18). In such a society “people live in what is already given” (Hajnal 1988: 18).

In other words, sooner or later development inevitably maneuvers itself into some blind alley,<sup>14</sup> and the only way out is to go backward. It already means a regression when during the crises “scientists usually develop many speculative and unarticulated theories” (Kuhn 1970: 61), and it is well-known that “Copernicus did not continue where Ptolemy left off; he went back two thousand years, to Aristarchus” of Samos (Koestler 1967: 233). Kuhn called scientific revolutions “those non-cumulative developmental episodes in which an older paradigm is replaced in whole or in part by an incompatible new one” (Kuhn 1970: 128). Clearly, giving up the idea of the cumulative development of science and the

<sup>13</sup> „As a rule, then, the beginning is made by some great trauma... This followed by a traumatic neurosis, which really constitutes the incubator of the new being. By it everything is thrown about and broken down into a labile, anarchic, chaotic condition” (Friedell 1953: 54).

<sup>14</sup> In the longer run every course is a blind alley. All this raises the possibility that the real problem in Hungarian history is not its blind-alley character, but that in the sequence of fresh starts the road ends too soon.

incomparability of the rival paradigms endow scientific revolutions with the same legislative beginnings accompanied by the violation of the law as hold for political revolutions today. As a matter of fact, originally, the word revolution was an astronomic term, which referred to the regular rotating movement of celestial bodies. It is particularly noteworthy that in the seventeenth century, when we first encounter the word as a political term, it was still used to describe the *return* to some predetermined point (Arendt 1963).

Stops, blind alleys in social development also “mean disintegration, the increasing shallowness of the driving forces, the absence of the need for productive effort” (Hajnal 1988: 18). Only “societies building on new, perhaps *more primitive*, but different, more viable fundamentals” (Hajnal 1988: 18) can offer a way out, societies, which are able to “reach down [*step back?* my interjection, *N. F.*] into the deep, irrational world of man’s and the society’s life, seizing therefrom the new life material, as it were, for elaboration” (Hajnal 1988: 25) [*italics added, N. F.*].

It is commonplace that “creativity in the sciences and the arts often depends on seeing analogies” (Szathmáry–Maynard Smith 1999: 145), since “the creative act ... does not create something out of nothing, it ... combines, synthesizes” (Koestler 1964: 140). Moreover, “it was the evolution of language that broke down the barriers ... and made possible the continuous cultural evolution that followed” (Szathmáry–Maynard Smith 1999: 145). At the same time, it is well-known that in the course of individual discoveries we have to abandon (*stepping back?*) just this language in order to make our new ideas clear (Koestler 1967: 235).

In other words, we may end up in a blind alley by way of the one-way process of – scientific, biological, social – overspecialization. Backing out of it is accompanied by the undoing of specialization, that is, in this sense, by disintegration. Backing out may be not only concrete but also figurative. Koestler cites a notion from 1928 “according to which ... vertebrates... descend from the *larvae* of a primitive echinodermatous form reminiscent of sea urchins or sea cucumbers” (Koestler 1967: 213) [*italics added, N. F.*]. This may be a fine example for the developmental model “of getting out of the blind alley by backing out firmly and jumping high,” which, according to Koestler, “can be identified at every important and definitive evolutionary juncture” (Koestler 1967: 216).

The development of the eukaryotic cells offers an especially exciting example, which is considered a major evolutionary leap even by those who do not necessarily share Szathmáry’s and Maynard Smith’s concept of big leaps in evolution. These cells are real factories with specialized organelles. Their nuclei are separated from the surrounding cytoplasm, they have a number of cell organelles – such as the mitochondrion considered the cell’s energy factory – and pigmented plastids. They are much more complex than the prokaryotic cells (essentially bacteria) and, on the average, about ten times bigger. But the most important difference between prokaryotic and eukaryotic cells is that the former are surrounded by a rigid cell

wall. We do not know how it happened, but we may assume that losing the cell wall triggered the series of changes leading to the development of the eukaryotic cells. Undoubtedly, this in itself created a very *unstable* situation for the cells. In the absence of walls the cells became highly vulnerable, which led to the extinction of the lines of descent in question, one after the other. Therefore, the loss of the rigid cell wall is, in itself, a risky backward step, and, presumably, occurred repeatedly. Then, at some stage, this more unstable situation made possible the development of a new nutritional method, and “*forced* the ancestral eukaryotes to evolve a new way of segregating of their chromosomes” [italics added, *N. F.*]. The resulting “mitosis can be seen as *something forces* on the eukaryotes because the old, prokaryotic mechanism was no longer effective” (Szathmáry–Maynard Smith 1999: 68) [italics added, *N. F.*]. On the other hand, it is a fact that this new method was the prerequisite to the appearance of multicellular organisms, so we need not dwell on the evolutionary significance of this step.

It seems that there really exists a general model in the circumstances of the origin of significant innovations impacting on the whole of some level of a hierarchic system. The innovation may emerge amid the *dissolution, disintegration, anarchy, crisis, anomaly, revolution, or bisociation* resulting from the backing out of some blind alley. Henceforth *instability* will be used as a collective for these conditions. This will not only simplify our phraseology, but also remove the direct or indirect political connotations of the terms conventionally used in social scientific sources, making it clear that we treat these terms as dynamic categories for describing the circumstances of the origin of innovations and not as political categories.

At the same time, I prefer to use the attribute *unstable* as a metaphor, to exploiting its exact dynamic meaning. This is not to say that there is no connection at all with this specialized scientific meaning. The connection with the modern chaos theory is particularly exciting, since in the course of the so-called chaotic behavior the system approaches always new unstable conditions, “chaos is, therefore, permanent instability” (Tél–Gruiz 2002: 80). It is also noteworthy that complex systems may be permanently in a state of nonequilibrium (instability), which may serve as a fundamental source of endogenous creativity in the system. Furthermore, the fact that unstable chaotic systems are extremely sensitive to external influences is decisive in making virtually impossible the prognostication of the long-term consequences of highly significant innovations. All this leads one to look on all kinds of social engineering with very strong reservations. However, thinking through how the above natural scientific findings apply to social systems requires further research.

We may establish that big, that is, dramatic changes affecting the whole of one system (subsystem) are possible if the said system (subsystem) becomes unstable; with this we come to our proposed resolution of the essay’s central problem. Since we have answered the initial question, our study ends at this point.

Nevertheless, I feel the need to continue, because I would like to make it absolutely clear that in the foregoing we have explored only the *necessary* conditions of the origin of innovations affecting a whole system. The kind of constructivist activism characterized by “the worse, the better,” or “let’s foment a revolution, provoke turmoil, bring about a state of disorder” in the interest of some hoped for utopia does not in any way follow from our findings. The fact is that instability by no means guarantees the emergence of innovations, and I am inclined to believe it is impossible to explore the sufficient conditions even theoretically. In any case, at this point I have certainly come to the limits of my knowledge. Of course, not knowing something does not mean we cannot give account of the nature of this not knowing, which is what I shall try to do in the remaining part of the essay.

In both Hajnal’s and Bibó’s perception of the peculiarities of European social development special emphasis is attached to their use of the phrase *a state of ‘creative anarchy’* (Lakatos 1998: 59) to characterize the circumstances after the fall of the Roman Empire, which they – and, as we have seen, Szűcs too – regarded as extraordinary. Koestler used the very same phrase when he said that any new scientific synthesis surfaces from “creative anarchy” (Koestler 1964: 230) which recurs from time to time in the history of every discipline and which corresponds to the “incubation” period seen in the case of individual discoveries. When, in connection with scientific development, Kuhn discusses the “destructive-constructive paradigm changes” (Kuhn 1984: 97), then, though in different words, he not only qualifies very similarly the circumstances of the birth of something new, but also calls to mind the well-known concept of *creative destruction* (Schumpeter 1976). We know Schumpeter used the phrase in order to grasp the endogenous process of economic development, wherein the enterprises, regarded as the incessant source of innovations, create something new by destroying old structures – in the field of consumer goods, production and transport methods, markets, or organizations.

Moreover, this may explain how during the past couple of hundred years western capitalism was able to introduce important technological, technical, organizational, and social innovations without cataclysms. In my opinion, the concept of *creative destruction* is but the economic equivalent of the microanalytical perspective cited from Mérei in the early part of the study. This is a likely interpretation in light of the fact that Schumpeter regarded as entrepreneurs not only the independent businesspersons of the market economy, but every economic subject assigned the function of achieving new combinations (Schumpeter 1968: 120). Furthermore, competition among entrepreneurs acts as the economic counterpart of *tension* in the inseparably connected modeling and alteration process, which makes it a permanent source of economic innovations. Accordingly, the peculiarity of the western-type of market capitalism lies in the frequency with which innovations

emerging in a micro-level state of instability become macro-level results, rather than in their birth under not unstable circumstances.<sup>15</sup>

Naturally, none of the above mentioned authors averred that anarchy itself is fertile and destruction is inevitably creative, since they all qualified the given unstable situation as creative in retrospection, aware of the outcome. (Lakatos 1998: 59) If it is not always so, we justifiably ask: how can anarchy, destruction, instability actually be creative in a given case?

## Is evolution the token of success?

First of all, let us make it clear that, contrary to our usage so far, innovations are by no means one-time acts, as priority disputes often accompanying scientific discoveries indicate, among others. These disputes conceal not only the existing personal aspirations and career ambitions of scientists, but also the real difficulties in deciding priority. This is very vividly described in Kuhn's analysis of the priority dispute surrounding the discovery of oxygen. As he noted, "Priestley's claim to the discovery of oxygen is based upon his priority in isolating – in 1774 – a gas that was later recognized as distinct species" (Kuhn 1970: 54). On the other hand, Antoine Lavoisier was the first to reach the conclusion in 1777 that "the gas was a distinct species, one of the two main constituents of the atmosphere, a conclusion that Priestley was never able to accept" (Kuhn 1970: 54). Evidently, "though undoubtedly correct, the sentence '*oxygen was discovered*' misleads by suggesting that discovering something is a single simple act" (Kuhn 1970: 55). "Discovering a new sort of phenomenon is necessarily a complex event, one which involves the recognizing both that something is and what it is" (Kuhn 1970: 55). Undoubtedly, Priestley chanced upon oxygen, but did not really understand its nature, therefore, "we can safely say that oxygen had not been discovered before 1774, and we would also say that it had been discovered by 1777 or shortly thereafter" (Kuhn 1970: 55).

The process of individual discoveries is just as complex with multiple steps as that of the discoveries of the various disciplines. As for how many iterative steps it goes through, how many loops it contains, how many insights it requires depends on the depth and range of the subject. Incubation lasts sometimes years and sometimes only a couple of hours (Csikszentmihalyi 2008: 88). Or much less, we might add, considering that some of the great orators did not know what they were going to say when they opened their mouths to speak (Feyerabend 1993: 51).

So far we have referred to the development of eukaryotic cells as taking place in one big step, while clearly this evolutionary step comprised a whole se-

<sup>15</sup> I do not have the answer as to the how of it, but I believe its explanation would require a much more thorough study of the literature on economics, we have deliberately avoided here. It is also clear that the subject matter dealt with here is one of the reasons why the sphere of validity of the model I have discussed remains uncertain. In the course of writing the essay I have repeatedly stressed my special interest in significant innovations, but what qualifies as significant in a hierarchical system strongly depends on one's point of view. It may easily be that an innovation proves important on the level of some subsystem, but from the point of view of higher organizational levels it may be unnoticeable even.

ries of events. This included the development of a new nutritional method, the emergence of an internal cellular structure and a new method of locomotion, the development of chromosomes with several replication points, and the appearance of organelles. "The fascinating thing about this story is the way in which many apparently unconnected changes, setting the scene for all subsequent evolution, were in a sense forced on the cell by the loss of the cell wall, an event that might have seemed at the time both trivial and regressive" (Szathmáry–Maynard Smith 1999: 78).

Thus we can see that innovations are not one-time acts but complex events, that is, multiple-step, feedback processes. This in turn, specifically in connection with this process character, makes it possible for us to take the success of some novelty, initiative, or innovation to mean the prevalence over the unknown future. That is to say, if we were to imagine an artificial life form, which has the greatest chances of survival in a complex and unfathomable environment, then it would not be an engineering, constructional problem as commonly understood. (Csikszentmihalyi 2008: 116). Nor can we propose some ready-made solution, instead we may only conceive that course which stands the greatest chance – still far from certainty – of leading to some kind of a, by necessity, temporary solution.

What we can build on is that the conceived life form must be capable of *"tackling most unexpected situations"* and also of *"taking advantage of as many opportunities as possible"*. On the one hand, in accordance with this dual expectation, it is obviously advantageous for the entities of the life form to strive for some measure of stability, and, with a certain degree of conservative attitude based on the principle of *"if it isn't broken, don't glue it together"* (Mokyr 2004: 211) and relying on past experiences, to try to find the best possible solution to everything, because it is *"routine that keeps us prepared for the next challenge"*.<sup>16</sup> On the other hand, it would be expedient for "one or another of the entities to have a regulatory system which gives a positive signal every time they discover something". It is particularly important that the discovery be valuable in itself and the "organism receive reinforcement not only for useful discoveries, otherwise it would be overpowered in the struggle with the future" (Csikszentmihalyi 2008: 116). In order to find a solution which is in some way successful, we must lay down certain rules and leave the rest to time, more exactly, to the Blind Watchmaker, the Darwinian evolution. However, this does not mean that in the foregoing we have offered logical arguments in favor of the application of the Darwinian evolutionary view. We have merely unfolded what our chosen definition of success implicitly contained already. If survival is success, then Darwinian evolution is the "answer".

It is noteworthy that in European social theory there is a trend of thought which preceded Darwin, indeed exerted an influence on him through Malthus, and which works with a development concept much like Darwin's. This "shows

<sup>16</sup> The captain of the world's biggest container carrier said this on the National Geographic Channel.

how complex, orderly and, in a precisely defined sense, suitable institutions may develop through interpersonal relations, which owe little to planning, which we do not make up, but derive from the autonomous actions of many people, who were acting unawares" (Hayek 1960: 58–59). Since the development of this social order would be "the result of adaptive evolution" (Orthmayr 2002: 90), we must consider whether the elementary conditions of the unfolding of evolution based on natural selection could be realized on the social level.

Speaking about conditions, the first thing usually emphasized is that the process of evolution based on natural selection requires the periodic appearance of mutations, that is, new inheritable variations, in the course of procreation. Therefore, heredity cannot be perfect, because its occasional errors (mutations) create the diversity wherein selection may take effect. Let us note, however, that initially, at the beginning of life on earth "the problem would have been too much mutation, and not too little" (Szathmáry–Maynard Smith 1999: 34). If there are too many mutations, then selection cannot sustain the original message for the necessary length of time. Clearly, in the course of Darwinian evolution prevailing in the organization of society the problem arises more from the inaccuracy than from the high accuracy of replication.

At this point the case of the emergence of the West as innovation becomes particularly interesting. Let me say in advance that I do not propose to give an exhaustive description of the development of the western-type of societies in the following discussion. Let us consider it as a highly stylized case study, in which, still within the conceptual framework of the school of the dawn of social history, I shall analyze the applicability of Darwinian evolution to this concrete case.

In Lakatos' opinion, expressed in his study reconstructing Hajnal's developmental theory, "the most important thing in the whole of European development – what really makes it 'European' and 'humane' – is that due to the peculiar – otherwise intimidating, but from the point of view of subsequent development actually advantageous – circumstances at the beginning of the development, over a long historical period the social actors, even the smallest, had the opportunity to shape the life forms and techniques as best suited them, and not the powers that be, not money, not rational thinking could (greatly) interfere" (Lakatos 1998: 60).

I tend to interpret all this as the change in the modes of storage of information found in human communities, which recalls the circumstances of Szathmáry's and Maynard Smith's evolutionary big leaps. The conditions of the Darwinian evolutionary process, however, are created by the self-governing character of the world of customs. Lakatos summarizes it as follows: "(1) Initially strong intellectualism did not yet exist, therefore, life forms were expressed in customs; (2) the social actors themselves expressed their living conditions in customs; (3) customs protect against external forces...; (4) ... this complexity of the world of



customs offers protection against the commanding reason" (Lakatos 1998: 60).

Consequently, through the transmission of life forms developed by exploratory human endeavors, customariness ensured the necessary accuracy of replication, and made it possible for evolution based on natural selection to function on the level of the smallest social actors as well. I think Lakatos formulated essentially the same idea when, summarizing Hajnal's thoughts on ideal social development, he underlined the role of time next to the constraint of individual achievement and the unconditional recognition of spontaneous life forms. Thus, the secret of the success of medieval European development lies in the unfolding of evolution at this lowest autonomous social level.

The above paragraph introduced the *ideal* social development beside the *successful* one from the point of view of survival, whereby we have also entered the realm of value judgments. Obviously, in judging the success of innovations we may also consider, beside survival, a great variety of moral, religious, political, economic or short-term efficiency criteria. Their application will not make evolution teleological, but by way of certain – mostly "we don't do such things" type of prohibitive – (moral, religious, political, economic etc.) rules we may nevertheless set the general course for social development.

It is in terms of this value content that Lakatos queries: "is Hajnal right in suggesting that the best results are accomplished when things are entrusted to the people? Is it true that left to themselves people will create a world worthy of man?" (Lakatos 1998: 63). Well, "evolution by natural selection lacks foresight" (Szathmáry–Maynard Smith 1999: 25), consequently, it fails to guarantee anything. At most we can say is that "amid the disintegration ... civilizational decline, re-agrarianization, and protracted political anarchy" (Szűcs 2006: 37) characterizing the first half-millennium of the West, certain previous constraints ceased. When, "after the near complete breakdown of the old forms, this peculiar, mosaic-like, ceremonious, mannered feudal world ... after its own fashion, set out to establish some new kind of a relationship between society and state..., the administrative, military, fiscal, jurisdictional functions of the state *sinking* stepwise, as it were, *dispersed* in the "feudal society", "sovereignty was absorbed *piecemeal*, as it were, by a newly formed political sector of society" (Szűcs 2006: 42) (italics added, N. F.).

The secret of European development lay not simply in the evolutionary process, but also in the noteworthy fact that basically the struggle for survival was no longer between empires, states, city states or dynasties. Selection shifted to the level of "many small provinces each governed by its own customary law" (Szűcs 2006: 42), more specifically, to the level of (micro-)communities functioning under the constraint of individual performance. For this reason, if this evolution was able to "invent" anything – for which there were, of course, no prior guarantees at all – it could only be autonomous society itself.

Thus the "exploratory endeavors" (Orthmayr 2002: 90) of human micro-

communities gave rise to “something greater than the individual human spirit” (ibid). Consequently, “every idea, plan rests on enormous, tested empirical material. The organization of European society ... transmitted knowledge, from manual labor to abstract thoughts, on a never before seen scale. All work and enterprises are deeply embedded in the labor structure of society, they become real, possessing a sense of assurance based on the experience of generations, they do not rely merely on theoretical learning and individual abilities, daring” (Hajnal 1988: 7).

It seems that in the course of the emergence of the western-type of organization of society the “means whereby information is stored and transmitted (Szathmáry–Maynard Smith 2000: 15) in human communities have also changed. This is why I consider this innovation within the framework of social evolution to be a major evolutionary step as Szathmáry and Maynard Smith understood it.

## Concluding remarks

Our initial motivation was to explore the circumstances of the origin of innovations leading to big leaps. Then we narrowed this, saying, we were interested in the possibility of dramatic changes affecting the whole of some organizational level of a hierarchic system. On the basis of the direction our modified phrasing has taken and the terminology – tension, instability, complex hierarchic systems, chaos – gradually introduced along the way, our analysis has taken on the aspect of a systems-theory approach. In view of this, the absence of systems-theory references may be surprising. The reason is precisely that I consider this connection to be of fundamental importance. Therefore, a systems-theory analysis of the questions discussed here merit a separate study. By citing sources far removed from systems-theory approaches, the aim of the present essay was merely to demonstrate the presence of this view in such works as well. I plan to write a sequel based on the theory of dynamic systems, in which I would like to pair Niklas Luhmann’s biology-inspired systems theory with the chemoton theory devised by Tibor Gánti, the leading Hungarian representative of systems chemistry.

## References

- Arendt, Hannah (1963): *On Revolution*. New York: Viking.
- Berlin, Isaiah (1999): *Concepts and Categories: Philosophical Essays*. Hogarth Press, 1978. Pimlico.
- Braudel, Fernand (1996): *The Mediterranean and the Mediterranean World in the Age of Philip II*. University of California Press.
- Cavafy, Constantine (1992): *Waiting for the Barbarians*. Edmund Keeley (trans.) In C. P. Cavafy. *Collected Poems*. Edmund Keeley and Philip Sherrard (trans.),

- George Savidis (ed.). Revised edition. Princeton, NJ: Princeton University Press.
- Coleman, S. James – Herbert Menzel – Elihu Katz (1966): *Medical Innovation: A Diffusion Study*. Indianapolis: Bobbs-Merill.
- Crozier, Michel (1963, 1964): *The Bureaucratic Phenomenon*. Chicago: University Chicago Press.
- Csikszentmihalyi, Mihály (2008): *Kreativitás*. Budapest: Akadémiai Kiadó.
- Feyerabend, Paul (1993): *Against Method*. London, Verso.
- Friedell, Egon (1953): *A Cultural History of Modern Age*. London: Vision Press.
- Fokasz, Nikosz (2006): *Növekedési függvények, társadalmi diffúzió, társadalmi változás. Szociológiai Szemle, 3*.
- Furet, Francois (1999): *The Passing of an Illusion: The Idea of Communism in the Twentieth Century*. Deborah Furet (trans.). Chicago: University of Chicago Press.
- Geroski, Paul A. (2000): Models of Technology Diffusion. *Research Policy*, 29: 603–625.
- Griliches, Zvi (1988 [1957]): Hybrid Corn: An Exploration of the Economics of Technological Change. *Technology, Education and Productivity: Early Papers with Notes to Subsequent Literature*, 27–52. New York: Basil Blackwell.
- Hajnal, István (1988): *Az újkor története*. Budapest, 1936. Reprint. Akadémiai Kiadó.
- Hayek, Friedrich A. (1960): *The Constitution of Liberty*. London and Henley: Routledge and Kegan Paul.
- Hegel, Georg Wilhelm Friedrich (1979): *The Science of Logic*. Budapest: Akadémia Kiadó.
- Jánosy, Ferenc (1979): *Az akkumulációs lavina megindulása*. Budapest: Magvető Kiadó.
- Konstandopoulos, Athanasios G. – Theodore Modis (2003): Urban Guerilla Activities in Greece. *Technological Forecasting and social Change*, 10 May.
- Koestler, Arthur (1964): *The Act of Creation*. London, Hutchinson & Co. Publishers
- Koestler, Arthur (1967): *The Ghost in the Machine*. Hutchinson & Co.
- Kornai, János (2010): Innováció és dinamizmus. *Közgazdasági Szemle*, (57): 1–36.
- Kuhn, Thomas (1970): *The Structure of Scientific Revolutions*. The University of Chicago Press.
- Lakatos, László (1998): Az eredeti felhalmozás. *Szociológiai Figyelő*, 1–2.
- Marchetti, Cesare (1986): Fifty-year Pulsation in Human Affairs. Analysis of Some Physical Indicators. *Futures*, 17(3): 376–388.
- Marchetti, Cesare (1997): Millenarian Cycles in the Dynamics of the Catholic Church. *Technological Forecasting and Social Change*, 55: 281–299.
- Marchetti, Cesare – Perrin S. Meyer – Jesse H. Ausubel (1996): Human Population Dynamics Revisited with the Logistic Model: How Much can be Modeled and Predicted? *Technological Forecasting and Social Change*, 52: 1–30.
- Maynard Smith, John and Eörs Szathmáry (1997): *The Major Transitions in*

- Evolution*. Oxford University Press.
- Mérei, Ferenc (1996): *Közösségek rejtett hálózata*. Budapest: Osiris.
- Modelski, George – Perry Gardner (2002): “Democratization in Long Perspective”, revisited. *Technological Forecasting and Social Change*, (69): 359–376.
- Mokyr, Joel (1990): *The Lever of Riches: Technological Creativity and Economic Progress*, Oxford University Press, New York-London
- Nora, Pierre (1974): Le retour de l'événement Faire de l'histoire. In Le Goff, Jacques–Nora, Pierre: *Faire de l'histoire*. Vol. 1. Paris: Gallimard.
- Orthmayr, Imre (2002): Polányi Mihály és a társadalmi evolúció elmélete (Mihály Polányi and the theory of social evolution). *Polanyianna* 1–2.
- Popper, Karl (1997): *Megismerés, történelem, politika*. Budapest: AduPrint.
- Rogers, Everett M. (1962): *Diffusion of Innovations*. New York: Free Press.
- Schumpeter, Joseph (1947): *Capitalism, Socialism Democracy*. New York: Harper.
- Schumpeter, Joseph (1968): *The Theory of Economic Development*. Harvard University.
- Simon, Herbert (1982): The Structure of Complex Systems. In *Models of Bounded Rationality*. Vol. 3. MIT Press.
- Szathmáry, Eörs – John Maynard Smith (1999): *The Origins of Life*. Oxford University Press.
- Szűcs, Jenő (2006): Vázlat Európa három történelmi régiójáról. In Mátyás Domokos (ed.): *A Magyar esszé antológiája* (Anthology of Hungarian Essays). Budapest: Osiris.
- Tél, Tamás – Márton Gruiz (2002): *Kaotikus dinamika*. Budapest: Nemzeti Tankönyvkiadó.

# Commuting Time, Wages and Reimbursement of Travel Costs. Evidence from Hungary<sup>1</sup>

Tamás BARTUS

tamas.bartus@uni-corvinus.hu

**ABSTRACT:** The paper explores the hypothesis that high costs of commuting are responsible for the persistent unemployment of Hungarian villages. An attempt is made to estimate the compensating wage differential associated with commuting time using individual-level data, taken from a survey conducted among workers who have left the unemployment register and got a job in March 2001. The empirical analyses are motivated by a simple wage posting model, which predicts a positive effect of commuting time on wages and explicit reimbursement of travel expenses, which is conditional on the unemployment rate at place of work. We find that the unemployment rate in settlements where jobs are located lowers the positive effect of commuting time on wages, but it increases the probability of receiving some reimbursement of travel expenses, conditional on high unemployment at place of work. The findings suggest that wages paid by employers located in high unemployment areas do not compensate for costly commuting. Our study therefore supports the hypothesis that persistent unemployment is maintained by high costs of commuting, relative to wage advantages.

**Keywords:** Commuting, Spatial Mismatch Hypothesis, Compensating Wages

## 1 Introduction

One of the unique features of the Hungarian labor market is the persistence of regional differences in economic prosperity and unemployment. Unemployment rates are substantially higher in the villages situated in the North-Eastern and the Southern part of Hungary, and the rate of unemployment has been persistently high since the early 1990s. Similar to the logic of the well-known spatial mismatch hypothesis, which claims that the suburbanization of job opportunities accounts for the high unemployment rate among black inner-city residents (Kain 1992; Ihlanfeldt–Sjoquist 1998), it was proposed that the high costs of daily commuting to urban labor markets accounts for the high unemployment rates in villages

<sup>1</sup> The research was financially supported by the project “Efficient Government, Professional Public Administration and Regional Development for a Competitive Society”, which is part of the TÁMOP project 4.2.1/B-09/1/KMR-2010-0005 of the Corvinus University of Budapest. The paper was prepared within the research group “Social and Cultural Resources, Development Policies and Local Development”, led by Zoltán Szántó. English revision by Andrew Ryder.

(Köllő 1997, 2006; Kertesi 2000). Using a transportation database with settlements as units of observation, Köllő (1997, 2006) showed that in the absence of public transportation linkages, commuting with cars would use up a substantial part of the expected wages. Public transportation links are especially underdeveloped in regions where villages with high unemployment rates are typically situated. He also estimated lower-bounds of travel expenses. Kertesi (2000) relied on these estimates when analyzing the 1996 micro-census of the Hungarian Statistical Office and found that the probability of commuting decreased with commuting costs.

In this paper, we make an attempt to estimate the compensating wage differential (Leigh 1986; Zax 1991) associated with costly commuting using individual-level data. The positive correlation between wages and commuting time is often explained with the help of search theory: prospective commuters are more critical towards low wage jobs, because the reservation wage is expected to increase in commuting distance or commuting time (Rouwendal 1999; Manning 2003b). However, standard urban economic theory offers another explanation in terms of residential choices. There is a trade-off between costs of commuting and costs of housing, and high-wage employees might prefer residential locations which are far away from their job (Dargay-van Ommeren 2008), while low-income employees are forced to live close to their jobs. The relationship between commuting time and wages arises as a by-product of choosing the best residential location. The Hungarian labor market provides a unique opportunity to assess the explanatory power of the former approach. Labor mobility in Hungary is substantially constrained by the small housing rental market; the vast majority, more than 90 percent of houses are owner-occupied. Besides, housing transactions involve substantial transaction costs and a bad decision may put more than the annual income at risk (Hegedűs 1994). Moving is more characteristic for high-income families, which is evidenced in the suburbanization process especially around the capital city Budapest but also around other larger towns. (Budapest lost about 15 percent of its inhabitants during the 1990s). Given the difficulties associated with changing residence, we expect that people who wish to improve their labor market situation will choose commuting, especially those who are unemployed or often face the risk of unemployment, and mostly think of commuting as the means of improving their economic condition.

The paper is organized as follows. Section 2 presents a simple model of wage determination, which allows one to study the relationship between commuting distance and wages. The key hypothesis derived from the model is that the positive effect of commuting time on wages is conditional on unemployment rate at place of work. More specifically, the higher the unemployment rate, the larger are the returns to commuting. The remainder of the paper is devoted to empirical analyses. Section 3 describes the data and variables used in the subsequent sec-

tions. Section 4 begins with the examination of the relationship between commuting time, wages and reimbursement. Then we provide estimates for the returns to commuting time in terms of both wages and the probability of receiving reimbursement of travel expenses. Section 5 concludes.

## 2 Theoretical background

### 2.1 The wage posting model

In recent years, research into the relationship between commuting and labor market outcomes has been increasingly influenced by modern theories of monopsonistic labor markets (Manning 2003a, 2003b; Rouwendal–van Ommeren 2007). This line of literature argues that commuting costs are an important source of monopsony power. Theories of monopsonistic behavior rely on wage posting models, which elaborate the simple idea that even high wages may maximize profit, provided high wages guarantee a steady flow of applicants and reduce the probability of job separations. Wage posting models assume that wages are posted by employers before workers and jobs meet and that the respective rates at which worker–job matches are formed and terminated are related to wages.

Our study of the relationship between commuting time and wages will be motivated by a simple wage posting model. The wage setting problem is formally described and analyzed in the Appendix. Here we restrict ourselves to an informal presentation.

In essence, the maximization problem consists of maximizing the expected profit associated with holding a vacancy. If vacancies were always filled, the profit would be maximized by paying the lowest possible wage, since profit in this case is the difference between the productivity of the worker and the wage. However, vacancies are not filled with probability one because workers are to some extent selective. Obviously, low wages are rejected if unemployed workers can find better jobs. Therefore, employers take into account the probability with which job seekers accept a wage offer. We refer to this probability as acceptance probability throughout this paper.

Since the acceptance probability reflects the desirability of the job offer, it is assumed to depend on the wage offer, the reservation income and the unemployment rate. First, the better the job offer, the higher is the acceptance probability. Second, independent of the wage offer, the acceptance probability is an increasing function of unemployment rate. That is, unemployed people are more likely to accept a job if unemployment is serious; and they become more selective as unemployment decreases. Even the minimum wage is acceptable, provided that the unemployment rate is sufficiently high. Finally, the acceptance probability depends on the reservation income: job seekers reject all wage offers which are below the income they can earn while being unemployed.

Under these assumptions, the profit maximizing wage depends on the productivity of the representative worker, denoted by  $y$ , the reservation income  $z$ , and the unemployment rate. More specifically, the profit maximizing wage,  $w^*$  can be shown to be the weighted sum of productivity minus search costs  $-k$  and the reservation income:

$$(1) \quad w^* = \frac{\theta(y+k)+z}{1+\theta},$$

where  $\theta$  is the ratio of the number of vacancies to the number of unemployed job searchers. (For the derivation, consult the Appendix.) Although we assume an unilateral wage posting process, the profit maximizing wage is the solution of the generalized Nash bargaining process where the worker and the employer are characterized with the respective threat points  $z$  and  $-k$ , and the „bargaining power” of the worker is  $\frac{\theta}{1+\theta}$ .

The study of the compensating wage differential paid to commuters relies on the simple assumption that the reservation income depends on commuting distance. The starting point is the assumption that workers living at distance  $t$  evaluate jobs on the basis of the difference between wage  $w$  and commuting costs (Manning 2003b). That is,  $w_t = w - ct$  where  $c$  captures both the monetary cost of travel and the monetary value of time associated with travel. As shown in the Appendix, the profit maximizing wage in spatial labor markets is

$$(2) \quad w_t^* = \frac{\theta(y+k)+(z+ct)}{1+\theta}.$$

The result is simple. In spatial labor markets, commuters behave as if their reservation income were the sum of the true reservation income and commuting costs. This implies that the profit maximizing wage offer must be higher for commuters, since the optimal wage offer is a function of the reservation income. Since our model implies that the effect of reservation income on wages is conditional on unemployment, we hypothesize that returns to commuting are larger in labor markets where unemployment is high. Employers are willing to pay higher wages to commuters because they perceive the reservation income of commuters as the sum of the true reservation income and commuting costs. This is consistent with search theoretic models claiming that the reservation wage is an increasing function of commuting distance (Rouwendal 1999).

If employers were free and able to bargain wages on an individual basis, they would pay different wages to commuters and to local residents. The wage premium, denoted by  $\Delta w(t)$ , is simply the difference between the optimal wage offers  $w_t^*$  and  $w_0^*$ , the later denoting the optimal wage as defined by equation (2). That is:

$$(3) \quad \Delta w(t) = w_t^* - w_0^* = \frac{ct}{1+\theta}.$$

If employers were free to set different wages to workers with different commuting time, the theoretical compensating wage premium would be observable



in wages. However, employers rarely pay different wages to workers with similar productivity but different residential location. Employers often avoid intra-firm wage discrimination because wage differences, often perceived as “unfair” by employees put team performance at risk (see, for example, Akerlof–Yellen 1986). Employers therefore aim at establishing a wage structure, which is relatively independent of the personal characteristics of the employed or prospective workers. Assuming that there are no systematic differences in the productivity of workers with different commuting time, employers cannot discriminate on a spatial basis. While employers are “spatially blind” during their decisions with respect to the wage structure, they might legitimately reimburse some of the travel expenses (Rouwendaal–van Ommeren 2007). Indeed, employees of various European countries, including Hungary, receive explicit reimbursement of travel expenses, the extent of which being often stipulated in collective wage agreements or in labor law. Our paper does not assume that such agreements and legal rules are binding. Instead, we argue that reimbursement equals the profit-maximizing compensating wage; that is, the theoretical compensating wage premium is paid in the form of explicit reimbursement. Similar to the argument developed in the previous paragraph, the effect of commuting distance on reimbursement must be larger in settings where unemployment is more pronounced. While employers should remain “spatially blind” during their decisions with respect to the wage structure, they can discriminate on a spatial basis and offer reimbursement to travel expenses. In short, the compensating wage premium, as defined by equation (3) is paid in the form of explicit reimbursement.

Whatever interpretation of equation (3) is adapted, the result is that returns to commuting depend on the ratio of the number of vacancies to the number of unemployed workers. Employers reimburse, either implicitly in the form of higher salaries or explicitly, the fixed fraction  $\frac{1}{1+\theta}$  of commuting costs. If the number of vacancies and the number of job seekers were the same, half of the travel expenses were reimbursed.<sup>2</sup> Travel expenses are fully reimbursed when the labor market becomes extremely tight ( $\theta \rightarrow 0$ ). Contrary to this, employers are not willing to reimburse travel expenses if there is substantial labor shortage ( $\theta \rightarrow \infty$ ). The logic is as follows. Labor shortage induces firms to increase wages up to the point where wages equal productivity. The side effect of wage competition is that wages become independent of the reservation income. In the eyes of employers, commuting costs are a component of the reservation income, and not of productivity. If firms compete with offering better reimbursement schemes, it is because there is an excess supply of labor.

The negative effect of the improvement of labor market conditions on reimbursement can be illustrated by two additional results. First, the ratio of reimbursement to the wage is  $\frac{ct}{\theta(y+k)+z}$ . Keeping commuting time constant, the rela-

2 Rouwendaal–van Ommeren (2007) arrive at the same result, albeit using a different reasoning.

tive amount of explicit reimbursement becomes smaller as the relative number of vacancies increases. Second, note that employers are not willing to reimburse the costs of too long commutes. Letting the (instantaneous) profit zero, we obtain the largest commuting time  $t^{max} = \frac{y - \theta k - z}{c}$ .

The largest commuting time reimbursed would be independent of tightness and wage distribution if search costs for new workers were zero. Otherwise the largest commuting time eligible for reimbursement becomes shorter as the ratio of vacancies to unemployed workers increases. Thus the improved chances of earning high wages reduce the chances of receiving reimbursement for longer commutes. The surprising implication is that employers are more willing to reimburse travel expenses if the unemployment rate is higher.

## 2.2 Empirical model

In this subsection, we elaborate the empirical implications of the wage posting model. To arrive at a tractable empirical model, we first rewrite Equation (2) as

$$(4) \quad w_t = \frac{u_0}{u_0 + v} [z_t + ct - y - k] + (y + k),$$

where  $u_0$  and  $v$  denote the number of unemployed people and the number of vacancies at the place of work, respectively. Since the fraction appearing in the right-hand side increases with unemployment at a decreasing rate, we use the approximation  $\frac{u_0}{u_0 + v} \approx \ln u_0 - \ln v$ .

Substituting this approximation into (4), we obtain

$$w_t = [\ln u_0 - \ln v][z + ct - y - k] + y + k.$$

In our dataset, which will be described in the next section, there is no information about the number of vacancies in local labor markets and search costs. Substituting constants for unobserved variables leads to the linear model

$$(5) \quad w_t = \beta_0 + \beta_t t + \beta_z z + \beta_u \ln u_0 + \beta_y y + \beta_{ut} t \ln u_0 + \beta_{uz} z \ln u_0 + \beta_{uy} y \ln u_0.$$

We have shown in the previous subsection that employers are not willing to compensate for long commutes that would turn the instantaneous profit to zero or even negative. In order to account for this ceiling effect, the square of commuting time can be added. The revised version of the above model is

$$(6) \quad w_t = \beta_0 + \beta_t t + \beta_{t^2} t^2 + \beta_z z + \beta_u \ln u_0 + \beta_y y + \beta_{ut} t \ln u_0 + \beta_{ut^2} t^2 \ln u_0 + \beta_{uz} z \ln u_0 + \beta_{uy} y \ln u_0$$

Interest centers on the coefficients (main effects) of commuting time and on the coefficients of the product terms. The compensating wage approach to com-

muting time research usually boils down to estimating regression models which involve commuting time as well as personal and eventually firm-level and regional characteristics (Leigh 1986; Zax 1991; Manning 2003b). Our empirical model differs from earlier models in two important respects. First, we also add the square of commuting time to allow for a ceiling effect, which is related to the fact that neither workers nor employers tolerate too long commuting. Second, we also include interaction terms between unemployment and human capital characteristics, meaning that an increase in unemployment should reduce the returns to productivity. If our model is correct then previous regression models are misspecified and regression estimates of the compensating wage differential are biased.<sup>3</sup>

Explicit reimbursement, denoted by  $R_t$  will be studied using the same logic. Using the above approximation, equation (3) becomes

$$R_t = [\ln u_0 - \ln v][z + ct].$$

Given the limitations of our data, our empirical reimbursement model is

$$(7) \quad R_t = \beta_0 + \beta_t t + \beta_z z + \beta_u \ln u_0 + \beta_{ut} t \ln u_0 + \beta_{uz} z \ln u_0.$$

If there is a ceiling effect, the model is extended to

$$(8) \quad R_t = \beta_0 + \beta_t t + \beta_{t2} t^2 + \beta_z z + \beta_u \ln u_0 + \beta_{ut} t \ln u_0 + \beta_{ut2} t^2 \ln u_0 + \beta_{uz} z \ln u_0$$

### 3 Data

In April 2001, a survey was conducted among registered unemployed who were entitled to unemployment benefits (N=105,924) and eventually found a job between the 18th of March and 7th of April 2001. The primary purpose of data collection was the evaluation of the effect of the dramatic rise of the minimum wage on changes in unemployment.<sup>4</sup> In the above mentioned period, 9474 people got a job, out of which 8339 people completed the questionnaire (Köllő 2002). The questionnaire contains both retrospective questions about the previous job and questions about the new job. This paper will use a subset of the data, consisting of 801 observations.

Survey data are rarely free of data problems. In our dataset, two problems are of special interest. First, respondents who were reemployed by the former employer were not asked about the receipt of reimbursement. Since commuting costs cannot be assumed to remain constant, these cases must be excluded. Our sample therefore is restricted to job changers. Second, when asked about the prospective job, respondents were asked to estimate the lower and the upper bounds of the salary. Unfortunately the reported minimums and maximums differ substantially

3 The expectation that unemployment rate should have a positive regression coefficient does not contradict the fact that unemployment is negatively associated with individual wages. A small increase in log unemployment changes wages by  $\beta_u + \beta_{ut} t + \beta_{ut2} t^2 + \beta_{uz} z + \beta_{uy} y$ , which can (and should) be negative.

4 In January 2001, the minimum wage rised from 25.5 to 40 thousand HUF

in a considerable proportion of cases. We omitted respondents where the difference between the maximum and the minimum exceeds 10 thousand HUF.

Since our focus is on the effect of commuting, and migration might disturb the empirical relationship between commuting time and commuting decisions (Ihlanfeldt–Sjoquist 1998), we exclude those unemployed who changed their place of residence during their unemployment spell. Since we wish to generalize our results to the population of job seekers with low education, we omitted respondents with college or university education. The sample selected for analyses include full-time employees aged 15–74 in 2001, who travel to work and back no more than four hours. Note that the sample includes cases where none of the variables take missing values. As a result of these decisions, we are left with 783 observations for further analyses.

Our interest centers on the relationship between wages, commuting time and reimbursement. The hourly wage variable is the reported gross monthly salary and is measured in thousand HUF. Commuting time is the time spent on traveling on an average day. Reimbursement is a dummy variable indicating respondents who either received some reimbursement of travel expenses or were transported to work on the cost of the employers. Note that we do not know the exact amount of money received by the workers.

The productivity of workers is captured by gender, a dummy indicating general high-school education and experience. The latter variable measures the number of years elapsed since the first entry to the labor market, minus the years having been unemployed. The reservation income is captured by the last gross wage (measured in thousands of HUF) and the unemployment rate at the place of residence. In our paper, all unemployment figures were computed using the 2000 wave of the TSTAR database of the Hungarian Statistical Office. They actually measure the average number of registered unemployed divided by the size of the active population. The hourly wage variable is the reported gross monthly salary and is measured in thousands of HUF.

*Table 1* shows the means and standard deviations of the variables used in subsequent analyses. The average wage exceeds the minimum wage by 7.5 thousand HUF among women and 12 thousand HUF among men. 44 percent of women and 52 percent of men receive some compensation for travel expenses. Average commuting time is 0.88 hours (53 minutes) among women and one hour among men, the grand mean being 56 minutes. The average commuter thus does not travel more than one hour per day.

Table 1 Means and standard deviations of variables

Variable	Women (N=344)		Men (N=439)		Total (N=783)	
	mean	sd	mean	sd	mean	sd
<i>Dependent variables</i>						
gross monthly wage	47.536	12.333	55.764	17.658	52.149	16.064
log monthly wage	3.836	0.212	3.978	0.283	3.916	0.264
Receipt of reimbursement	0.445	0.498	0.517	0.5	0.485	0.5
<i>Independent variables</i>						
Commuting time	0.881	0.638	0.987	0.714	0.941	0.683
Commuting time squared	1.182	1.893	1.483	2.283	1.351	2.124
Last monthly wage	45.163	48.419	52.417	28.782	49.23	38.798
Unemployment rate at place of residence	5.488	2.77	5.979	3.512	5.764	3.215
Log unemployment rate at place of work	0.74	0.159	0.753	0.158	0.747	0.158
<i>Interaction of log unemployment at workplace with</i>						
Commuting time	0.631	0.447	0.722	0.516	0.682	0.489
Commuting time squared	0.82	1.264	1.06	1.58	0.955	1.453
Last monthly wage	32.595	33.343	38.674	21.476	36.003	27.478
Unemployment rate at place of residence	4.387	3.284	4.877	4.246	4.662	3.858

Our theoretical assumption is that persistent unemployment is maintained by the lack of spatial mobility. A brief comparison of our estimates to estimates presented in other studies shows that Hungarian workers do not lack spatial mobility in international comparison. Using Dutch aggregate statistics, van der Vlist (2001) reports an average commuting distance of 17.5 km among men and 11.0 km among women (the gross average being 15.3 km) for 1997. In Hungary, traveling 15 kilometers using public transportation costs about 30 minutes, so the approximately one hour commuting time seems to be consistent with the Dutch findings. Using data from another Dutch survey conducted in 1998, Rouwendal-van Ommeren (2008) report an average of one hour for workers with reimbursement and half an hour for workers without reimbursement. Since 46% of the sample received reimbursement, the sample average is about 40 minutes. Almost the same figure, about 45 minutes is reported by Manning (2003) using the British Labour Force Survey for 1993-2001 and the British Household Panel Survey for 1991-2000. To summarize, the workers we study do not travel less than workers in Britain or the Netherlands. This is striking because our sample does omit people with good education and high earnings, who tend to commute larger distances (see, for example, van der Vlist [2001]).

## 4 Empirical analyses

### 4.1 *The relationship between commuting time, wages and reimbursement*

Before making any attempt to explain the relationship between commuting and wages, we first examine the question whether there is any relationship to explain. Table 2 shows the distribution of daily commuting time, as well as mean wages and the level of reimbursement of travel expenses as a function of commuting time. The distributions are presented separately for female and male workers. The vast majority of workers (80 percent of women and 74 percent of men) do not travel more than one hour, and the proportion of workers commuting more than 2 hours is very low. The distribution of commuting distances resembles the exponential distribution among women and the log-normal distribution among men. This pattern is not surprising: since the household is supposed to be run by women, they find longer commutes more costly than men. Our descriptive findings are similar to those reported in the literature on commuting<sup>5</sup>

Table 2 Distribution of wages and reimbursement by commuting distance

Commuting time	Women			Men		
	N	Mean wage	% receive reimbursement	N	Mean wage	% receive reimbursement
1–30 minutes	144	45.94	15.97	156	51.97	19.87
31–60 minutes	133	46.99	55.64	167	56.49	58.68
61–90 minutes	27	46.59	70.37	44	57.31	72.73
91–120 minutes	26	51.6	92.31	50	60.3	90.00
121–180 minutes	12	66.04	91.67	19	63.45	94.74
181–240 minutes	2	47.5	100.00	3	65.67	100.00

The distribution of commuting time is not surprising if we look at the relationship between wages and commuting time. While average wages are monotonically increasing with commuting time among men, an inverted U shaped pattern describes the relationship among women. Given our estimates, men and women should follow different commuting strategies in order to realize the highest marginal increase in wages. Men might find it rational to commute 31–60 minutes instead of 1–30 minutes because this change improves the wage of the average male worker by about 10 percent. However, women do not gain anything from commut-

5 For the United Kingdom, Manning (2003b) finds that about 80 percent of employees commute less than or equal to one hour; for selected cities, an average of 4.5 km is found (Frost–Linneker 1998). For the Netherlands, van Ommeren (1996) found that half of the workers commute less than 8 kilometers and only 10 percent of workers commute more than 32 kilometers. In terms of commuting time, half of the workers commute less than 20 minutes. (These estimates use the so-called Enquete Beroepsvolking, having been conducted in 1992.) For the United States, estimates are 8.7 miles (Hamilton 1982), estimates in commuting time are 22.5 minutes (White 1988). A recent overview of empirical findings (Rodríguez 2004) suggest that average commuting time in US cities ranges between 14 and 23 minutes and average commuting distance ranges between 14 and 25 kms.

ing 31–60 minutes or even one and half hours; if they wish to improve their wages substantially, they should commute 2–3 hours.

The proportion of people receiving any reimbursement is an increasing function of commuting time among both sexes. Especially commuters traveling more than one and half hours receive some reimbursement with a probability equal or larger than 90 percent. However, short travels, not exceeding the half an hour value, are not likely to be covered by employers. Unlike wages, the relationship between reimbursement and commuting time does not differ substantially between men and women.

*Table 2* also shows that reimbursement and wages are positively correlated among men. Keeping daily commuting time constant, the higher the wages, the higher the probability of receiving some reimbursement. This finding might seem to contradict our theoretical model assuming the absence of intra-firm wage discrimination. If workers oppose intra-firm wage discrimination, employers are forced to pay the compensating wage premium in the form of explicit reimbursement, being independent of actual wages. The assumption of wage bargaining, which allows for intra-firm discrimination, seems to be more capable of explaining the positive correlation. Employers ready to pay high wages can pay a relatively small proportion of wages in the form of explicit reimbursement. The underlying incentive is tax-evasion: explicit reimbursement in Hungary is not taxed, therefore both employers and workers are interested in receiving a part of the wage in the form of reimbursement (see for example, Rouwendal–van Ommeren 2008). Since firms paying (and workers receiving) high wages gain more from tax evasion, the correlation between wages and reimbursement must be positive. However, the conclusion that a positive association between wages and reimbursement is at odds with the assumption of no intra-firm wage discrimination is premature. First, the positive correlation between wages and reimbursement might be due to the presence of a common cause. For instance, larger firms pay higher wages, and these are large firms as well who are more likely to comply with the legal rules prescribing the reimbursement of travel expenses. It is also possible that respondents who received some reimbursement misinterpreted the survey question concerning the wage and reported a higher figure.<sup>6</sup>

## 4.2 The effect of commuting time on wages

We proceed with the regression analysis of the relationship between wages and commuting time in order to estimate the net effect of commuting time. We will estimate the models as specified in Equations (5) and (6), being labeled the linear and curvilinear specifications, respectively. The models are estimated using ordinary least squares. In the literature, returns to commuting time are often

<sup>6</sup> Unfortunately, the survey question did not make it clear to the respondents that they should not think of reimbursement when they estimate or tell their wages.

estimated using household or individual level fixed effects regressions. The aim of this modeling strategy is to minimize the bias arising from endogenous residential choices and to remove spurious correlations arising from the effect of unobserved characteristics on both wages and commuting time. Endogenous moving are not a concern here because our sample does not include people who have changed place of residence. We believe that the last wage, which is intended to capture the reservation income, also reflects unobserved personality traits. The assumption here is that employers can observe the personality traits that were hidden at the beginning of the match and update their beliefs about workers' productive abilities. The last wage variable refers to the end of a worker-job match, thus it can be expected to incorporate the employers' assessment of productive abilities. We therefore use simple ordinary least squares instead of using fixed-effects regressions.

Estimation results are presented in *Table 3*. While the coefficient of commuting time lacks statistical significance in the linear specification (Model 1), it is significant in the curvilinear specification (Model 2). The same applies to the interaction between commuting time and log unemployment at place of work. The significance level of other variables is not affected by the choice of specification. The interpretation of the results therefore is based on the estimates of the curvilinear specification (Model 2). As we well see, the results from Model 2 also explain why we failed to find a significant effect of commuting time in Model 1.



Table 3 OLS estimates of log monthly wages

Variable	All		Women		Men	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
<i>Main effects</i>						
Commuting time	0.0428 (0.262)	0.4536* (0.003)	-0.0031 (0.486)	0.5083* (0.032)	0.0391 (0.348)	0.4695* (0.02)
Commuting time squared		-0.1572* (0.006)		-0.2141* (0.037)		-0.1502* (0.018)
Last monthly wage	0.0003 (0.457)	0.0003 (0.447)	0.0042 (0.171)	0.0045 (0.161)	0.003 (0.125)	0.003 (0.123)
Unemployment rate at place of residence	-0.0288* (0.003)	-0.0313* (0.001)	-0.0027 (0.425)	0.0007 (0.48)	-0.0272* (0.017)	-0.0311* (0.01)
Experience	0.0022 (0.428)	0.0022 (0.425)	-0.0111 (0.272)	-0.0082 (0.331)	0.001 (0.478)	-0.0011 (0.476)
Experience squared	0.0000 (0.45)	0.0000 (0.479)	0.0004 (0.26)	0.0003 (0.323)	0.0001 (0.436)	0.0001 (0.405)
High-school education	0.201* (0.022)	0.2135* (0.016)	0.0441 (0.374)	0.0775 (0.288)	0.3332* (0.024)	0.329* (0.025)
Male	0.3898* (0.000)	0.3796* (0.000)				
Log unemployment rate at place of work	-0.1649 (0.183)	0.0543 (0.394)	-0.1217 (0.358)	0.2629 (0.241)	-0.3134 (0.103)	-0.1047 (0.344)
<i>Interaction of log unemployment at workplace with</i>						
Commuting time	0.033 (0.354)	-0.5137* (0.01)	0.0855 (0.229)	-0.7053* (0.041)	0.0369 (0.39)	-0.4746 (0.055)
Commuting time squared		0.2154* (0.011)		0.3357* (0.037)		0.1842* (0.036)
Last monthly wage	0.0014 (0.325)	0.0013 (0.335)	-0.0054 (0.192)	-0.0059 (0.181)	0.0003 (0.464)	0.0003 (0.464)
Unemployment rate at place of residence	0.0217* (0.012)	0.024* (0.007)	-0.0053 (0.364)	-0.0098 (0.262)	0.0202* (0.042)	0.024* (0.026)
Experience	-0.0004 (0.49)	-0.0006 (0.484)	0.0284 (0.111)	0.0236 (0.16)	-0.0034 (0.437)	-0.0008 (0.486)
Experience squared	-0.0001 (0.381)	-0.0001 (0.412)	-0.0009 (0.115)	-0.0007 (0.171)	-0.0001 (0.464)	-0.0001 (0.43)
High-school education	-0.1038 (0.201)	-0.1225 (0.161)	0.1344 (0.221)	0.0865 (0.313)	-0.3394 (0.055)	-0.3369 (0.055)
Male	-0.3189* (0.001)	-0.3069* (0.002)				
Constant	3.8573 (0.000)	3.691 (0.000)	3.8007 (0.000)	3.5484 (0.000)	4.046 (0.000)	3.8663 (0.000)
N	783	783	344	344	439	439

Notes: Numbers in parentheses are p-values. Coefficients significant at the 5 percent level are marked by asterisk.

Since our regression models include interaction terms, the positive main effect of commuting time together with the negative main effect of its square does not imply that there is an inverted U shaped relationship between commuting time and wages. The main effects of commuting time variables are meaningful in an economy where unemployment rate at place of work is 1 percent.<sup>7</sup> If this were the case, there is indeed an inverted U shaped relationship; the wage-maximizing commuting time is about 71 minutes among women and 94 minutes among men. Note, however, that the interaction between log unemployment and commuting time is positive and the the interaction between log unemployment and commuting time is negative. This means that as unemployment at place of work increases, the relationship between commuting time and wages first becomes more and more flat then U shaped. Another implication is that the wage-maximizing commuting time decreases as the unemployment at place of work increases. Among women, the predicted optimal commuting time is zero (or lower than zero) if unemployment at place of work is 5 percent or higher. Among men, the predicted optimal commuting time reduces to zero if unemployment at place of work is about 10 percent or higher. In our sample, the average of unemployment at place of work is about 5 percent, which explains why the commuting time variable lacked statistical significance in the linear specification (Model 1).

Our wage posting framework implies that unemployment at place of work modifies the returns to human capital and the reservation income. More specifically, if unemployment in the center of local labor markets increase, returns to human capital should decrease but returns to the reservation income should increase. Our results do not support this prediction unambiguously. The main problem is with commuting time. Unemployment does affect returns to commuting time, but the direction of the effect is negative instead of being positive. Commuting time is a component of the reservation income in the model, therefore the returns to commuting time should increase with unemployment at place of work. The evidence presented here clearly contradicts this expectation. Within our theoretical model, the unexpected negative interaction effect can be explained in three different ways. First, one might assume that commuting increases productivity: since unemployment lowers the returns to productivity, the negative interaction effect between commuting time and productivity obtains. The assumption of a positive relationship between commuting distance and productivity were realistic in a sample of qualified white-collar workers who moved to suburbs and commute to well paid jobs, or in a sample of urban residents who work in rural areas. Since our sample includes people with low educational levels, and mainly rural residents who work in urban areas, this explanation can be rejected. Second, one might argue

7 The unemployment at the place of work variable ( $u$ ) was logarithmized using the transformation  $\log_{10}(u+1)$ . The wage maximizing commuting time is defined by  $\frac{1}{2} \frac{b_1 + c_1 \log_{10}(u+1)}{b_2 + c_2 \log_{10}(u+1)}$ , where  $b_1$  is the main effect of commuting time,  $b_2$  is the main effect of the square of commuting time,  $c_1$  is the interaction between log unemployment and commuting time, and  $c_2$  is the interaction between log unemployment and the square of commuting time.

that searching for employees who live in spatially remote areas is more costly than to search for local workers. This assumption is standard in the literature on spatial mismatch (see Gobillon–Selod–Zenou 2007). The positive association between commuting distance and search costs is also consistent with the hypothesis that employers prefer informal employee referrals as the means of filling vacancies. If personal contacts are more likely to emerge with commuting distance, employers who wish to hire spatially remote workers cannot rely on referrals and must incur some recruitment costs. The final logical possibility is that one assumes that the reservation income decreases with commuting time. This assumption is realistic if the representative commuter is engaged in informal economic activities as well, so that she establishes a lower reservation level towards market income.

Another puzzling finding is the positive interaction effect between log unemployment at workplace and unemployment at residence among men. Our model implies that unemployment at place of work should decrease the (negative) effect of unemployment at place of residence on wages since the latter is negatively associated with the reservation income. The finding can be explained with the assumption that unemployment at place of residence also expresses a low level of productivity (Gobillon–Selod–Zenou 2007). One reason is that longer trips make workers tired, and commuters are more likely to be late, especially if public transport is bad. Another reason is territorial discrimination, emerging from the spatial segregation of ethnic minorities. In Hungary, a substantial proportion of the discriminated Roma minority lives in small villages far from urban areas, thus pessimistic expectations concerning the productivity of Roma should overlap with pessimistic expectations about the productivity of commuters. The spatial location of the worker thus signals not only a low reservation income but also a low level of productivity.

One should also note that unemployment at place of work does not seem to modify the returns to human capital, with the exception of gender in the full sample. A straightforward explanation of the lack of empirical support concerns the characteristic of our sample. First, the sample includes people who were successful in escaping unemployment. Since productive abilities deteriorate during unemployment, it might be the case that employers think of past unemployment of individuals as a dominant signal of productive abilities, which suppress other available information, like education and experience. Besides, the discovery of interaction effects is usually difficult in samples which are larger than our sample.

### 4.3 The effect of commuting time on reimbursement

Finally, we examine the effect of commuting time on reimbursement. Again, we estimate both a linear and a curvilinear specification, as described in Equations (7) and (8). The estimation method is logistic regression. We omit human

capital variables since they do not appear at the right-hand sides of the equations. But we do include monthly wages in order to check whether reimbursement is independent of the wage. Under the assumption of no intra-firm discrimination, the probability of receiving some reimbursement must be independent of wages. If, however, reimbursement is a form of tax evasion, we should observe a positive relationship between reimbursement and wages, since the gains from tax evasion increase with wages.

Table 4 Logistic regression estimates of receipt of reimbursement

Variable	All		Women		Men	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
<i>Main effects</i>						
Commuting time	0.2567 (0.758)	0.7274 (0.734)	1.6692 (0.193)	4.7338 (0.228)	-0.6737 (0.569)	-0.1338 (0.962)
Commuting time squared		0.0921 (0.901)		-1.1379 (0.449)		0.1064 (0.909)
Last monthly wage	-0.0056 (0.777)	-0.0043 (0.83)	0.0345 (0.423)	0.039 (0.37)	0.0012 (0.963)	0.0011 (0.965)
Unemployment rate at place of residence	0.6376* (0.000)	0.6118* (0.000)	1.0679* (0.000)	1.0099* (0.000)	0.5116* (0.006)	0.5015* (0.005)
Log unemployment rate at place of work	-4.172* (0.041)	-4.419 (0.074)	0.2036 (0.958)	1.4248 (0.767)	-4.130 (0.122)	-4.3302 (0.174)
Gross monthly wage	0.009 (0.144)	0.0097 (0.125)	0.0472* (0.003)	0.0500* (0.003)	-0.0059 (0.427)	-0.0059 (0.436)
Male	-0.831 (0.374)	-0.853 (0.364)				
<i>Interaction of log unemployment at workplace with</i>						
Commuting time	2.7352* (0.015)	4.1228 (0.171)	1.0587 (0.51)	-0.8962 (0.874)	3.9504* (0.018)	5.0329 (0.199)
Commuting time squared		-1.0486 (0.345)		0.5465 (0.813)		-0.9616 (0.497)
Last monthly wage	0.0091 (0.749)	0.0074 (0.799)	-0.0553 (0.377)	-0.0612 (0.333)	0.0095 (0.778)	0.0095 (0.778)
Unemployment rate at place of residence	-0.4532* (0.000)	-0.4241* (0.001)	-0.8071* (0.002)	-0.7468* (0.003)	-0.3655* (0.028)	-0.3488* (0.025)
Male	1.2174 (0.336)	1.2361 (0.33)				
Constant	-0.9986 (0.486)	-1.4744 (0.398)	-6.6813 (0.015)	-8.373 (0.015)	-0.1575 (0.936)	-0.6072 (0.794)
N	783	783	344	344	439	439
Log-likelihood	-397.624	-392.626	-162.534	-160.141	-225.606	-223.159
$\chi^2$	121.568	178.736	64.322	88.852	67.162	93.937

Notes: Numbers in parentheses are *p*-values. Coefficients significant at the 5 percent level are marked by asterisk.

The estimated coefficients are shown in *Table 4*. Contrary to the previous wage equations, the linear specification seems to perform better than the curvilinear one. Although the main effect of commuting time lacks statistical significance, its interaction with unemployment at place of work is positive and significant in the sample of men. This means that longer commutes imply a higher probability of receiving reimbursement only if the unemployment rate is sufficiently high at the place of work. Our estimates suggest that sufficiently high means an unemployment rate of about ten percent or higher. This pattern clearly supports the theoretical prediction that unemployment should increase the returns to commuting in terms of reimbursement. Unfortunately, the interaction effect is not significant among women.

The interaction between the two unemployment variables is also significant. The negative sign is consistent with the assumption that the unemployment rate at place of work is a negative indicator of reservation income, but contradicts the assumption that it is a negative indicator of productivity. In the previous analysis, we interpreted the unexpected positive interaction of the unemployment variables using the assumption that unemployment at place of residence expresses a low level of productivity. The change in the interpretation seems to be a contradiction. However, one should keep in mind that the decision on reimbursement differs from wage setting in an important respect: employers do consider productivity when deciding on wages, but they do not consider productivity when making decisions on reimbursement. If workers oppose intra-firm wage discrimination and employers establish a wage structure, the decision to reimburse travel expenses is made independently of productivity. Therefore, the finding of a negative interaction between the unemployment variables in the reimbursement regression does not contradict the positive interaction between the same variables in the wage regression, since only the latter is affected by the fact that employers think of unemployment at place of work as a negative signal of productivity.

## 5 Conclusions

Observers of the persistent regional differences in unemployment argued that high costs of commuting prevent residents of high unemployment areas from finding employment in other areas. This paper examines the relationship between commuting distance, on the one hand, and wages and the receipt of explicit reimbursement, on the other. We develop a simple wage posting model, which implies not only a positive effect of commuting time on wages and reimbursement, but also that returns to commuting must be larger if the unemployment rate where jobs are situated is large as well. We test our predictions using data on low educated workers who were registered unemployed and got a job in march 2001.

We find some evidence that returns to commuting is indeed conditional on

unemployment rate. First, we find a positive effect of commuting time on wages, which, however, decreases with unemployment at place of work. This conditional effect contradicts our expectation of a positive impact of unemployment on returns to commuting time. However, the found pattern can be explained with the assumption that commuting time is also a signal of low productivity: if this is the case, the observation is consistent with the wage posting model, which hypothesizes a negative effect of unemployment on returns to productivity. We also find evidence among male workers that commuting time increases the probability of receiving some reimbursement of travel expenses, conditional on high unemployment at place of work. This finding is consistent with our theoretical framework.

The research outlined in this article is an attempt to contribute to the explanation of persistent regional inequalities in Hungary. In regions where unemployment is high, the unemployment rate in the large towns, which can be considered as centers of local labor markets, have an unemployment rate of approximately 10 percent. The findings suggest that in such a labor market commuting time is not associated with a compensating wage premium, but it is associated with a higher probability of receiving reimbursement. Although we did not observe the actual value of reimbursement, it is safe to assume that all commuting costs are never reimbursed. First, labor law prescribes that employers must reimburse a percentage of the monetary costs of travel, which is lower than 100 percent. Second, employers probably do not wish to reimburse the monetary value of time spent on traveling. These facts together imply that employers do not compensate for costly commuting. The present study therefore supports the conclusion of previous studies: commuting is too costly to induce people living in high unemployment regions to find a work in urban areas (Köllő 1997, 2006; Kertesi 2000).

The findings might suggest that reimbursement of expenses on the part of employers is a necessary condition for the reduction of persistent regional inequalities. This conclusion, however, neglects the possibility that employers will reduce labor demand as a reaction to increases in labor costs. If employers cut labor demand, it is difficult to predict the net effect of reimbursement of expenses on regional differences in unemployment rates. Without knowing the precise effect of reimbursement of travel expenses on labor demand, it is impossible to formulate firm policy recommendations.

A fundamental limitation of our study is absence of information on the level of reimbursement. Our theoretical model predicts a trade-off between wages and reimbursement. An exact test of the model predictions requires data about the amount of reimbursement received. Unfortunately, we do not have such data at our disposal. Therefore, our interpretation of the evidence is not the final word on the subject.

Our study has limitations because of the sample and the estimation method we used. A substantial limitation of our study is that our sample is probably not free of

sample selection problems (Cooke–Ross 1999). Our sample was taken from a survey of unemployed, and unsuccessful job searchers are not included in the sample. This might lead to the problem of self-selection if unobserved factors determining the success of job search (getting a job) are correlated with unobserved determinants of wages or commuting decisions.

## Appendix. The wage posting model

In the labor market, employers face the problem of setting a profit maximizing wage. Output is the sum of individual outputs, thus profit maximization boils down to maximizing the expected discounted lifetime value of a job. In other words, we proceed as if firms were sums of one-job firms (Pissarides 2000).

Jobs are either vacant (state 0) or filled (state 1). If a job is vacant, employers search for workers and incur fixed cost  $k$ . Vacant jobs are contacted by unemployed workers at the exogenous arrival rate  $\lambda(\theta)$ , where  $\theta$  denotes the ratio of the number of vacancies to the number of unemployed workers. If vacancies and unemployed meet each other randomly,  $\lambda(\theta) = 1 - \exp(-\frac{1}{\theta})$ .

Unemployed workers accept the wage offer  $w$  with acceptance probability  $\alpha(w)$ . The choice of the functional form is motivated by the assumption that the acceptance probability must reflect the utility of the representative worker. This is achieved by normalizing the value of job offers to the unit interval. Let  $z$  be the income received while unemployed. It is reasonable to assume that wages have an upper bound  $\bar{w}$  which is the revenue of the most productive firm within the industry under study. In a vivid labor market, wages must fall into the  $[z, \bar{w}]$  interval. The acceptance probability is defined as

$$(A1) \quad \alpha(w) = \frac{w-z\sigma^\theta}{\bar{w}-z}$$

where  $\sigma$  captures the shape of the wage distribution. For analytical simplicity, we assumed that the relative frequency is weakly decreasing in wages, so that  $0 < \sigma \leq 1$ .

Our specification of the acceptance probability reflects not only the fact that better offers are more likely to be accepted but also the fact that a particular job offer becomes more attractive as the labor market becomes less tight or the wage distribution becomes more dispersed. Equation (A1) implies that the maximum wage offer  $\bar{w}$  is always accepted. Another implication is that unemployed workers will accept job offers equaling the reservation income, provided that there is no wage dispersion ( $\sigma = 0$ ), or there are no alternative offers so that the ratio of vacancies to the number of unemployed is (close to) zero.<sup>8</sup>

The acceptance probability plays a similar role to that of reservation wage in standard search models. Standard search theory argues that the reservation wage

<sup>8</sup> The acceptance probability might also reflect preferences towards risk if  $\sigma^\theta$  is multiplied by a parameter  $\rho$  so that  $\rho < 1$  indicates risk aversion and  $\rho = 1$  indicates risk neutrality

equals the reservation income  $z$  plus the product of the arrival rate  $\lambda(\theta)$  and the expected value of the (truncated) wage distribution above the reservation wage. The expected value of the truncated wage distribution is an increasing function of the dispersion of wages, since workers can expect a higher return to job search if wages are more dispersed. Thus, reservation wage is proportional to the product of the arrival rate and wage dispersion, and effect which is captured by the shape parameter  $\sigma\theta$  in Equation (A1).

To compare the acceptance probability with the traditional reservation wage, it is useful to examine the effect of labor market tightness on both the reservation wage and the acceptance probability. First consider the case when there are no vacancies, implying that  $\theta = 0$ . In this case, the reservation wage equals the reservation income  $z$ , implying that if a vacancy were created, it would be taken by the worker, regardless of the wage. Our acceptance probability implies the same since  $\alpha(w|\theta = 0) = 1$  for all  $w$ . As the number of vacancies increase, the arrival rate of job offers to workers increases as well, which implies an increase in the reservation wage. The increase in the relative number of vacancies changes the shape of the acceptance probability so that bad offers are accepted by a nonzero probability. If the number of vacancies equal or exceed the number of job searchers, the probability of accepting bad offers approaches zero. Note that the reservation wage can be related to the acceptance probability by assuming that wage offers equal to the reservation wage are accepted by probability  $\frac{1}{2}$ .

When a match is formed, the worker produces  $y$  units sold at unit price. Matches break up at the endogenous rate  $\lambda(\theta)[1 - \alpha(w)]$ .<sup>9</sup> That is, keeping the arrival rate constant, job separations are more likely in jobs that pay low wages.

Employers aim to establish a wage which maximizes the net present value of holding a vacancy for an infinitely long period of time. The respective Bellman equations describing the net present value of vacant and filled jobs are

$$(A2) \quad rV_0(w) = -k + \lambda(\theta)\alpha(w)[V_1(w) - V_0(w)]$$

$$(A3) \quad rV_1(w) = (y - w) + \lambda(\theta)[1 - \alpha(w)][V_0(w) - V_1(w)]$$

Substitution of (A3) into (A2) and the assumption that  $r^2=0$  yields the following expression for the value of vacant jobs:

$$(A4) \quad rV_0(w) = -r\lambda(\theta)^{-1}k + [\alpha(w)(y + k - w) - k].$$

Interpretation of Equation (A4) is straightforward. Assuming zero search costs, the the net present value of a job, independently of being vacant or filled,

<sup>9</sup> Usually, search models assume an exogenous separation rate or a combination of exogenous and endogenous components (for example, Manning 2003a). The implications of the model presented here are not affected by neglecting the exogenous part of job separation rate.



equals the expected instantaneous profit. Since the probability of filling the vacancy increases in the wage, which, in turn, decreases the instantaneous profit, there is an inverted U shaped relationship between wages and the value of the job.

Assuming that the discount parameter is close to zero, differentiating (A4) with respect to  $w$  leads to the first order condition

$$y + k - w^* = \frac{\alpha(w^*)}{\alpha'(w^*)} = \frac{w^* - z}{\sigma\theta}, \text{ which immediately yields the solution}$$

$$(A5) \quad w^* = \frac{\sigma\theta(y+k)+z}{1+\sigma\theta}.$$

Equation (1) in the main text obtains after constraining the shape parameter to one. This constrain is imposed only for analytical simplicity.

In spatial labor markets, the net wage is the gross wage minus the costs of commuting. The acceptance probability for wage offers at distance  $t$  can be defined as

$$(A6) \quad \alpha(w) = \frac{w-ct-z\sigma\theta}{w-z},$$

where  $c$  is the cost associated with traveling one hour. After inserting (A6) into equations (A2) and (A3), we proceed as before. We again substitute equation (A2) into (A3) and then differentiate a modified version of equation (A4) with respect to  $w$ . The solution for the profit maximizing wage is:

$$(A7) \quad w^* = \frac{\sigma\theta(y+k)+(z+ct)}{1+\sigma\theta}.$$

Equation (2) in the main text obtains after constraining the shape parameter to one.

## References

- Akerlof, George A. – Janet L. Yellen (eds.) (1986): *Efficiency Wage Models of the Labor Market*. Cambridge: Cambridge University Press.
- Cooke, Thomas J. – Stephen L. Ross (1999): Sample Selection Bias in Models of Commuting Time. *Urban Studies*, 36(2): 1597–1611.
- Dargay, Joyce M. – Jos van Ommeren (2008): The Effect of Income on Commuting Time Using Panel Data. 45th Conference of the European Regional Science Association.
- Frost, Martin – Brian Linneker (1998): Excess or Wasteful Commuting in a Selection of British Cities. *Transportation Research A*, 32(7): 529–538.
- Gobillon, Laurent – Harris Selod – Yves Zenou (2007): The Mechanisms of Spatial Mismatch. *Urban Studies*, 44(12): 2401–2427.
- Hamilton, Bruce W. (1982): Wasteful Commuting. *Journal of Political Economy*, 90(5): 1035–1053.

- Hegedűs, József (2004): The Housing Market and Residential Regional Mobility in the 1990s. The Case of Hungary. 73–80. In *The Hungarian Labour Market. Review and Analysis*. K. Fazekas – J. Koltay – Zs. Cseres-Gergely (eds.). Budapest: Institute of Economics, HAS and Hungarian Employment Foundation.
- Ihlanfeldt, Keith R. – David L. Sjoquist (1998): The Spatial Mismatch Hypothesis: A Review of Recent Studies and their Implications for Welfare Reform. *Housing Policy Debate*, (9): 849–892.
- Kain, John (1992): The Spatial Mismatch Hypothesis: Three Decades Later. *Housing Policy Debate*, (3): 371–460.
- Kertesi, Gábor (2000): Ingázás a falusi Magyarországon. Egy megoldatlan probléma. *Közgazdasági Szemle*, (47): 775–798.
- Köllő, János (1997): A napi ingázás feltételei és a helyi munkanélküliség Magyarországon: számítások és számpéldák. *Esély*, 2.
- Köllő, János (2002): Az ingázási költségek szerepe a regionális munkanélküliségi különbség fenntartásában. *Budapesti Munkagazdaságtani Füzetek*, 2. Budapest: MTA Közgazdaságtudományi Kutatóközpont.
- Köllő, János (2006): A napi ingázás feltételei és a helyi munkanélküliség Magyarországon. Újabb számítások és számpéldák *Budapest Working Papers on the Labour Market*, 1.
- Leigh, J. Paul (1986): Are Compensating Vages Paid for Time Spent Commuting? *Applied Economics*, (18): 1203–1213.
- Manning, Alan (2003a): *Monopsony in Motion: Imperfect Competition in Labor Markets*. Princeton, NJ: Princeton University Press.
- Manning, Alan (2003b): The Real Thin Theory: Monopsony in Modern Labour Markets. *Labour Economics*, (10): 105–131.
- Pissarides, Christopher (2000): *Equilibrium Unemployment Theory* (second edition). Cambridge (MA): MIT Press.
- Rodriguez, Daniel A. (2004): Spatial Choices and Excess Commuting: A Case Study of Bank Tellers in Bogota, Columbia. *Journal of Transport Geography*, (12): 49–61.
- Romani, Jordi – Surinach, Javier – Manual Artis (2003): Are Commuting and Residential Mobility Decisions Simultaneous? The Case of Catalonia, Spain. *Regional Studies*, 37(8): 81–826.
- Rouwendal, Jan (1999): Spatial Job Search and Commuting Distances. *Regional Science and Urban Economics*, (29): 491–517.
- Rouwendal, Jan – Jos van Ommeren (2007): Recruitment in a Monopsonistic Labour Market: Will Travel Costs be reimbursed? *Tinbergen Institute Discussion Paper*, TI 2007-044/3.
- van der Vlist, Arnold J. (2001): *Residential Mobility and Commuting*. Amsterdam: Thela Thesis.
- van Ommeren, Jos (1996): *Commuting and Relocation of Jobs and Residences: A*

*Search Perspective*. Amsterdam: Thesis Publishers.

White, Michelle J. (1988): Urban Commuting Journeys are Not “Wasteful”. *Journal of Political Economy*, 96(5): 1097–1110.

Zax, Jeffrey S. (1991): Compensation for Commutes in Labor and Housing Markets. *Journal of Urban Economics*, (30): 192–207.

# Recruitment Channels and Strategies for Employing the Low-skilled Workers in Europe<sup>1</sup>

Katalin Tardos – Jacob J. Pedersen

tardosk@socio.mta.hu; jacobbj@dps.aau.dk

**ABSTRACT:** This paper looks into the recruitment processes of a number of European companies and puts particular emphasis on the importance of the use of informal recruitment channels and of direct applications combined with waiting lists. Through this we seek to generate new knowledge about how European companies advertise vacant positions and recruit new employees, particularly low-skilled employees; and seek to understand whether the present recruitment strategies impact the employment of disadvantaged groups on the labour market. The research concludes that formal, informal and direct recruitment strategies need to be differentiated with regard to their impact on the employment of various groups of low-skilled facing difficulties on the labour market, and direct application seems to be a channel by which low-skilled jobseekers at a disadvantage on the labour market could achieve relatively balanced and promising results.

## Introduction

Designing an effective recruitment strategy for different groups of future employees is a major business challenge for employers. Employers need to balance several factors that might contradict each other, such as costs of the recruitment methods, acceptable vacancy duration, time spent on implementing the recruitment process, resources available inside the company for implementing recruitment, the quality of the generated pool of applicants, time and cost involved by the screening and selection process necessary after the recruitment. Employers engage in a complex, often implicit, optimisation process that to a great extent is based on previous experience related to recruitment channels and assumptions made in regard to the benefits and risks associated with the various channels.

<sup>1</sup> The research has been conducted in the framework of the Reconciling Work and Welfare in Europe (RECWOWE) EU Sixth Framework Programme, STRAND 1: Job protection, activation and employers' recruitment decisions project. Project participants included Christian Albrekt Larsen, Michel Berclaz, Giuliano Bonoli, Nevenka Černigoj Sadar, Karl Hinrichs, Mirosljub Ignjatović, Vera Messing, Jacob J. Pedersen, Valeria Sparano, Katalin Tardos, Patrik Vesan, A. Caroline Warfelmann, and Sabine Wichmann. English revision by Caroline Bodoczky.

At the same time as companies are looking for people to fill vacant positions, individuals are looking for jobs. The choices of employers play a key role in defining who gets employed and who stays unemployed. If job information is not equally accessible for all, this may have some important consequences at community level. It is possible that this will lead to structured social inequality in the sense that access to jobs is not equal (see Braddock–McPartland 1987; Neckermann–Kirchenmann 1991; Reskin–Ross 1990; Windolf 1986).

It is therefore important to understand the actions of the employers in relation to their choice of recruitment channels and their practices in combining the different search methods. To date relatively little is known about how different recruitment channels are bundled by companies to reach the desired optimum of balancing the above-mentioned factors and how these impact the employment of various groups of jobseekers, in particular groups with a disadvantage.

This paper looks into the recruitment processes of a number of European companies and puts particular emphasis on the importance of the use of informal recruitment channels (through current employees and other contacts), and the use of direct applications (generated by unsolicited inquiries made by jobseekers) combined with waiting lists. These are recruitment channels that are often overlooked even though information about the vacant position is not made freely available in the job market. Through this we seek to generate new knowledge about how European companies advertise vacant positions and recruit new employees, particularly low-skilled employees; and seek to understand whether the present recruitment strategies impact the employment of disadvantaged groups on the labour market. We aim to test the hypothesis that informal recruitment channels put problem group members at a disadvantage, which subsequently leads to lower levels of employment of such groups.

## Literature Review

There exist several theoretical and empirical standpoints on how job seekers find jobs. Research in employer's recruitment is far less developed (Behrenz 2001; Marsden–Gorman 2001). However the development in job search literature from one-sided partial search models (e. g., McKenna 1985) to equilibrium search models (that include both sides of the labour market – e. g., Burdett–Mortensen 1998) has resulted in an increased number of studies taking the recruitment actions of employers into consideration. However still recruitment studies outside the job search literature are still at a minimum and comparative research in the field of how companies recruit is still almost non-existent. Nevertheless, based on the available country specific studies, it is possible to generate some expectations in relation to why different companies recruit in different ways.

When trying to explain companies' choice of recruitment channels it is important to understand what these decisions are based upon. Viewed from the corporate side, depending on the current circumstances, there may well be a number of elements to take into account. Based upon existing literature the following three key factors help determine how a given company, in a given situation, chooses to search for new employees: 1. experience from previous use of different recruitment channels; 2. the economic climate and degree of sector unemployment; 3. what kind of labour the company needs.

The choice of search strategy will typically be based on experience from previous recruitment processes, for instance, local conditions may need to be taken into considerations (Russo et al. 1997). This will ensure that the company is adapting to the specific circumstances characteristic of the industry, geographic location, and the size of the company. At the same time this also depends on the conditions in the labour market. Russo et al. (1997) have shown that low unemployment leads to the use of many different search channels, while the opposite is true in a recession with high unemployment. Osberg (1993) underlines that the role of the employment services must be considered in conjunction with the economic climate and the current degree of unemployment. In a situation with a tight labour market, companies will thus be more likely to use recruitment methods that trigger a larger number of candidates than otherwise. This is supported by studies that suggest that firms' choice of recruitment channel partly depends on the local labour supply, and the type of labour that the company demands (Gorter-van Ommen 1999).

Also, important is the kind of position that is vacant. Barron et al. (1997), in a study of the optimal search strategy for employers, demonstrated that it is important to distinguish between positions where much training/education is needed, on the one hand, and those where this is not the case, on the other hand. Companies will, from a rational perspective search more intensively and extensively to fill positions when looking for workers with higher education or more experience than usual.

The essence of the above is that a number of factors affect how useful it is for the company to provide free access to information that a position is vacant. This is crystallized in the companies' choice of recruitment channel. In the literature on recruitment, three major approaches are identified for ways of finding a candidate for a vacant position, as the most common (Granovetter 1974):

**1. Formal methods of recruitment.** In this case the employer publicly makes the information on the job vacancy available through an intermediary service, like the public and private employment agencies, advertisements in newspapers, and journals or on the Internet.

**2. Informal methods of recruitment.** Using this approach to recruitment the employer will secure potential job candidates through recommendations from current employees or other contacts in the sector, or among personal networks. In this case, information on the job vacancy will remain restricted to the public.

**3. Direct application and waiting lists.** In this case the employer has a passive role, and is relying on the job search activity of jobseekers who submit an unsolicited résumé to the employer through a personal visit, by post or through the company website, thus creating the basis for a pool of candidates the employer can rely on should there be a job vacancy.

Formal channels have a larger audience and increase the chance of getting in contact with a high number of applicants. It also increases the chance of hiring a previously unemployed applicant. If the companies are considering this, their use of different channels will be affected by different circumstances, such as the economic climate, and country specific factors such as unemployment numbers. To some degree, formal channels are the best choice for companies willing to take risks; large companies might best afford this.

According to the literature, the use of informal recommendations give employers certain advantages: low costs for recruitment, screening, training and control as well as low turn-over (Rees 1966; Rees-Schultz 1970; Fernandez et al. 2000). A number of studies have however also stipulated that the use of informal channels results in better applicants (Granovetter 1995; Marsden-Campbell 1990; Simon-Warner 1992). Similarly, Gorter et al. (1996) shows that employers using formal channels need some time to assemble a pool of job applicants, while the process is quicker if they are found informally. In view of this, informal recruitment should play a role in most companies that have this option.

A number of studies have underlined the importance of having social ties: 40–50 percent of jobs found in the US (Granovetter 1995) – and 25–60% of jobs in different European countries (ISSP 2001) – are found through information from friends and relatives. A research conducted in Sweden (Behtuoi 2008) on the use of informal recruitment methods and the implication for immigrants showed that immigrants are less likely to be able to find their jobs through informal methods as the networks of the poor are less likely to include individuals who are employed.

Detailed single-firm studies, see below, have also shown that referred applicants had a higher probability of being hired than the non-referrals. Fernandez and Weinberg (1997) studied a bank, Fernandez et al. (2000) studied a phone centre, and Petersen et al. (2000) studied a high-technology firm. These are important findings and indicate that the way companies recruit should be investigated more closely. The current empirical basis is limited to very few companies – and European studies are even fewer – but recruitment processes seem able to define a social inequality in the job market.

Formal channels can be expected to reach a larger audience and generate more applications than informal channels. On the other hand informal channels are a way of generating a smaller portion of applications, as well as securing applicants who have some sort of informal knowledge or relationship to the company in advance. Both formal and informal recruitment channels are associated with advantages and disadvantages for the company (Mencken–Winfield 1998), making it rational for most companies to combine formal and informal recruitment channels since these complement each other.

## Research Methodology

This research is based on a sample of 41 companies located in 6 European countries, Denmark, Germany, Hungary, Italy, Slovenia, and Switzerland. Among the selected countries Switzerland, Denmark, Germany, and to a lower extent, Slovenia represent labour markets with employment rates above the European Union average for the 15–64 age group. On the contrary, the Hungarian and the Italian labour markets are characterised by low employment rates coupled with relatively high unemployment compared to the European Union average (Eurostat 2011). Considering the employment rate of low-skilled workers in the selected countries (pre-primary, primary & lower secondary education – ISCED levels 0–2) Denmark and Switzerland represent the higher end of the scale (62–69% in 2010); in Germany, Italy and Slovenia employment levels of the low-skilled are conducive to the European average (50–55%) while the low-skilled workers have the lowest chances to find employment in Hungary (38%) (Eurostat 2011).

In each country 7 companies (six in Slovenia) were selected from both the service and the industrial sector. Selected companies typically belonged to industries such as cleaning, food, retailing, construction, and manufacturing. A special criterion for the selection of companies was the high ratio of low-skilled workers among the workforce. On average, the investigated companies had a ratio of nearly 60 percent of low-skilled employees. Concerning the size of the companies, nearly a quarter were small businesses with less than 50 employees, less than half medium size organizations employing between 51 and 500 employees, and finally one-third of the companies in the sample employed more than 500 employees.



Data collection took place in 2009. Personal interviews were carried out in each organisation mainly with the Human Resources manager or, in the case of smaller companies, with the owner and CEO using a semi-structured questionnaire thus collecting qualitative data in addition to some quantitative figures.

## Dominant Recruitment Channels for the Low-skilled Across Europe

The research has revealed a surprising convergence of the most frequently used recruitment methods in all the six examined European countries. The two most frequently used methods in each country were *direct application or waiting list* and *recruiting through current employees*. There was only one exception to this rule, namely Switzerland, where the second most frequently used search method turned out to be the private employment agencies. 33 out of 41 companies (82 percent) used *direct application* and *waiting list* methods often or very often when seeking low-skilled applicants to job vacancies. A similarly high number of companies (30, 73 percent) used their *current employees* as sources to find new applicants.

A second group of recruitment channels was formed by those search methods that only a quarter to one-third of the companies used often or very often. *Private employment agencies* (14, 34 percent) and *public employment agencies* (11, 27 percent), and *other contacts in the sector* (11, 27 percent) belonged to the moderately used recruitment channels. Finally, a third category of rarely used recruitment methods was identified including search methods such as *Internet tools provided by non-public agencies* (7, 17 percent), *News-papers and other written media* (7, 17 percent), and *Internet tools provided by public employment agency* (5, 12 percent).

In the next section of the paper both frequently and moderately used Recruitment Channels will be discussed in detail to show the rationale of employers' preferences in their decisions to use each of the recruitment methods during their search process for new employees, particularly the low-skilled.

### Direct application and waiting list

Direct application can take the form of applying directly by phone, mail, e-mail, or on the company website, and sending an unsolicited CV and/or application form to the company. Most companies, large or small, receive a relatively high number of direct applications. "*We receive more than 2000 direct application a month at the national level. Most often people directly go to the local shops*" (CH5 RETAIL). It is important to note that direct applications are per se low cost solutions to find applicants, so companies can save considerably on choosing this form of recruitment. If direct applications are combined with building up a waiting list database, companies can procure a low-cost and relatively fast recruitment source for

themselves. An HR manager in Denmark explains the system of using unsolicited applications: *“When we get uninvited applications, we have received a lot of those lately, then we keep them for a couple of months, and if a position opens up within this period, we look at the applications we have to see if there is someone suitable, and then we call people in for a preliminary interview”* (DK2 RETAIL).

For many companies, a person handing in an unsolicited application is considered to be a positive signal for future motivation to work. In the case of low-skilled jobs, as qualification and experience is less meaningful, companies are usually looking for other ways to predict future performance. *“There are indeed many people who come to us in person and hand in an unsolicited application. With these people you recognize right away that they are looking for a position and really want to work”* (GE3 LAUNDRY).

Furthermore, there is clearly an increase of direct applications due to increased levels of unemployment on local labour markets. The Recruitment Officer of a large food chain in Hungary reports on the link between the dominant recruitment channels and the labour market as follows: *“Previously we used the newspaper frequently as a channel for recruitment. Now, due to the economic crisis more people take the initiative to find employment for themselves, and we do not have to find them as they contact the company. Until last year, this happened the other way round”* (HU4 FOOD).

## Recruitment through current employees

A major reason for choosing this informal method of recruitment is an underlying belief that referrals made by present employees will lead to candidates whom the employer can trust and who will perform their jobs to levels equivalent to employees making the recommendations. *“When we are looking for staff, we find someone relatively fast through word of mouth. We’ve had the best experiences this way”* (GE1 CLEAN). A major benefit, stated by several companies, linked to recruitment through current employees is the higher level of trust from the very beginning of the employment relationship. *“You cannot hire a complete stranger, only somebody by recommendation. The job itself requires this”* (HU2 SUPERMARKET). Recruiting through current employees is not necessarily an explicit recruitment policy, but in many cases it takes place spontaneously using the word of mouth of current employees and thus generating a necessary pool of applicants. The method of recruiting through current employees works even better, if employees of the company are satisfied with their job, so an indirect benefit of investing into some area of human recourse policy can generate substantial cost saving for the company in terms of recruitment costs. *“We can count on an important number of potential workers because our own employees recommend them a lot, because they are satisfied by the working conditions. In relation and in addition we have of*

*lots of direct application*" (CH2 CLEAN). Some companies formalise their informal recruitment policy to an extent that they match a financial system to go with it. *"If you recommend a friend or an acquaintance, and they stay here for longer than the initial trial period [3 months], we reward the employee [€ 400]"* (DK7 CALL). Interestingly, recruiting through current employees is not only used by small firms with a less formalised HR policies in general, but medium-size and large firms with more than 500 also utilise this informal recruitment method extensively, too.

## Private employment agencies

Many HR managers believe that private employment agencies are efficient in pre-selecting candidates and that they represent a faster and more flexible approach compared to public employment offices. Private agencies are generally seen as a good source for getting appropriate workforce for their needs being highly qualified professionals for job matching: *"ADDECO and others they do their work professionally. Their goal is to achieve the profit. Therefore they send to employers the best workers so the employers would be satisfied with them"* (SL4 HOUSEHOLD-APP). According to the data, larger companies are the ones that typically use private employment agencies. Temporary agency contracts are often used by firms as a longer probation period when recruiting workers. It usually happens, in fact, that companies firstly employ workers on temporary agency contracts, especially in case of recruiting from low-skilled groups, and after a longer period of time they in-source the best workers. *"My strategical idea is to first lease employees, and use the "try and hire" approach, meaning that after half a year or a year we are providing possibilities for employment for the best leased workers"* (HU6 ELECTRONICS). Many companies are using leased employees as a pool of tested employees from which they can hire whenever they need to *"this allows us to evaluate the workers and when we need to employ new workers we usually choose these people, that is people we have already known and evaluated"* (IT1 BEVERAGES). Another benefit mentioned in terms of using private employment agencies is the flexibility and ease with which to terminate the contracts.

Several companies have indicated that they have stopped using private employment agencies in the last two years. The rationale for suspending or terminating the usage of private employment agencies was different in each case, either the service was considered too expensive and inefficient (HU1 CLEANING), or it was used for the initial mass recruitment of large number of employees (HU7 INDUSTRIAL), or used as an easy way to lay off employees (HU6 ELECTRONICS), and finally, in the fourth case, problems of motivation occurred with the leased employees. *"We had an initiative to lease employees through the HumanRent Holding. In my opinion, it raises a problem for the employee, as it is not clear for the employee who the employer is. I work in one place, but all the administrative tasks are conducted*

*by the other organisation. You hate your workplace, but which one? The one that administers the pay, or the one where I work? A leased employee cannot be loyal to the company as he/she doesn't know whom to be loyal to" (HU4 FOOD).*

As discussed above, using the private employment agencies may considerably decrease the risks of employing low-skilled workers. As a matter of fact, half of the interviewed companies were, in principle, in favour of using private employment agencies as a means to increase the employment chances of unemployed groups.

## Public employment agencies

According to the research data, European companies are unanimously dissatisfied with the services of the public employment agencies. Less than one-third of companies use public employment services often or very often to recruit the unskilled, but being a user of its services, or reporting job vacancies to the public employment agency does not necessarily indicate satisfied clients of the institution. Employers' opinion ranges from advocacy to the complete close down the employment agency at one end of the scale, to considering the cooperation with the public agency as an act of social responsibility, at the other. A small Hungarian company presents the most negative view: *"The problem with this is that if we need a new employee, we need it immediately. The reason why I do not want to maintain a relationship with a state employee is that I cannot explain the weight of my problem, they do not have such empathy to understand why this is so important for me. I just do not believe in the system. (...) In my opinion the public employment agency should be closed completely"* (HU5 CONSTRUCTION). The inefficiency of the public employment agency is raised by other companies as well. *"This is a high amount of work. And not much comes out of it. Therefore, I'm always weary of the recommendations from the public employment agency because the expenses are so high"* (GE3 LAUNDRY). On the other end of the scale we find a few companies who raise the issue of their social responsibility. *"We want to expand further, we want to be a socially responsible organisation, especially due to the size of the organisation, and we have had a few annoying incidents, where you've tried some different schemes, with annoying negative outcomes, but we have also had positive stories of course"* (DK4 MACHINE).

*The companies elicited several reasons for why they were not using the public employment agency more frequently. The first reason mentioned was the agency's low performance in job matching. "I consider it pretty important that the employment agency employs 'field representatives' and that they travel around, participate in conversations, check out the firms, talk with the companies and ask, 'what do you guys actually need?' And then promptly send the right person for the right job"* (GE4 SUPERLARGE). Secondly, employers have previous negative experience with the unemployed sent by the public employment agency, and thus assume that the un-

employed are not motivated to work. Being sent by the public employment agency is often considered a negative signal. *"(...) I also fear that they will give me the one that they want to get rid of the most, if I contact the employment agency"* (DK3 CAKE). *The third problem mentioned in relation to the public agency was its too bureaucratic nature. "We have tried the public employment agency, but it was very complicated. It was too bureaucratic. There were a lot of requirements, administration, and paper work. This was 2-3 years ago, when I wanted to arrange for some kind of state subsidy for employing unemployed people, but I decided not to do so"* (HU3 Security). The fourth problem mentioned was that the level of service depends on the person, and the service of the agency is not standardised sufficiently. Another factor considered to be deficient was that the agency does not try to understand the needs of the employer. Employers raised concerns about the amount of information required by the agency, stating that the agency wanted too much information on employers. Another employer mentioned its dissatisfaction with the fact that they could not keep their own format for the job ads, so that all ads look uniform to the unemployed. Finally, employers complained of various aspects of inefficiency of the services offered by the public employment agency. Some employers found it too slow, others mentioned that too many applicants were sent for a single vacancy, while others just concluded that it is not worth the effort.

The above-mentioned characteristics explain why the majority of the companies are reluctant to use the services and employ the referees of the public employment agency despite the low-cost nature of this recruitment channel and why the more expensive private employment agencies are chosen more frequently even in the case of the unskilled workforce. Among the companies using the public employment agency we only find medium-size and large companies.

## Using other contacts in the sector

Relying on other contacts in the sector is a moderately frequent recruitment channel mostly used by small and medium size firms. This recruitment channel was found to be more extensively used in Switzerland and Denmark. A major advantage of this source of recruitment is the likelihood of the person making the referral to understand the nature of the job and thus the needs of the employer. This recruitment channel can usually provide candidates for the job in a short period of time and at low cost. *"A channel that is commonly used in order to recruit workers is represented by recommendations and suggestions on the part of other companies. When we had to recruit new workers we usually asked other firms to recommend us someone and we often received calls from other firms, which were considering employing a worker previously employed in our firm, asking us if she/he was a valuable and reliable person"* (IT5 ELECTRONICS).

To summarise the pros and cons of each of the recruitment channels we have prepared a table indicating their major characteristics in the case of searching for low-skilled workers in the context of high labour supply in the labour market as it is the case at present in Europe (see Table 1). The summary table indicates that the highest number of positive characteristics can be associated with the two most frequently used recruitment channels, namely, waiting list and direct application, as well as recruiting through current employees. The importance of the various factors would be different for highly qualified jobs in contrast to low-skilled ones. In the case of low-skilled workers, the factors with the highest rank tend to be speed, low costs, positive signals on future motivation and performance, and having a smaller pool of applicants in order to minimise the screening time and effort needed to find the person who best fits the requirements of the job.

Table 1: Summary Table on Recruitment Channels

	Waiting list or direct application	Current Employees	Private employment agencies	Public employment agencies	Other contacts in the sector	Internet tools provided by non-public agencies	News-papers and other written media	Internet tools provided by public employment agency
Speed	Fast (+)	Fast (+)	Fast (+)	Slow (-)	Fast (+)	Fast (+)	Slow (-)	Slow (-)
Cost	Low (+)	Low (+)	High (-)	Low (+)	Low (+)	High (-)	High (-)	Low (+)
Signal	Positive (+)	Positive (+)	Neutral	Negative (-)	Positive (+)	Neutral	Neutral	Negative (-)
Pool of applicants	Small (+)	Small (+)	Large (-)	Large (-)	Small	Large (-)	Large (-)	Moderate
Amount of screening needed	Low (+)	Low (+)	Low (+)	High (-)	Low (+)	High (-)	High (-)	High (-)
Availability without extra resources	High (+)	High (+)	Moderate	High (+)	Moderate	Moderate	Moderate	High(+)
Likelihood of having a good person/job match	High (+)	High (+)	High (+)	Low (-)	High(+)	Low (-)	Low (-)	Low (-)
Number of positive features	7	7	3	2	4	1	0	1

## The Strategies to Bundle Recruitment Channels

In the large majority of cases companies do not use single recruitment channels, but flexibly bundle several recruitment methods. On average, companies use approximately 3 recruitment methods often or very often. Nevertheless, companies have experience with a higher number of recruitment channels, which they might use in special circumstances on a rare basis. Considering the total array of

recruitment methods companies use at least 5 on average. Thus, most companies rely on more than one method, but little is known on how they combine the various methods.

First, let's consider the three major approaches to recruitment (direct, formal and informal methods). Our research results highlight that overall formal recruitment channels are less frequently used (24, 60 percent) than direct application (33, 82 percent) and informal recruitment channels (31, 77,5 percent) in case of recruiting the low-skilled labour. Hence, our original assumption that in a majority of cases job vacancies would not be officially made public to jobseekers was reinforced by the research data. Large companies employing more than 500 employees more typically use formal channels of recruitment. Also in Germany and Switzerland, formal channels tend to be more frequently used.

In order to reveal more on the strategies to combine specific recruitment methods we have prepared a typology of recruitment strategies that focuses on combinations of formal versus non-formal recruitment methods used often or very often. There are only very few companies that rely uniquely on one of the types of recruitment approaches (7, 17 percent) (see Table 2). Among those companies relying on only one type of approach, it is either the direct application (3, 7,5 percent) or the formal recruitment methods (4, 10 percent). *It is important to stress that in the sample no companies used the informal channels independently from other recruitment approaches.* A large group of the companies combine at least two of the recruitment approaches, most typically *direct application and informal channels* (13, 33 percent). We named this type of recruitment strategy "*Simple & non-formal*". However, "*Complex & balanced*" recruitment strategies characterise more than one-third of the companies (15, 37,5 percent) that combine all three types of recruitment approaches – direct, informal, and formal – in their recruitment strategy.

Table 2: Typology – Direct & Informal & Formal Recruitment Channels often or very often used

	Frequency	Percent
Only direct application	3	7,5
Only Informal	0	0,0
Only Formal	4	10,0
Direct & Informal	13	32,5
Direct & Formal	2	5,0
Informal & Formal	3	7,5
Direct & Informal & Formal	15	37,5
<b>Total</b>	<b>40</b>	<b>100,0</b>

## Do Recruitment Channels Have an Impact on Employing Problem Groups?

A minority of employers (one-third) acknowledged that the segment of the labour market they were functioning in was to a great extent composed of so called problem groups, and did not consider group membership to particular socially disadvantaged groups as a special risk factor. *“People who don’t have any ‘problems’ don’t want to work in a laundry”* (GE3 LAUNDRY). Nevertheless, two-thirds of the companies have identified problem groups whose employment represents risks for the employer. The semi-structured interview explicitly inquired about three typical problem groups, *unemployed above 50 years of age, unemployed with ethnic minority background, as well as long-term unemployed*. From these three groups long-term unemployment seems to be the strongest signal for employers in terms of risks for future employment (42 percent), whereas only a quarter of the companies considered unemployed of ethnic minority or immigrant background and aged above 50 years to be a risk. In addition, companies also mentioned other problem groups. Young people, people with disabilities, drug addicts, people with a criminal record were among the groups mentioned by the employers.

To summarize the action of employers in relation to problem groups, we collected data on the ratio of companies hiring from the three most affected problem groups along with the average number of employees hired from the given problem group. The trend is consistent insofar as the group that was considered the most frequently as risky (the long-term unemployed) is the least frequently employed by companies. Approximately half of the companies are ready to employ a person with a record of long-term unemployment whereas 80 percent of the companies reported employing from the group of unemployed of ethnic minority background or above 50 years of age. Furthermore, the average number of long-term unemployed hired in the last two years was also significantly lower (approximately 3) than for unemployed with ethnic minority background or of mature age (8 on average). Nevertheless, in absolute terms, the numbers are very low. On average, companies have employed 16 employees from the three largest problem groups, representing 1,5 percent of the average labour force (1122 employees).

As a matter of fact, in relation to the perception of problem groups, the research indicates a relatively high variety among countries. In Slovenia, Hungary and Denmark companies mention a higher number of problem groups on average (2,14–3,33 groups), while in Germany, Italy and Switzerland companies typically mention less than one group on average (0,43–0,71). However, according to the data, not labelling explicitly as risky the employment of a given disadvantageous group on the labour market does not necessarily lead to higher employment ratios, conversely, explicitly mentioning risks of employment does not necessarily indicate lower employment rates of problem groups. The data indicates a gap between the actual discourse and employment practices in both directions.



Comparing the research data with the European labour market statistics, we can differentiate among the countries where the average number of problem groups mentioned is congruent with the evidence of the labour market statistics and those where it is not. Statistics exist for the employment rate of the mature age population (55–64 age group) and long-term unemployment rates. Italy, Hungary and Slovenia have an especially low employment rate (34–36%) for the 55–64 age group, whereas Denmark, Switzerland and Germany have considerably higher employment rates for the same age group (57–68%) (Eurostat 2011). With regards to the unemployment rate of the long-term unemployed, Denmark and Slovenia have the best performance, while Switzerland and Germany join the negative trends of Italy and Hungary in this respect. Comparative statistics for employment levels of immigrants and ethnic minorities, especially the Roma in Hungary and Slovenia could not be found, but research evidence revealed that mostly the Roma origin was considered explicitly to be a risk. For Switzerland and Germany, countries with a typically high employment rate both in general and for the low-skilled, mentioning less than one problem group in average, represents congruent practices. Both in Switzerland and Germany employment of the long-term unemployed caused a problem. At the other end of the scale, for Hungary mentioning higher number of problem groups on average is congruent with the overall low levels of employment, in particular of the low-skilled, the 55–64 age group, the Roma and the long-term unemployed. Slovenia represents a mixed case, as compared to the European Union average it has good labour market indicators, however both mature age and Roma origin lead to lower employment levels. Denmark and Italy represent incongruent cases as in Denmark all available labour market indicators are among the best in Europe (general employment rate, employment rate of the low-skilled, mature aged, rate of long-term unemployed), nevertheless employers explicitly mentioned a higher number of problem groups on average. Italy represents a case of inconsistency in the opposite direction. The Italian labour market indicators (general employment rate, employment rate of the low-skilled, mature aged, rate of long-term unemployed) were below the European average in 2009; nevertheless, employers did not articulate their reservations towards the above mentioned problem groups.

In the following section we shall focus on the relationship between the recruitment channels and the employment of the problem groups. For this purpose we shall use two indicators, namely the average number of employees hired from problem groups in the last two years, and the average number of hires from problem groups in the last two years per 100 employees. Firstly, we shall investigate whether the average number of hires from problem groups differ in the case of the frequent usage of given recruitment channels. Secondly, we shall take a closer look at whether the employment chance of members of problem groups on the labour

market relative to the size of companies differ in case of recruitment channels. The first indicator focuses on the absolute numbers, while the second indicator assesses differences in relative terms.

Considering individual recruitment channels, we do find some differences among companies with regard to the average number of employees that they have hired from problem groups. Interestingly, companies using the typical formal recruitment channels as Internet tools, public and private employment agencies, advertisements in newspapers tend to employ higher number of employees from problem groups, on average. On the other hand, companies using informal recruitment methods and direct application, on average, employ a lower number of employees from problem groups. Hence, in absolute terms, employment through formal recruitment channels seem to produce a higher number of hires from problem groups. These results correlate with the literature that emphasizes that the unemployed are more frequently recruited through formal channels of recruitment and the usage of informal recruitment methods tend to exclude those who have weaker social ties and networks, as problem groups among the unemployed typically do. However, if we consider the indicator of average number of hires from problem groups in the last two years/100 employees, we find that companies using informal or direct recruitment channels perform better in relative terms in employing from the problem groups. This is especially true for using direct application and waiting lists and contacts in the sector (see Table 3).

Table 3: Average number of employees hired from problem groups and average number of employees hired from problem groups/100 employees by companies using the different recruitment channels often or very often.

	Average number of hires from problem groups in the last two years	Average number of employees hired from problem groups / 100 employees	Number of companies
Internet tools provided by non-public agencies	44,0	3,9	N=7
News-papers and other written media	25,5	6, 2	N=7
Public employment agencies	25,1	7,4	N=11
Internet tools provided by public employment agency	24,6	10,4	N=5
Private employment agencies	19,3	3,74	N=14
Waiting list or direct application	16,2	10,47	N=33
Current employees	15,5	8,5	N=30
Other contacts in the sector	8, 5	12,5	N=11
<b>Total</b>	<b>15,9</b>	<b>9,7</b>	<b>100% (N=41)</b>

To summarise this section, due to the low case number in the sample, it is very difficult to evaluate whether there is a clear-cut relationship of the recruitment channels of firms to the chances of problem group members gaining employment. We can surely state that the question needs to be tested on a larger sample. Never-

theless, the data does show some differences among average number of hires from problem groups across the frequently used recruitment channels, but it is difficult to give a precise indication of its degree. Our explorative research indicates that the impact of recruitment channels on the employment of problem groups should be tested both in absolute and relative terms and controlled for other variables, too. According to our assumption, other factors, like the prejudices of the recruiter, the inclusiveness of the company, previous experience with problem groups, methods of selection, to name some of them, will influence the final outcome of whether problem group members will be employed with equal chances or not.

## Conclusions

In the first part of the paper we investigated the dominant recruitment channels used by European companies to find job applicants for low-skill jobs, and concluded that direct and informal recruitment channels were the most frequently used across the countries with a surprising convergence. We analysed the positive and negative factors for employers related to each recruitment channel and concluded that waiting list and direct application along with recruiting through current employees represented the largest benefits for employers in the case of searching for low-skilled employees, in terms of low cost, speed, and positive signals. Thus employers follow a rational choice in the implicit optimisation process of choosing recruitment channels.

We have also investigated the bundles of recruitment channels used by companies, and found that two major strategies co-exist: a “Complex & balanced” recruitment strategy used by more than one-third of the companies and a “Simple & non-formal” recruitment strategy performed by closely one-third of the interviewed companies. In view of this, we need to differentiate our initial assessment of the dominant role of non-formal recruitment channels. Nevertheless, a large segment of the labour market is operated through asymmetrical information on job vacancies. Hence, we aimed to test the hypothesis that informal recruitment strategies put problem group members at a disadvantage, which subsequently lead to lower levels of employment of such groups. From the three groups, long-term unemployed, unemployed with ethnic minority background and unemployed above 50 years, the long-term unemployed were considered to be the most risky by employers and, consistent with the employers’ perception, members of this group were employed in the lowest numbers and by the fewest companies.

Related to the relationship of recruitment channels and employment of problem groups our results were mixed. In absolute terms, employment through formal recruitment channels produced a higher number of hires from problem groups, on average. On the other hand, when considering the indicator of the average number of hires from problem groups in the last two years/100 employees we found that

companies using informal or direct recruitment channels were performing better in relative terms in employing from the problem groups. Thus the hypothesis that informal strategies put problem group members at a disadvantage was not unanimously reinforced, at this stage. Interestingly, the case of *direct application and waiting lists* somewhat stands apart, as results showed both in absolute and relative terms an average or above average chance of employment for members of problem groups through this channel. Hence, we hereby conclude that formal, informal and direct recruitment strategies need to be differentiated with regard to their impact on the employment of various groups of low-skilled facing difficulties on the labour market, and direct application seems to be a channel by which low-skilled jobseekers at a disadvantage on the labour market could achieve relatively balanced and promising results. Finally, we need to emphasise that the relationship between recruitment channels and strategies and employment of problem groups must be further tested on larger samples.

## References

- Barron, John M. – Mark C. Berger – Dan A. Black (1997): *On-The-Job-Training*. Michigan: Upjohn Institute for Employment Research.
- Behrenz, Lars (2001): Who Gets the Jobs and Why? An Explorative Study of Employers' Recruitment Behavior. *Journal of Applied Economics*, (4)2: 255–278.
- Behtoui, Alireza (2008): Informal Recruitment Methods and Disadvantages of Immigrants in the Swedish Labour Market. *Journal of Ethnic and Migration Studies*, (34)3: 411 – 430.
- Braddock, Jomills H. – James M. McPartland (1987): How Minorities Continue to be Excluded from Equal Employment Opportunities: Research on Labor Market and Institutional Barriers. *Journal of Social Issues*, (43)1: 5–39.
- Eurostat (2011): Employment statistics. [http://epp.eurostat.ec.europa.eu/statistics\\_explained/index.php?title=File:Employment\\_rate,\\_age\\_group\\_15-64,\\_2010\\_\(%25\).png&filetimestamp=20111117142647](http://epp.eurostat.ec.europa.eu/statistics_explained/index.php?title=File:Employment_rate,_age_group_15-64,_2010_(%25).png&filetimestamp=20111117142647). [Accessed on November 18, 2011.]
- Fernandez, Roberto M. – Nancy Weinberg, (1997): “Sifting and Sorting: Personal Contacts and Hiring in a Retail Bank”. *American Sociological Review*, 62: 883–902.
- Fernandez, Roberto M. – Emilio Castilla – Paul Moore (2000): Social Capital at Work: Networks and Hiring at a Phone Center. *American Journal of Sociology*, (105)5: 1288–1356.
- Gorter, Cees – Peter Nijkamp – Piet Rietveld (1996): Employers' Recruitment Behaviour and Vacancy Duration – An Empirical Analysis for the Dutch Labour Market. *Applied Economics*, 28: 1463–1474.
- Gorter, Cees – Jos N. Van Ommeren (1999): Sequencing, Timing and Filling Rates of

- Recruitment Channels. *Applied Economics*, (31)10: 1149–1161.
- Granovetter, Mark (1974): *Getting a Job. A Study of Job Contacts and Careers*. Cambridge: Harvard University Press.
- Granovetter, Mark (1995): Afterword, 1994 – Reconsiderations and a New Agenda, Mark Granovetter: *Getting a Job: A Study of Contacts and Careers*, 2. udgave, USA: University of Chicago Press, 139–182.
- Lindeboom, Maarten – Jan van Ours – Gusta Renes (1994): Matching Employers and Workers: An Empirical Analysis on the Effectiveness of Search. *Oxford Economic Papers*, 46: 45–67.
- Marsden, Peter V. – Karen E. Campbell (1990): Recruitment and Selection Processes: The Organizational Side of Job Searches. Ronald L. Breiger (red.): *Social Mobility and Social Structure*. Cambridge: Cambridge University Press.
- Marsden, Peter V. – Elizabeth H. Gorman (2001): Social Networks, Job Changes and Recruitment. Ivar Berg. – Arne L. Kalleberg (red.): *Sourcebook on Labor Markets: Evolving Structures and Processes*. New York: Kluwer Academic/Plenum Publishers.
- McKenna, Christopher J. (1985): *Uncertainty and the Labour Market: Recent Developments in Job-search Theory*. Harvester Press.
- Mencken, F. Carson – Idee Winfield (1998): In Search of the “Right Stuff.” – The Advantages and Disadvantages of Informal and Formal Recruiting Practices in External Labor Markets. *American Journal of Economics and Sociology*, (57)2: 135–153.
- Neckerman, Kathryn M. – Joleen Kirschenman (1991): Hiring Strategies, Racial Bias, and Inner-City Workers. *Social Problems*, (38)4: 433–447.
- Osberg, Lars (1993): Fishing In Different Pools: Job Search Strategies and Job-Finding Success in Canada in the Early 1980s. *Journal of Labor Economics*, (11)2: 348–386.
- Petersen, Trond – Ishak Saporta – Marc-David Seidel (2000): “Offering a Job: Meritocracy and Social Networks”. *American Journal of Sociology*, 106: 763–816.
- Rees, Albert – George P. Shultz (1970): *Workers and Wages in an Urban Labor Market*, Chicago, University of Chicago Press.
- Rees, Albert (1966): Information Networks in Labor Markets. *American Economic Review*, (56)2: 559–566.
- Reskin, Barbara F. – Patricia A. Ross (1990): *Job Queues, Gender Queues: Explaining Women’s Inroads into Male Occupations*. Philadelphia: Temple University Press.
- Russo, Giovanni – Cees Gorter – Ronald Schettkat (1997): *Searching, Ranking and Hiring*. Free University Amsterdam, Faculty of Economics.
- Simon, Curtis J. – John T. Warner (1992): Matchmaker, Matchmaker: The Effect of Old Boy Networks on Job Match Quality, Earnings, and Tenure. *Journal of Labor Economics*, (10)3: 306–330.
- Windolf, Paul (1986): Recruitment, Selection, and Internal Labour Markets in Britain and Germany. *Organizational Studies*, (7)3: 235–254.

# The Prominent Csángó Hub

## Key Actors in the Network Structure of the Csángó Elite

Eliza BODOR-ERANUS and Hanna KÓNYA

eliza.eranus@uni-corvinus.hu; hanna.konya@uni-corvinus.hu

**ABSTRACT:** The main question this article addresses is what characteristics should actors have to be recognized as prominent figures by the Csángó minority in the collaborating network of “Csángó matter”. To answer the question, we have analyzed the relations between actors – individuals and organizations – of this ethnic minority. The results show that a small and prominent hub of the Csángó elite has a remarkable role within the collaboration network of the “Csángó matter”, and their network position correlates with the interlock organizational membership as well.<sup>1</sup>

“It should be noted that to say we are studying social structures is not exactly the same thing as saying that we study social relation.”

A. R. Radcliffe-Brown, 1940

## Introduction

The Csángós are an ethnic minority in the eastern part of the Carpathians in Romania. There seems to be more or less agreement in today’s scientific literature that Catholics from Moldova are of Hungarian origin, but are strongly assimilated into Romanian culture. At least three features need to be considered when speaking about the identity of the Csángó minority from Moldova: geographic location, religion and language, and self-identification. There are about 250.000 Catholics in Moldova, out of which approximately 60.000 speak Hungarian.

Scientific interest toward Hungarians from Moldova arose in the 18<sup>th</sup> century and for more than a century it only focused on their origin (Benda 1989:12). When trying to present an overview of the Csángó origin based on scientific literature, it is inevitable to speak very briefly about the history of Romanians and Hungarians from Moldova. According to Benda (1989) Romanians have been known of in

<sup>1</sup> This research is part of 4.2.1/B-09/VKMR-2010-0005. TÁMOP program, under the subproject, “*Hatékony állam, szakértő közigazgatás, regionális fejlesztések a versenyképes társadalomért*”, and workshop, “*Társadalmi és kulturális erőforrások, fejlesztéspolitikák, helyi fejlődés*”. The research coordinator is Zoltán Szántó. Eliza Bodor-Eranus wishes to thanks for the support of the Magyar Állami Eötvös Fund. Hanna Kónya wishes to thanks for the support in her research of the Ferenc Deák Fund. We thank to Jose Luis Molina and László Letenyei, for helping our work with precious professional observations.

Moldova since 1164, in the 12<sup>th</sup> century “a group of people called Vlach” were the majority in Moldova. The first certified mention of Hungarians in Moldova is dated to the 13<sup>th</sup> century i.e. 1225, when Hungarian King Endre II established a military presence of Hungarian mercenaries and their families. By the last quarter of the 14<sup>th</sup> century Petru Schiopu estimated that there were 20.000 Hungarians in Moldova. At the very beginning of the following century their number rapidly declined due to the Tartar ravages, even though the Hungarians populated more than a dozen of towns and 16 villages. In the middle of the 17<sup>th</sup> century there were about 12000 Hungarians for whom the ruler of Moldova requested a religious leader from Rome. Survivors of the 1764 Siculicidium<sup>2</sup> fled to Moldova and Bucovina. At the beginning of the 19<sup>th</sup> century the Austrian Consul from Jászvásár, reported to Vienna about 21.307 people; twenty years later their number was more than double. In the 1905 yearbook of the Romanian Academy, Radu Rosetti estimated that there were about 50–60.000 Hungarians in Moldova.

However, there is considerable debate in the historical literature concerning the identity of Csángós (Benda 1989; Baker 1997; Mikecs 1989; Domokos 1940, 1987; Cosa 2007; Mărtinaş 1985). While some researchers are convinced about the Hungarian origin of Csángós, mainly due to cultural aspects, others emphasize their assimilation into the society of the Romanian majority, or even acclaim their Romanian origin.

It is due to these investigations, as well as to a wide number of ethnographic researches, that we have an extensive knowledge of the minority's origin, historical background, cultural heritage, language - especially their characteristic dialect -, identity, and the institutions that play an important role in their life: church, school, etc. (e. g. Arens–Bein 2004; Kinda–Pozsony 2005; Pozsony 1996, 2003, 2005; Peti 2006; Tánczos 1996, 1997; Halász 2002).

As far as the migration of the Csángó population is concerned, two studies need to be mentioned: the one conducted by Larissa Adler-Lomnitz among Csángó guestworkers living in Budapest, focusing on their information channels (Adler-Lomnitz–Gonzalez 2007) and the empirical study by Mohácsék–Vitos (2005) on the consumption patterns of migrants from one particular Csángó village, differentiating between productive and unproductive personal consumptions.

In spite of these very informative investigations, very little is known about the structural characteristics of this society, the people who play important roles in the preservation and the transfer of the Csángó identity and culture. Sociological literature is very limited regarding the Csángó minority, especially the characteristics of its social network.

The 1989 regime change in Romania can be considered as a historical fact which favored Csángó's self identification, when members of this community mi-

2 Also called the Massacre from Mádéfalva (mádéfalvi veszedelem), when under the order of Maria Theresa, about 400 Széklers were massacred. The reason was, that they were not obeying the order of the Habsburgs, and they resisted the forced military draft and organized a revolt.

grated abroad either to work or to study at Hungarian universities. Some of these people settled in Hungary others returned to Moldova, and in many cases they identify themselves with the cause they call the “Csángó matter”. The “Csángó matter” includes a vast array of events ranging from the international Csángó ball to various debate forums: cultural events, exhibitions and publications. All these activities are claimed to serve the purpose of preserving the Csángó-Hungarian society, its language and cultural heritage. Several formal or informal groups and organizations are active in the Csángó matter in Romania and Hungary as well.

Neglecting outsiders, or people who have a certain level of involvement, but who are not really actively involved, the most active groups in considering the Csángó matter are as follows: The first group needed to be mentioned is the formal organization representing Csángó Hungarians, the MCSMSZ, (Association of Csángó-Hungarians in Moldova) which first of all focuses on the Hungarian education of the young generation. The second group is the KEMCSE, geographically situated in Hungary, (the Association of Godparents for Csángó-Hungarians from Moldova) a support institution for the large educational program. Nearby there is a branch of MCSMSZ called Pusztinai Házért Egyesület, having even common members with it. Another social group is the Szeret-Klézse Alapítvány also involved in Hungarian education and extra-curricular activities in Klézse. Beside KEMCSE other organizations such as the AMMOA (Foundation for the Hungarian education from Moldova) and the Foundation of Pál Péter Domokos are also sustaining the economic aspects of the educational programmes We can find the social group consisting of those young Csángós who settled in Hungary (mostly in Budapest) upon graduation, keeping in regular contact with Moldova as well: the Csángók a Csángókért Baráti Kör. Several members of this later group are involved in the publication of the only Csángó journal, the Csángó Tükör, which is entirely written and edited by Csángós. An important event that brings together people and also sponsors interested in the matter is the Csángó Ball, organized yearly in Budapest, bringing together artists, musicians, social scientists and everyone supporting the preservation of this culture. Again there are people involved in the organization of the Village days in Magyarfalu and the Village Week of Somoska (both held in Moldova). One important issue of the Csángó matter – the most argued as well – is the introduction of the Hungarian language into religious life. While many support the idea of having Hungarian masses, there are just as many if not more who are in opposition, arguing that people would not understand it. Two important actors are the two priests who were delegated by the Episcopate of Iasi to Budapest in order to pave the way for introducing Hungarian masses in Moldova. Priests always had and have an important role in Csángó communities, yet as several interviewees recall, they often see things differently and are not necessarily supporting the so called Csángó matter. Another important group or organization is the Dumitru Mărtinaş Association that represents the idea that Csángós are of Romanian ori-



gin. Foundations like the Pro Minoritate and the Teleki Foundations are also very active in the Csángó matter, organizing festivals, conferences, publications that focus mostly on cultural aspects.

It can be concluded from the broad number of organizations and events involved, that lots of different people contribute to the preservation of the Csángó culture and language, yet there is no example in the literature about a thorough investigation concerning the social network that maintains and operates the Csángó matter. The aim of this article is to analyze the people engaged in activities related to the Csángó matter from the point of view of their social network. In order to meet this goal we study both the relationship among people and among different organizations involved.

During the first (exploratory) stage of the research, actors were selected through 'snowballing', namely a nonprobability sampling technique where existing study subjects recruit future subjects from among their acquaintances. Thus the sample group appears to grow like a rolling snowball; further interview subject were identified through Csángós living in Budapest and in Moldova. During this research we use 'Csángó elite' as a term, for those who were named and identified as elite through the snowball sample. Elite definitions and expected characteristics were also collected from all interview subjects.

Our hypothesis in this article is that the network properties of opinion leaders and brokers are in strong relation with the organizations and events they are enrolled in, so that those people will be the most visible in the Csángó matter, who are involved in the most organizations, thus providing the social structure of the "Csángó matter". As such we suppose that there will be a couple of individuals who are active on several fronts and will be seen as key figures – named by many individuals as collaborating with – rather than those who are concerned on only one aspect even if they are working on that everyday. To test our hypothesis, we first examine the relation between the most active members of the Csángó elite – a group defined with snowball sample through group members, – then among the most active organizations, and finally, we try to combine these two types of relations, presenting the relation between people and social groups that constitute the social structure of the Csángó matter. Our results show that few members of the elite with significant influence – the opinion leaders (see Katz-Lazarsfeld 1955) – also have an important role through organizations.

The next section includes the literature review, then we present the methodology and data, followed by the discussion of the results and conclusion.

## Theoretical background

Similarly to Szántó-Tóth (1993), in our approach micro level includes the relations among individuals and it is based on the theories about social capital, embeddedness, opinion leader and brokerage. Meso level is constituted by the social network theory frame and the social network of individuals and organizations themselves, while Macro level includes the relation between social groups within the larger society and the relationship between the two states involved.

The relation between social interactions and social structures is firstly associated with Radcliffe-Brown (1940) and the Manchester school, who defined societies as the system of social relationships, emphasizing that observing events in a specific social group cannot clarify the social structure when the society crosses group boundaries. In order to reveal the actual system functions we should concentrate on people and their interactions.

At Manchester University in the 1940s and 1950s Barnes (1949), Mitchell (1969) and Bott (1955) developed theories based on the structural models of social relations, interpreted the social structure as a network of relations and reloaded the network analysis methods with real sociological content. They collected data focusing on the relations of a central individual, instead of concentrating on the whole network i.e. their primary interests were one's personal relationships. The complete network of a society is built on ego-centered networks, a concept suggested by Mitchell who discovered how interactions and relations could be analyzed from micro to the macro level This influence can be observed in further literature (as it is described by Némedi 2008), especially from the Harvard school, e. g. Mayo (1945), who focused on the social problems of the industrial civilization and Warner-Lunt (1941, 1942), who researched the principles of community organizations.

Previously, Jacob Moreno contributed to the development of the field with the idea of representing social interactions with dots and arrows: the sociogram (1937). Moreno's starting point was the informal networks in educational institutions. He presented these relations in graphs and this has become the basis of social network analysis applied in social sciences. (In other sciences like mathematics, graph theory appeared sooner).

Relevant literature also takes into consideration that people can be engaged in more social structures or organizations, for example Scott (1997) who investigates corporate power and the formation of the business elite, and in Hungary Gerő-Vedres (1997), using block model analysis to determine the relationship of a company with politics and finance, focusing on board interlocks rather than ownership ties.

Another aspect of relations between individuals is given by embeddedness and social capital theories (Granovetter 1973; Coleman 1990). Granovetter's theory of embeddedness, specially the idea of the strength of weak ties, says that people

who are weakly embedded in society are more likely to spread information among different groups of the society, namely they fulfill the bridge role in a society. As about Coleman's social capital concept, integration is one important dimension, since it is significant in the flow of information. For instance, in a more or less close group (such as a club or class) the actors are more quickly informed than non members. With the spread of the social network analysis paradigm, social capital theories were reconsidered, and they originated social capital in social relations (e. g. Lin [2001]).

Social capital and embeddedness can have another meaning in studies concerning the main actors of networks, such as opinion leaders and brokers. Opinion leaders were first associated with the two-steps flow of information theory: opinion leaders spread information about politics, fashion, movie attendance from media to their closest "ties": their family, relatives and friends (Katz-Lazarsfeld 1955). In other domains, such as medicine (Coleman et al. 1957), the role of opinion leaders proved to be very important. While first researches related to opinion leaders focused on classification (Katz 1957; Merton 1949; Katz-Lazarsfeld 1955; Katz 1963; later Weimann 1994), and on methodology (Rogers 1962), with the spread of the social network analysis paradigm the concept of opinion leaders was revisited. As such, the opinion leaders were defined as central figures of the group (for example Burt 1999, 2005), and besides them another important actor emerged: the broker, who can spread information between different social groups (Fernandez-Gould 1994). Ronald Burt (1999) introduced the term opinion broker, regarding the role of actors, not only on an individual level, but also among groups. According to this concept, opinion brokers are those persons who are using structural holes to spread information from one opinion group to the other. The specific feature of opinion brokers is that, while they are active in their own group, they have strong connections with other groups as well. Thus opinion brokers benefit from structural holes, just like the *tertius gaudens* (ex. Merton 1957).

Groups of people, such as organizations (for example at Kilduff-Tsai 2003) or even states (Walker 1966; or Gray 1973), can have the same network function as individual actors, and this leads back to the meso and macro level of relations, mentioned by Szántó and Tóth (1993). When interpreting social structures as a set of relations, the aim of the research is more and more a description and scientific explanation of social networks (Szántó-Tóth 1993: 45). The authors differentiate three types of networks: when the social entities are persons, they speak about networks of interpersonal relations; when the entities are social groups they refer to the network among social groups, classes or organizations, and when the entities are countries or regions, the network is among these entities.

This article takes into consideration both the relation between people and relation between social groups, emphasizing the role of key actors. The data and methodology used during the investigation at micro and meso level can be read as follows.

## Methodology and data

27 subjects were questioned with life-path interviews, using the snowball method among the Csángó elite, between November 2009 and August 2010. At the end of the interview, relational data was collected: all subjects named their relations with whom they work together regularly at events of Csángó concern. The total number of relations mentioned was 342. Subjects were also asked to name organizations and foundations they are members of. During the interview it became evident that being very active in the Csángó matter is not a prerequisite of being member of the elite, hence only the indegree (number of relations showing to an actor) indicator was used in the social network analysis. Costenbader and Valente (2003) debate the importance of indegree indicator which can be used even at a low sampling level, although the authors speak of randomly missing data.

During the data collection our questions referred to four important aspects: 1. who can be considered elite or opinion leader among Csángós. 2. who plays an important role in the Csángó matter, 3. whom the others engaged in the Csángó matter regularly collaborate with, and 4. which organizations are they the members of.

As a result, we could outline three networks: relations between the members of the elite, relations between most important people involved in the Csángó matter, and relations between organizations and social events (referred to as social groups) they are involved in. In this article we use social network analysis as a representation; the explanatory causes can be found qualitatively in the life interviews of the questioned people. Network analysis was performed with UCINET. In identifying the key figures of the Csángó matter we used centrality and betweenness (betweenness centrality of an actor is the number of shortest paths that pass through that actor). As far as the methodological aspects are concerned, we relied heavily on Letenyei (2005).

## Results

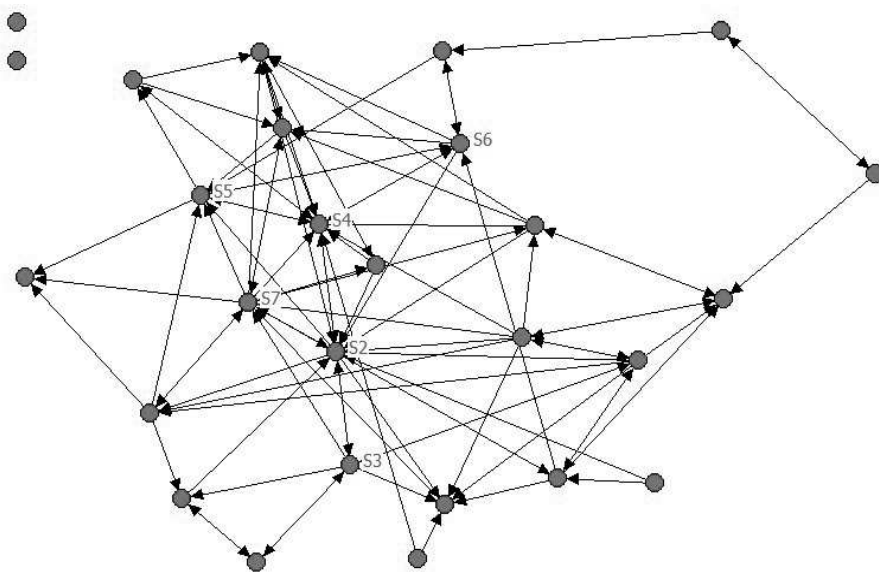
Taking into consideration that the Csángó matter mainly contains cultural events and organizations with the goal to maintain and preserve, Csángó people who identify themselves and others as elite, during the fieldwork and interviews, were mainly part of the Csángó cultural elite. That explains why our research primarily focuses on the Csángó elite and their relations. Nevertheless, with the interpretation of the results, we should always remember that the Csángó matter has a

considerable number of members and sympathizers, and if we extend the analysis to all people involved in the Csángó matter (not only Csángos), the patterns might, and probably will change. That is why our results must be interpreted only in the limit of the Csángó elite involved in Csángó matter. At the same time, since the Csángó matter is so diversely interpreted, we consider not only the relation between people, but also the relation between social groups (such as organizations and events), representing the social structure of the Csángó matter together, which consists of the same or similar goals, of people and social groups that take an active part.

## Micro level analysis

The first network represents the relation between the interviewed Csángó elite. The network consists of 27 subjects and a total number of 96 directed relations. Key actors were coded S1, S2, S3, S4, S5, S6, and S7.

Figure 1: Cooperation network among questioned elite actors



Source: Social network analyze based on data collected November 2009–August 2010. The illustration was made with NETDRAW.

According to Costenbader and Valente (2003), when we analyze the key actors, we take into consideration the indegree network indicator which is usable even with significant missing data. The key actors in the cooperation network

of the Csángó elite regarding the Csángó matter are S2 (11), S4 (11), S5 (7), S7 (7). The centrality (betweenness) indicator partly supports this: S2=126,692, S7=111,132, S3=61,142, S6=59,958, S4=57,183. In Burt's (1999) terms S2 and S7 can be considered both opinion leaders and opinion brokers as well. The number of non-directly connected pairs was high in the case of S5 (17) and S6 (16).

Table: Indicators of cooperation networks related to Csángó matter

Indicators	Figure 1	Figure 3
Centrality (degree) According to the total number of relations (assuming symmetric networks)	S2=14 S7=13 S4=11 S5=9	S2=34 S7=32 S3=31 S1=22
Centrality (In-degree)	S4=11 S2=11 S7=7 S5=7	S4=11 S2=11 S7=7 S5=7
Centrality (betweenness)	S2=126,692 S7=111,132 S3=61,142 S6=59,958 S4=57,183	S2=1604,735 S3=1001,724 S7=842,698 S4=581,405 S6=527.856
Egonet Broker: the number of non-directly connected pairs.	S5=17 S6=16  S7=49 S2=44 S4=33	S5=89 S3=75 S6=52  S2=241 S4=107 S7=171

Source: Based on data collected November 2009–August 2010

According to the relation between the interviewed people we can say that these actors have a prominent role in the Csángó matter: they are responsible for spreading information from one social group to another, while being the opinion leaders of their own social groups as well. It is natural to question the reason why these particular actors have a key role in the Csángó matter. What kind of qualities they have when compared to others (also important figures, also part of the elite)? To satisfy partly the curiosity of the reader, we give an overview of the most important actors. S1 after working together with an organization in Moldova, started his own foundation in a neighboring village, focusing on cultural – mostly folk music and dance – aspects. S2 after graduating, settled in Hungary, regularly works together with the Pro Minoritate Foundation and the Teleki Foundation, and has an active role in the organization of the Csángó Ball as well as several other festivals and exhibitions concerning Csángós. S2 also participates in the organization of village days in Moldova on a yearly basis, keeps regular contact with Moldova, and participates in events both in Moldova and Hungary, building a bridge between the geographic regions. S3 is a teacher in the MCSMSZ, has an

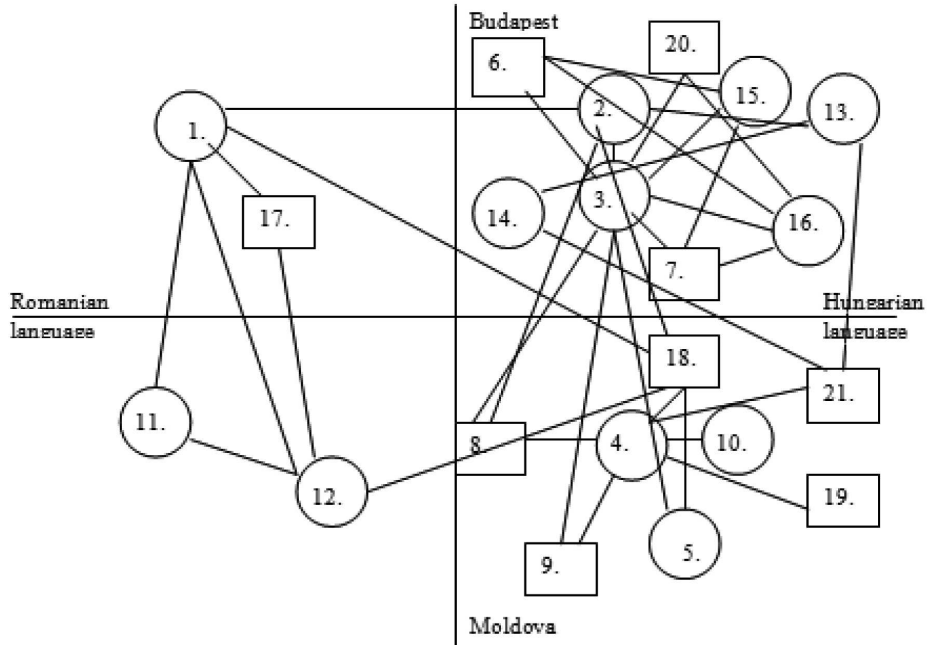
active role in the organization of the Csángó Ball and in village days, representing Csángós at festivals on a regular basis. S4 is the Csángó representative of the Maltese Charity Service in Hungary, organizes village days in Moldova and was one of the main figures in the Csángó Tükör Journal. (Later he started the first Csángó radio, broadcasting via the internet). S5 also settled in Hungary, published several books regarding Csángós, represents them in several occasions – even on a political level – regularly wrote in the journal. S6 is a priest in Hungary, who in some aspects builds a bridge between the Csángós in Hungary and the MCSMSZ. Participates in the organization of village days, conferences and writes articles on religious topics concerning Csángós. S7 is probably of the highest importance as he represented Csángó interests in political spheres, and participates actively in the organization of the Csángó Ball, linking the expatriate Csángó community and Csángós in Moldova.

It is obvious that the active membership in different social organizations and events is very important for each opinion leader and broker. This is why we first examine the meso level of our theoretical approach: the relation between social groups, such as organizations and events related to Csángó matter.

## Meso level of analysis

Based on interviews and field knowledge we could draw the relation between different organizations related to the Csángó matter. The lines between different organizations and events aim to show the regular cooperation among them, yet it is vital to mention that these relations are often sustained on a personal level. Therefore, the arbitrary aspect of this drawing needs to be clarified, the goal is not to accurately represent statistical values, but to visualize the organizational structure of the Csángó matter.

Figure 2: Location and relation among organizations and events concerning the Csángó matter



Source: Analysis based on data collection, November 2009–August 2010

Legend: **Organizations:** Csángók a Csángókért Baráti Kör (3), MCSMSZ (4), Szeret-Klészse Alapítvány(5), Pusztinai Házért Egyesület (10), Dumitru Mártinas Association (11), KEMCSE (13), AMMOA(14) Episcopate of Iasi (12), Pro Minoritate Foundation (15), Teleki Foundation (16.) **Events:** Csángó Tükör (6), Csángó Ball (7), Village days of Magyarfalu (8), Village week of Somoska (9), Romanian masses in Budapest (17), Asking for Hungarian masses in Moldova (18), Folk song competitions (19), Táncház in Hungary (20)

The prominent importance of some groups was confirmed by the CONCOR analysis of UCINET. This analysis could easily identify MCSMSZ (4), Pro Minoritate Foundation (15), Csángó Ball (7), KEMCSE (13) groups, which can be considered the largest ones, from the point of view of attendance and popularity. Other groups were mixed together or split apart, and upon analysis of qualitative data it can be seen that these groups did not have a large number of nominations from our interviewed actors. But the nomination cannot be considered in correlation with their importance, so it can be concluded that CONCOR analysis shows only the popularity of the biggest organizations and events. The question what the real relations between prominent Csángó people and prominent organizations are, remains.

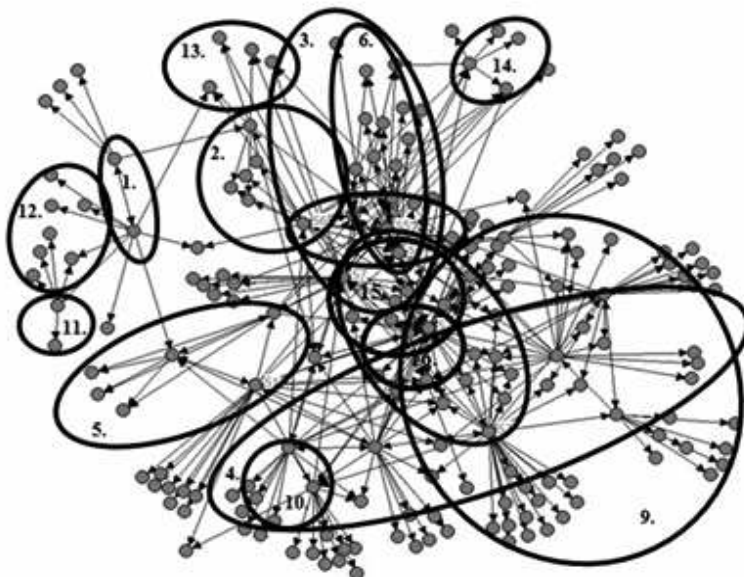
The following figure contains those social groups that appeared in the ego networks of the interviewed elite. The total number of actors is 183 and there



are 342 nominations. In the network, we rounded manually all social groups the actors are enrolled in.

Taking into consideration ego networks, it came to light, that actors are seen as important among elites, are also recognized by other people as prominent in the Csángó matter. The indegree indicator of key actors was the same in this network, compared with the cooperation network of elites only. The centrality of S2 (34), S7 (32), S3 (31), S1 (22) is the highest. The number of non-directed ties is higher in the case of S5 (89), S3 (75), S6 (52). The betweenness highlights the role of S2 (1604,735), S3 (1001,724), S7 (842,698), S4 (581, 405), S6(527,856). We can say that our results, confirmed Burt's (1999) observation that if opinion leaders are responsible for diffusion of information between social groups, they will have an opinion broker role, too. Moreover, in some cases these actors do not simply link groups of people, but they serve as bridges between Budapest – Csángós living in Budapest – and Moldova.

Figure 3: Cooperation network concerning the Csángó matter (actors and social groups)



Source: Social network analyze based on data collection, November 2009–August 2010. The illustration was made with NETDRAW

It can be observed that key actors of the network, opinion leaders and opinion brokers, have three main characteristics.

First, prominent actors are members of the greatest number of social structures. The activity of these people is emphasized by their organizational membership: every key actor of the network is member of at least 3 important organizations or events related to the Csángó matter. There is a special case, that of S1, who

can be considered a key actor only in this network. The key position of S1 is due to the fact that after being active in one organization, started his own foundation in a neighboring village in Moldova, focusing on teaching folk songs and folk music to the young generation. S1 is in good relations with the local political leaders and that proves to be beneficial. We can conclude that in this particular network the importance of the location of groups is noticeable – the two locations Moldova and Budapest, are far from each other – and although the key actors are considered key actors in Hungary, their inherence must be encouraged in Moldova, too.

Secondly, prominent actors are members of several organizations or events that have the greatest popularity among the Csángós, such as MCSMSZ (4), Pro Minoritate Foundation (15), Csángó Ball (7), KEMCSE (13). Meanwhile, others participate in less events and organizations. For example MCSMSZ (4), the Association of Csángó-Hungarians in Moldova, is the largest in number, with the support of KEMCSE (13), the Association of Godparents for Csángó-Hungarians from Moldova, has a very large educational program. The Csángó ball (7) is popular not only among Csángós, since it is open to everyone interested, therefore, it can be considered an important yearly event in Hungary. Pro Minoritate Foundation (15) has also a very important role in the field of culture and science.

Finally, the most remarkable feature of this network is that even the Csángó elite itself considers only a few active members in relation to the Csángó matter. The number of those people who are seen as active by many others is quite low, although it is obvious that the Csángó matter involves many people and has even more supporters and sympathizers, but being very active and known as such by others is restricted to only a few individuals. As we observed in Figure 1, opinion leaders' and opinion brokers' roles are fulfilled by the same people in the large cooperation network. It seems that social involvement correlates with people's opinion: the most active people are nominated as key actors.

## Discussion and conclusion

Fieldwork and interviews underline what can be seen in the network analysis as well i.e. that the entire issue of the Csángó matter as such is multi-segmented and cross-layered. On one hand there is the identity dimension dividing the population based on whether they identify themselves as Romanian or Hungarian. In this article we focused mostly on the latter group, with Hungarian self-identification. On the other hand, there is the geographical dimension that places the Csángó elite in Moldova and in Hungary, mostly in Budapest. According to linguistic and geographic location, we can find numerous organizations and events forming different social groups of different people that share the same goal: to preserve the Csángó culture and slow down assimilation. This interconnectivity, or in Scott's terms interlock, of people and social groups can be perceived as the social structure of the Csángó matter.

In this social structure we can find some key actors that have three main characteristics:

1. They are members of the greatest number of social groups.
2. There is at least one very popular social group among these social groups.
3. The central figures of a social group (the opinion leader) have a central position among various groups, too (they are also opinion brokers).

Because of the very intensive activity of key actors we call them the “prominent Csángó hub”. Moreover, their network position is recognized by other Csángós, and they serve as a link between Hungary and Moldova – regardless of their residence – by attending and organizing events on both geographic locations. Formal and informal activity is equally important for key actors: it seems that besides being the member of formal organizations and foundations, it is equally important to participate in the organizations of the Csángó ball, the publication of the Csángó Tükör, in organizing village days and artistic exhibitions, etc.

Social network analysis along with the life-path interviews made clear that elite definition is not necessarily formed by using several fronts of objective measures – such as level of education, income, wealth, etc. – but is just as important to take part in as many organizations and events related with Csángó matter as it is possible (including very popular organizations and events), for being capable to promote the Csángó matter in order to be considered central. Those key actors, who have a central position in the Csángó matter social structure, are the most visible persons. Their activity is recognized not only in the social groups they participate in, that is on a micro level, but between social groups (on a meso level) and between countries (on a macro level), too.

This research shows one important direction to follow regarding the investigation of the prominent Csángó hub. Taking into consideration that the Csángó matter is geographically widespread, it would be interesting to research whether the strength and direction of relations between geographic regions will modify the network structure presented in this article, which could be the extension of the research to the macro level.

## References

- Adler-Lomnitz, Larissa – González, Asaul Nuche (2007): Peasant Ethnic Migration from a Post-communist Society to Another. The Case of the Csángó Migrants from Moldavia (Romania) to Hungary. *Review of Sociology*, (13)1: 67–88.
- Arens, Meinolf – Daniel Bein – Radu Rosetti (2004): Katolikus magyarok Moldvában. In Miskolczy Ambrus (szerk.): *Rendhagyó nézetek a csángókról*. Budapest: ELTE Román Filológiai Tanszék – KSH Levéltára.
- Baker, Robin (1997): On the Origin of the Moldavian Csángos. *The Slavonic and East European Review*, (75)4: 658–680.

- Barnes, John A. (1954): Class and Committee in a Norwegian Island Parish. *Human Relations*, (7): 39–58.
- Bott, Elisabeth (1955): *Family and Social Network*. London: Tavistock.
- Benda, Kálmán (1989): *Moldvai csángó–magyar okmánytár 1467–1706. I–II*. Budapest: Magyarországtudató Intézet.
- Burt, Ronald. S. (1992): *Structural Holes: The Social Structure of Competition*. Cambridge, Mass.: Harvard University Press.
- Burt, Ronald. S. (1999): The Social Capital of Opinion Leaders. American Academy of Political and Social Science. *Annals of the American Academy of Political and Social Science*. 566. The Social Diffusion of Ideas and Things, 37–54.
- Burt, Ronald. S. (2005): *Brokerage and Closure*. Oxford University Press.
- Coleman, James – Katz, Elihu – Menzel, Herbert (1957): The Diffusion of an Innovation among Physicians. *Sociometry*, 20(4): 253–270.
- Coleman, James S. (1988): Social Capital in the Creation of Human Capital. *The American Journal of Sociology*, 94. Supplement: Organizations and Institutions: Sociological and Economic Approaches to the Analysis of Social Structure, 95–120.
- Coleman, James (1994 [1990]): *Foundations of Social Theory*. Cambridge: Harvard University Press. In Lengyel György – Szántó Zoltán (eds.). *A gazdasági élet szociológiája*. Budapest: Aula, 99–129.
- Cosa, Anton (2007): Catolicii din Moldova in izvoarele Sfantului Scaun (Secolele, XVII–XVIII.). Iasi: Editura Sapientia.
- Costenbader, E – Valente, T. W. (2003): The Stability of Centrality Measures when Networks are Sampled. *Social Networks*, 25(4): 283–307.
- Domokos, Pál Péter (1940): *Adalékok Moldva történetéhez*. Cluj-Kolozsvár.
- Domokos, Pál Péter (1987): *A moldvai magyarság*. Budapest: Magvető Kiadó.
- Fernandez, Roberto M. (1991): Structural Bases of Leadership in Intraorganizational Networks. *Social Psychology Quarterly*, (54)1: 36–53.
- Fernandez, Roberto M. – Gould, Roger V. (1994): A Dilemma of State Power: Brokerage and Influence in the National Health Policy Domain. *American Journal of Sociology*, (99)6: 1455–1491.
- Gerő, Barnabás – Balázs, Vedres (1998): Interlocking Comrades. Paper Presented at the 11th International Conference of Europeanists, March 2, 1998. Baltimore. [http://www.personal.ceu.hu/staff/Balazs\\_Vedres/papers/Comrade/comrade1.html](http://www.personal.ceu.hu/staff/Balazs_Vedres/papers/Comrade/comrade1.html)
- Granovetter, Mark (1973): The Strength of Weak Ties. *American Journal of Sociology*, 78(6): 1360–1380.
- Halász, Péter (2002): *Bokrétába kötögetem vala*. Budapest: Európai Folklór Intézet.
- Letenyei, László (2005): *Településkutatás*. Budapest: L'Harmattan–Ráció Kiadó.
- Lin, Nan (2001): Building a Network Theory of Social Capital. In Nan Lin – Karen Cook – Ronald S. Burt (eds.). *Social Capital – Theory and Research*. New York: Aldine de Gruyter, 3–31.

- Katz, Elihu (1957): The Two-Step Flow of Communication: An Up-to-date Report on an Hypothesis. *Public Opinion Quarterly*, (21): 61–78.
- Katz, Elihu (1963): Traditions of Research on the Diffusion of Innovation. *American Sociological Review*, (28): 237–253.
- Katz, Elihu – Lazarsfeld, Paul F. (1955): *Personal Influence*. New York: Free Press.
- Kilduff, Martin – Tsai, Wenpin (2003): *Social Networks and Organizations*. Sage.
- Kinda, István – Pozsony, Ferenc (eds.) (2005): *Adaptáció és modernizáció a moldvai Csángó falvakban*. Kolozsvár: Kriza János Néprajzi Társaság.
- Mărtinaș, Dumitru (1985): *Originea Ceangailor din Moldova*. Bucuresti: Editura Stiintifica si Enciclopedica.
- Mayo, Elton (1945): *The Social Problems of an Industrial Civilization*. London: Routledge and Kegan Paul.
- Mikecs, László (1989): *Csángók*. Pécs: Optimum Kiadó.
- Merton, Robert K. (1968 [1949]): Patterns of Influence: Local and Cosmopolitan Influentials. In *Social Theory and Social Structure*, New York: Free Press, 441–474.
- Merton, Robert K. (1957): *Social Theory and Social Structure*. Glencoe, Illinois: Free-Press.
- Mitchell, Clyde (1969): *Social Networks in Urban Situations: Analysis of Personal Relationships in Central Towns*. Manchester: Manchester University Press.
- Mohácsék, Magdolna – Vitos, Katalin (2005): Fogyasztói szokások a magyarfalusi vendégmunkások köreiben. *Erdélyi társadalom*, (1): 2.
- Moreno, Jakob Lévy (1937): Sociometry in Relation to Other Social Sciences. *Sociometry*, 1 (1/2) 206–219.
- Némedi, Dénes (2008): *Modern szociológiai paradigmák*. Budapest: Napvilág Kiadó.
- Peti, Lehel (2006): A csángómentés szerkezete és hatásai az identitásépítési stratégiákra. In Jakab, Albert Zsolt – Szabó, Árpád Töhötöm (eds.): *LeNyomatok 5*. Fiatal kutatók a népi kultúráról. Kolozsvár: KJNT, 129–155.
- Pozsony, Ferenc (1996): Etnokulturális folyamatok a moldvai csángó falvakban. In Katona, Judit – Viga, Gyula (eds.): *Az interetnikus kapcsolatok kutatásának újabb eredményei*. Miskolc: Herman Ottó Múzeum, 173–179.
- Pozsony, Ferenc (2003): *A moldvai csángó-magyar falvak társadalomszerkezete*. Pro Minoritate Foundation. Nyár, 142–165.
- Pozsony, Ferenc (2005): *A moldvai csángó magyarok*. Budapest: Gondolat Kiadó–Európai Folklor Intézet.
- Radcliffe-Brown, A. R. (1940): On Social Structure. *Journal of the Royal Anthropological Institute of Great Britain and Ireland*, (70): 188–204.
- Rogers, Everett M. (2003 [1962]): *Diffusion of Innovations*. Glencoe: Free Press.

- Sabel, Charles F. (1993): Studied Trust: Building New Forms of Cooperation in Volatile Economy. *Human Relations*, (46): 9.
- Schultz, Theodore W. (1961): Investment in Human Capital. *The American Economic Review*, (51)1: 1–17.
- Scott, John (1997): *Corporate Business and Capitalist Classes*. Oxford: Oxford University Press.
- Szántó, Zoltán – Tóth, István György (1993): A társadalmi hálózatok elemzése. *Aula*, XV. 1: 30–55.
- Tánczos, Vilmos (1996): “Én román akarok lenni!” Csángók Erdélyben. In Tánczos, Vilmos: *Keletnek megnyílt kapuja*. Kolozsvár: KOM-PRESS, Korunk Baráti Társaság.
- Tánczos, Vilmos (1997): Hányan vannak a moldvai csángók? *Magyar Kisebbség*, III. évf. (1–2): 370–390. [Also available in english Tánczos, Vilmos (1998): Hungarians in Moldavia, <http://mek.niif.hu/00900/00983/pdf/moldvang.pdf>]
- Uzzi, Brian (1996): The Sources and Consequences of Embeddedness for the Economic Performance of Organizations: The Network Effect Author. *American Sociological Review*, (61)4: 674–698.
- Vedres, Balázs – Stark, David (2010): Structural Folds: Generative Disruption in Overlapping Groups. *American Journal of Sociology*, (115)4: 1150–1190.
- Waker, Jack (1969): The Diffusion of Innovations among the American States. *The American Political Science Review*, (63)3: 880–899.
- Warner, William L. – Lunt, Paul S. (1941): *The Social Life of a Modern Community*. New Haven: Yale University Press.
- Warner, William L. – Lunt, Paul S. (1942): *The Status System of a Modern Community*. New Haven: Yale University Press.
- Weimann, Gabriel (1994): *The Influentials. People Who Influence People*. New York: State University of New York Press.