

## LIST OF CONTRIBUTORS

- Howard E. Aldrich* Department of Sociology, UNC-CH,  
Chapel Hill, NC, USA
- Carol A. Caronna* Department of Sociology, Anthropology, and  
Criminal Justice, Towson University,  
Towson, MD, USA
- Jeannette A. Colyvas* School of Education, Stanford University,  
Stanford, CA, USA
- James A. Evans* Department of Sociology, University of  
Chicago, Chicago, IL, USA
- Phillip H. Kim* School of Business, University of Wisconsin,  
Madison, WI, USA
- Michael Lounsbury* University of Alberta School of Business,  
National Institute for Nanotechnology,  
Alberta, Canada
- Ching Lin Pang* Social and Cultural Anthropology (IMMRC),  
Catholic University of Leuven, Leuven,  
Belgium
- Alejandro Portes* Department of Sociology, Princeton  
University, Princeton, NJ, USA
- Walter W. Powell* Stanford University and SCANCOR,  
Stanford, CA, USA
- Jan Rath* Institute for Migration and Ethnic Studies  
(IMES), University of Amsterdam,  
Amsterdam, Netherlands
- Martin Ruef* Department of Sociology,  
Princeton University, Princeton, NJ, USA

*Steven Shafer*

Department of Sociology, Princeton  
University, Princeton, NJ,  
USA

*Luca Solari*

Dipartimento di Studi del Lavoro-Facoltà di  
Scienze Politiche, Università Degli Studi di  
Milano, Milano, Italy

*Jesper B. Sørensen*

Graduate School of Business, Stanford  
University, Stanford, CA, USA

*Hongwei Xu*

NUS Business School, National University of  
Singapore, Singapore

# EDITORIAL BOARD

Howard E. Aldrich  
*University of North Carolina*

Paul M. Hirsch  
*Northwestern University*

Stephen R. Barley  
*Stanford University*

Karin Knorr Cetina  
*University of Constance*

Nils Brunsson  
*Stockholm School of  
Economics*

Christel Lane  
*University of Cambridge*

John Child  
*Birmingham Business School*

Walter W. Powell  
*Stanford University*

Elisabeth S. Clemens  
*University of Chicago*

Hayagreeva Rao  
*Stanford University*

Barbara Czarniawska  
*Göteborg University*

Kerstin Sahlin-Andersson  
*Uppsala University*

Gerald F. Davis  
*University of Michigan*

W. Richard Scott  
*Stanford University*

Marie-Laure Djelic  
*ESSEC Business School*

Robin Stryker  
*University of Minnesota*

Frank R. Dobbin  
*Harvard University*

Haridimos Tsoukas  
*Athens Laboratory of Business  
Administration (ALBA)*

Royston Greenwood  
*University of Alberta*

Richard Whitley  
*University of Manchester*

# INTRODUCTION: THE SOCIOLOGY OF ENTREPRENEURSHIP

Martin Ruef and Michael Lounsbury

## ABSTRACT

*The sociology of entrepreneurship is a blossoming field of research, but its scholarly contribution has been critiqued for its lack of coherence and intellectual distance from the sociological mainstream. In this article, we critically examine the theoretical presuppositions of the field, trace its historical origins, and attempt to situate the sociology of entrepreneurship within the sociological canon. We place special emphasis on the contribution of Max Weber, whose early work provides a useful template for a comprehensive approach to understanding the context, process, and effects of entrepreneurial activity. We conclude by locating contemporary approaches to entrepreneurship – including the contributions in this volume – within this neo-Weberian framework.*

## INTRODUCTION

The sociology of entrepreneurship analyzes the social context, process, and effects of entrepreneurial activity. Within this perspective, “entrepreneurship” can be construed either narrowly as purposive action leading to the creation of new formal organizations, or more broadly as any effort to

---

**The Sociology of Entrepreneurship**

**Research in the Sociology of Organizations, Volume 25, 1–29**

**Copyright © 2007 by Elsevier Ltd.**

**All rights of reproduction in any form reserved**

**ISSN: 0733-558X/doi:10.1016/S0733-558X(06)25001-8**

introduce durable innovations in routines, technologies, organizational forms, or social institutions. Research in the sociology of entrepreneurship tends to differ from related work in industrial psychology and economics in three basic respects. First, it often targets levels of analysis beyond the individual entrepreneur, addressing the role played by interpersonal networks, organizational structure, population, and field-level processes, as well as the broader institutional environment. Second, it balances the common emphasis on material aspects of venture formation (e.g. market conditions and financing) with attention to the symbolic and cultural dimension of entrepreneurial activity. Third, it seeks to understand entrepreneurship in a diverse set of contexts, including arenas – such as science, health care, and the fine arts – that tend to elude simple market-based accounts.<sup>1</sup>

During the past few years, sociologists have produced a rapidly growing body of research on entrepreneurship, including a number of noteworthy efforts to survey this nascent field (e.g. Aldrich, 2005; Keister, 2005; Thornton, 1999). In this prolific context, the jaded reader may well ask why there exists a need for “yet another” review article on the sociology of entrepreneurship and another volume highlighting current empirical work on the topic. While we could offer any number of self-serving explanations, it seems sufficient to acknowledge that the promise of sociological research on entrepreneurship has thus far been tempered by a number of factors. Unlike other subfields of the discipline that have experienced explosive growth, such as economic sociology, there is little intellectual cohesion in research on entrepreneurship. Few scholars appear to agree on a common sociological conception of entrepreneurship, much less on a canonical history of work on the topic. The geographic myopia often maintained by existing research is equally problematic. Mirroring the parochialism of organizational studies more generally (see March, 2004, for a critique), American scholars tend to ignore relevant studies of entrepreneurship by their international counterparts, while international scholars feel rebuffed by Anglo-American publications.<sup>2</sup> Compounding issues of intellectual fragmentation, sociological observers of entrepreneurship have sometimes spilled a disproportionate amount of ink on phenomena that are peripheral to the experiences of “average” entrepreneurs and startups. For instance, a large number of recent studies have emphasized venture capital investments (Podolny, 2001; Sorenson & Stuart, 2001), even though this funding source comprises a meager three percent of the first-stage financing for nascent entrepreneurs in the United States (Kim, Aldrich, & Keister, 2004), and even less elsewhere. Meanwhile, the principal exchange mechanisms supporting new ventures – e.g. resource pooling among teams of entrepreneurs – have received far less attention in mainstream sociology journals.

Taken together, these considerations paint an image of a subfield that is somewhat adrift, with little sense of its own intellectual history, international research community, and core substantive problems. In this overview, therefore, we do not offer another synoptic review, but, rather, attempt to trace the origins of the field and to suggest a blueprint for progress in the future. We begin with a brief history of the sociology of entrepreneurship, placing emphasis on the scholarship of Weber and, in particular, his often forgotten dissertation on the *History of Commercial Partnerships*. We explore the evolution of the subfield after Weber on an inductive basis, using a JSTOR search of journals pertaining to sociology and organizational theory. Next, we consider whether the sociology of entrepreneurship appears to be a marginalized subfield of economic and organizational sociology. We attend to the international reception of entrepreneurship research, theoretical dissension concerning definitions, and ideological opposition outside the subfield. Finally, we turn to contemporary work in the sociology of entrepreneurship, introducing the empirical papers in this volume. In the interest of fostering collaboration in the field, we suggest how these contributions inform a common multilevel framework for the study of entrepreneurial phenomena.

## A BRIEF HISTORY OF THE SOCIOLOGY OF ENTREPRENEURSHIP

Concepts related to entrepreneurship appear with some frequency in the sociological canon. Weber's (1930[1904–1905]) *Protestant Ethic and the Spirit of Capitalism* is perhaps the most well-known exemplar, with its provocative thesis that worldly asceticism among certain Protestant sects (particularly, Calvinists) yielded an ethic of calculability, efficiency, and self-control that was essential to the rise of entrepreneurial capitalism in the 16th and 17th century.<sup>3</sup> Simmel's essay on "The Stranger" highlighted the relationship between outsider status and middleman entrepreneurs, who made a living from intermediate trade between otherwise closed societies (Simmel, 1990[1907]; see also translation in Wolff, 1950, pp. 402–404). And Durkheim's (1984[1893]) evolutionary account of the division of labor could be rendered as a contribution to the sociology of entrepreneurship, given its explanation of the decline of occupational generalism and the proliferation of autonomous, specialist producers (under conditions of organic solidarity).

A more critical examination of the canon, however, fails to yield a systematic theory of entrepreneurship and, in some cases, an outright avoidance of entrepreneurs as distinctive subjects of sociological inquiry. For

instance, in their discussion of the middle class (*petty bourgeoisie*), neither Marx nor Engels distinguished between its “entrepreneurial” faction (e.g. small business owners, artisans, independent professionals) and elements of the “new middle class” (clerical, technical, and administrative workers, largely employed in bureaucracies) (Bottomore, 1991). Simmel referred often to entrepreneurs in his empirical examples, but readily interchanged the category of entrepreneur with that of capitalist/owner (e.g., see his discussion of triads and the *tertius gaudens* in Wolff, 1950). Even Weber, who presented the most extensive treatment of entrepreneurship in the classic sociological literature, offers only sporadic attention to the topic in his magisterial *Economy and Society* (Swedberg, 2005). All of this may leave the contemporary scholar with some doubt as to whether any of the classic sociological works exhibit a deep interest in entrepreneurship, rather than broader phenomena such as capitalism, rationalization, and modernity.

We argue, to the contrary, that Weber in particular did lay the foundation for a sociology of entrepreneurship. To appreciate his contribution, however, one should not begin with the usual suspects – the *Protestant Ethic, Economy and Society*, or the *General Economic History* – but with his earliest work, the *History of Commercial Partnerships* (hereafter, HCP; Weber, 2003[1889]). The HCP was Weber’s J.D. dissertation and has often been ignored by contemporary scholars, for a number of reasons: pragmatically, because it was only recently translated from German; biographically, because it has been linked to Weber’s “lost decade” prior to the appearance of the *Protestant Ethic* (Käsler, 1988); and, substantively, because its interdisciplinary orientation entailed a mixture of sociology, history, law, and political economy. Nevertheless, the HCP initiated many of the theoretical themes that would be central to Weber’s lifetime of work. Here we find the first traces of his theory of rationalization, with a discussion of the historical separation of household and business (HCP, Chapter 3). And also key elements of his economic sociology, especially an emphasis on the construction of ‘modern’ forms of organization and authority that would later culminate in his ideal type of bureaucracy. But the core narrative in this work revolves around a comparative-historical approach to entrepreneurial activity.

Weber’s substantive interest in the HCP was a comparison of Roman and Germanic commercial laws, with an emphasis on their historical development and influence on the formation of commercial partnerships among medieval entrepreneurs. Weber based his analysis on a systematic review of Italian and Spanish legal charters and statutes from the 11th through the 16th centuries, with detailed case studies of Pisa’s *Constitutum Usus* and commercial law in Florence. In simplified form, his argument in the HCP

can be summarized as the first three rows of Table 1. Beginning with a discussion of Roman property law, Weber (2003, Chapter 1) argues that the legal differentiation of partnerships (*societas*) from individual entrepreneurs participating in them (*socius*) is largely nominal in this legal form: “the partnership, as merely a complex of obligatory relations among the *socii*, is of no concern to third parties; in its legal consequences, a transaction a *socius* makes on the account of the partnership is no different from any transaction made on a personal account” (2003, p. 54). This differentiation between individual and corporate personhood strengthened, however, with the growing needs of maritime and overland trade during the Middle Ages. Two new organizational forms – the *societas maris* and *societas terrae* – relied increasingly on a cash fund that was separate from the assets of entrepreneurs participating in the venture (2003, Chapter 2). As discussed in the HCP, a further development in corporate personhood was the emergence of the “joint household” in Germanic law. This organizational form introduced the concept of solidary liability, whereby the debt of a family or community member “encumbers the joint assets” of that community (2003, p. 98). The concept was quickly generalized in the Middle Ages to commercial partnerships, as joint households were not only defined in terms of kinship ties, but also in terms of cohabitation and “communities of labor” (e.g. craft guilds).

In broad strokes, then, what Weber was identifying theoretically in the HCP were a set of institutional conditions that would allow for the social construction of corporate persons: partnerships that had a legal, economic, and social existence apart from the entrepreneurs that constituted them. These conditions include:

- *Separable identity*: the identity of a corporate actor may be different than that of the participating entrepreneurs (a contemporary example is “assumed name” certificates for proprietorships).
- *Separable resources*: the assets of a corporate actor are separated (in an economic and legal sense) from those of participating entrepreneurs.
- *Solidary liability*: the entrepreneurs qua participants commit to *joint* responsibility for the actions and debts of a corporate actor.
- *Separable liability*: the responsibilities (debts and obligations) of the corporate actor are separated from the *personal* responsibility of the participating entrepreneurs.

While Weber did not analyze the last institutional condition in detail (and, thus, the basis for the modern corporation), he did emphasize that medieval conceptions of partnership based on solidary liability continued to



**Table 1.** A Comparison of the Legal Forms of Entrepreneurial Ventures (based on Weber's, 2003[1889] *History of Commercial Partnerships*).

Contemporary Legal Forms	Early Exemplars	Legal Characteristics of the Form				
		Separable Identity	Separable Resources	Solidary Liability	Separable Liability	
Proprietorship, contractual affiliation	Roman property law, <i>societas</i>	Maybe	No	No	No	↓ Continuity of Existence ↓
Limited partnership	<i>Societas maris, societas terrae, commenda</i>	Yes	Yes	No	No	
General partnership	Joint households, craft guilds	Yes	Yes	Yes	No	
Corporation	Joint stock merchant companies	Yes	Yes	Yes	Yes	

insist that “members are personally liable, as debtors, for one another” (2003, p. 98). As this notion of liability, grounded in communal labor, was replaced by a notion of joint liability, grounded in communal investment, the legal conditions emerged for corporate actors that were fully separate from the “natural” persons that created them (Coleman, 1974).

The sociological issues raised in the HCP should be ones of central concern to the sociology of entrepreneurship. Many of us – especially those raised in an Anglo-American context – take it for granted that individuals can readily construct autonomous, organizational entities to act on their behalf. But, consistent with the intuitions of contemporary neoinstitutional scholars (Meyer & Jepperson, 2000; Hwang & Powell, 2005), Weber’s comparative analysis suggests that societies differ greatly in the amount of agency they accord to entrepreneurs and their organizational ventures. Institutional frameworks strongly influence whether collective enterprises are short-lived affairs that are tied closely to the fates of their creators, or whether they are able to develop as independent and, potentially, perpetual legal fictions.<sup>4</sup>

In addition to studying the *context* affecting entrepreneurial activity, Weber’s scholarship in the HCP also addressed variation in entrepreneurial processes and organizational forms. One rich exemplar is his comparison of the unilateral *commenda*, where investment capital is only provided by a single party, and the bilateral *commenda*, which entails investment from (at least) two parties. Each organizational form involved both a *commendator* (or passive investor) and an entrepreneur known as a *tractator* (see Fig. 1). The unilateral *commenda* differed, however, in that all of the financial risk was born on the part of the passive investor, who contributed to a fund that would be managed by the *tractator*, serving as his or her agent. In contrast to prevailing legal wisdom, Weber (2003, pp. 135–136) argued that the lack of a separate corporate fund in the unilateral *commenda* (rather than a fund that simply contained personal assets of the *commendator*) meant that it was not an institutional precursor of modern partnerships. In fact, the weak organizational foundation that the unilateral *commenda* provided for entrepreneurial ventures (cf. Table 1) led to its replacement by simpler forms of financing, such as commercial loans. The bilateral *commenda*, by contrast, allowed for the existence of a fund that was separate from the assets of investors and entrepreneurs and, therefore, served as a legal template for a distinctive organizational form (the limited partnership).<sup>5</sup>

A final topic addressed by Weber in the HCP, albeit briefly, concerned the *consequences* of entrepreneurship. In surveying the differences between the limited and general partnership, Weber speculated about the implications

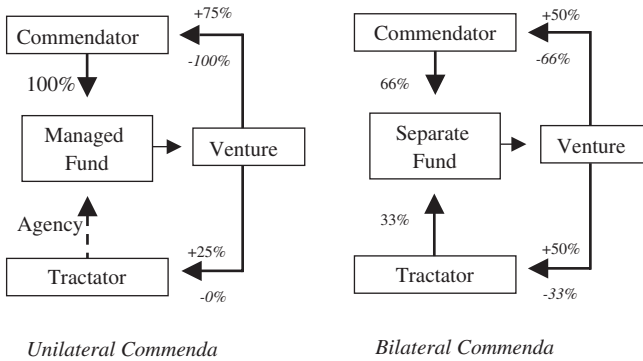


Fig. 1. The Organizational Structure of Two Medieval Forms of Entrepreneurial Partnership (based on Weber's, 2003[1889] *History of Commercial Partnerships*). Note: Positive percentages correspond to financial contributions and shares of profit. Negative percentages correspond to losses borne by each partner.

that these organizational forms held for inequality among participants. He thought that “the *commenda* and its derivative forms, including the limited partnership later on, derived from an association of people who were economically and, as one could say, socially unequal” (Weber, 2003, pp. 146–147). By contrast, joint liability “developed out of associations among equals and people who had an equal right to dispose of property” (*ibid.*, p. 147). Weber thus raised the possibility that different organizational forms had distinct implications for the ‘democratization’ of capital, a theme soon picked up and debated by students of corporate governance (Berle & Means, 1968[1932]; see Mizruchi, 2004 for a review). More generally, he initiated a linkage between the sociology of entrepreneurship and research on stratification, an idea that has only recently regained scholarly currency (Lippmann, Davis, & Aldrich, 2005).

### *The Sociology of Entrepreneurship After Weber*

How did the sociological literature on entrepreneurship evolve after Weber? To provide a systematic overview, we conducted an examination of six journals in JSTOR, beginning in 1895 with the first publication of *AJS*. We grouped the journals into three general categories, including top journals in organizational sociology and management (*Administrative Science Quarterly* and the *Academy of Management Journal*), in American sociology

**Table 2.** The Sociological and Management Literature on Entrepreneurship (1895–1999).

Source	Articles Referring to Entrepreneurial Processes <sup>a</sup>			Articles Referring to Founding Processes <sup>b</sup>		
	<i>ASQ, AMJ</i>	<i>AJS, ASR</i>	<i>BJS, ESR</i>	<i>ASQ, AMJ</i>	<i>AJS, ASR</i>	<i>BJS, ESR</i>
1895–1945	n/a	0	n/a	n/a	9	n/a
1946–1969	7	5	0	0	1	0
1970–1979	4	3	1	1	2	0
1980–1989	10	14	1	5	5	2
1990–1999	16	10	1	22	10	1
<i>Total</i>	37	32	3	28	27	3

<sup>a</sup>We enumerate all articles that have the keywords “entrepreneur(s)”, “entrepreneurial”, or “entrepreneurship” in their titles or abstracts.

<sup>b</sup>We enumerate all articles that have the keywords “founder(s)” or “founding” in their titles or abstracts, excluding the verb form of “founder”.

(the *American Journal of Sociology* and *American Sociological Review*), and in European sociology (the *European Sociological Review* and *British Journal of Sociology*). We began by assessing explicit references to entrepreneurial processes in the titles or abstracts of journal articles. The first three columns of Table 2 provide a quantitative summary of article counts over time.

During the period before World War II, the sociological literature on entrepreneurship was relatively barren. Several key works relevant to the topic – including Weber’s *Protestant Ethic* and Schumpeter’s (1934) *Theory of Economic Development* – were not translated from the original German until the 1930s (Swedberg, 2000). Moreover, the subfield of organizational sociology itself, which would provide much of the impetus to the study of entrepreneurship in the post-war years, existed only as a vague amalgam of scientific management and human relations thinking (Scott, 2003). Those desiring a dedicated introduction to the sociology of entrepreneurship would need to seek it in an unlikely source: a short, 52 page book published by the Austrian political economist Eugen Schwiedland in 1933. Entitled “*Toward a Sociology of Entrepreneurship*” (*Zur Sociologie des Unternehmertums*), Schwiedland’s monograph was primarily a policy tract favoring organized support for entrepreneurs under the auspices of a central state authority (see Hughes, 1934, for a critique). It was never translated and had no appreciable scholarly impact.

The post-war years witnessed a number of developments that were more favorable to sociological research on entrepreneurship. First, the burgeoning

field of organizational theory contributed to the founding of *Administrative Science Quarterly* (ASQ) in 1956 and the *Academy of Management Journal* (AMJ) in 1959. Starting with Heinz Hartmann's (1959) pathbreaking article, contributors to these journals began to examine the tension between formal authority and entrepreneurial tendencies in established bureaucracies (e.g. Becker & Gordon, 1966; Litzinger, 1963; Davis, 1968). What was perhaps most noteworthy about this research stream is that scholars quickly abandoned Joseph Schumpeter's well-known emphasis on the innovative function of entrepreneurship in favor of a Weberian approach, in which entrepreneurs were defined as holding a special relationship to formal bureaucratic authority. Hartmann, for instance, maintained that Weber's "sociological differentiation between the manager and the entrepreneur [was] a useful replacement for Schumpeter's scheme" (1959, p. 436). Empirical work highlighted the distinctive features of the entrepreneurial role (independence, nonconformity, attention seeking, etc.) and the difficulties inherent in transitioning from entrepreneurial to bureaucratic leadership in maturing organizations (Litzinger, 1963; Davis, 1968). Around the same time, members of the Tavistock group in Britain began to investigate structural differences between entrepreneurial and bureaucratic organization (Burns & Stalker, 1961; see Sine, Mitsuhashi, & Kirsch, 2006, for a critical appraisal).

A second propitious development was that the core sociology journals in the United States (*AJS* and *ASR*) also began to pay some attention to entrepreneurial phenomena. For instance, in her article on adaptation by Chinese immigrants, Lee (1949) documented the changing organizational structure of Chinatown businesses in Butte, Montana, thus initiating a productive research stream on enclaves of ethnic entrepreneurs. Other authors emphasized the distinct position of entrepreneurs in the status structure of occupations (e.g. Sarapata & Wesolowski, 1961), linking studies of entrepreneurs to mainstream sociological research on stratification. Still, entrepreneurship remained a peripheral topic for sociologists and much of the research invoking the label was merely metaphorical. The assessment of Everett C. Hughes (1952), that the object of sociological study "is more likely the slum-dwelling employee of the trader and the entrepreneur than the entrepreneur himself [*sic*]", appeared to ring true.

By the mid-1970s, sociological research that explicitly identified with the topic of entrepreneurship slowed to a trickle. The newly emerging paradigms in organizational analysis – especially, population ecology, and neo-institutional theory – downplayed attention to innovative or disruptive participants in organizations in favor of an emphasis on the environment. The earlier research stream that had highlighted the role distinctions

between entrepreneurs and managers was reformulated at the individual level, largely as a topic of social psychological inquiry (e.g. Ondrack, 1973; Brockhaus, 1980). Sustained attention to entrepreneurship within organizational sociology persisted primarily in the study of creative industries (e.g. popular music, movie production, etc.), leading to what is now known as the “production of culture” perspective (Peterson & Berger, 1971; Hirsch, 1972; see Peterson & Anand, 2004, for a review). A notable feature of this literature was its open rejection of a purely psychological view of entrepreneurial activity. As Peterson & Berger (1971, p. 98) argued, “entrepreneurship does not emerge automatically[.] persons with the psychology and motivation necessary for entrepreneurship must be in the strategically appropriate locations”. This literature also called attention to the routinized nature of most cultural entrepreneurship – rather than generating radical cultural innovations, the enterprises that persist in this arena filter out a large number of songs, films, books, and the like in order to offer a steady stream of mundane “hit” products (Hirsch, 1972).

In the United States, sociological research on entrepreneurship proliferated in the 1980s, but appeared to suffer from a lack of definitional coherence and common substantive foundation. Influenced by the ecological thrust in organizational theory (e.g. Aldrich, 1979), some scholars sought to explain different rates of startup activity in terms of the regional environment confronting entrepreneurs (Pennings, 1982). Others, continuing lines of thought initiated by Hartmann or Burns and Stalker, focused on the problem of intrapreneurship and organizational change (e.g. Burgelman, 1983; Hambrick & Schecter, 1983). To some extent, these lines of research converged in work that rediscovered Stinchcombe’s (1965) early statement on the propensity of organizations to retain the structural features adopted by entrepreneurs at their time of founding. For instance, Boeker (1989) found that the functional background of entrepreneurs who had founded semiconductor firms many years ago continued to affect the contemporary staffing and top-management structure of those organizations.

Around the same time, work in mainstream sociology began to show a renewed appreciation of the Weberian legacy of entrepreneurship research, especially with respect to comparative/historical scholarship (Collins, 1980). In this context, Eisenstadt (1980) and DiMaggio (1988) separately coined the term “institutional entrepreneurs” to refer to elites who were in a position to influence the culture and institutional structure of societal sectors or even entire societies. In developing this concept, DiMaggio drew on his earlier empirical work, which had explained the creation of high culture – with particular reference to art museums and symphony

orchestras – in the United States during the late 19th century (DiMaggio, 1982a, 1982b). His emphasis on societal (rather than narrowly organizational) effects of entrepreneurship remains a prominent feature of sociological scholarship.

Through the following decade, the sociology of entrepreneurship continued to exhibit a highly eclectic range of topics. However, sustained research programs began to emerge in a number of key areas, which we highlight as central themes in this volume. First, many sociologists recognized that mobility into and out of entrepreneurship was an important feature of stratification in modern societies (Granovetter, 1984), including those societies that were just beginning their transition to capitalism (Rona-Tas, 1994; Nee, 1996; see also Inkeles & Smith, 1974). Hout (1984) highlighted the intergenerational transmission of entrepreneurial attitudes and organizing skills as being of special relevance to stratification processes. Second, a number of scholars extended the emphasis on enclaves of immigrant entrepreneurs, as pioneered by Lee in the 1940s and Light in the 1970s (see Light, 2005, for a useful review). Modern empirical debates in the area often centered on the question as to whether the segregation and solidarity entailed by enclaves generated economic benefits for immigrants and whether these benefits differed between entrepreneurs and non-entrepreneurs (Sanders & Nee, 1987; Portes & Jensen, 1989; Portes & Zhou, 1996). Third, organizational and economic sociologists began to display considerable interest in the impact of interorganizational networks on entrepreneurial activity, including entrepreneurs' ties to suppliers and distributors (Uzzi, 1997), banks (Uzzi, 1999), venture capitalists (Stuart, Hoang, & Hybels, 1999), and academia (Louis, Blumenthal, Gluck, & Stoto, 1989). The university–industry interface emerged as a particularly promising locus for scholars interested in high-tech entrepreneurship (Owen-Smith, Riccaboni, Pammolli, & Powell, 2002). Finally, a number of entrepreneurship scholars sought to transcend developments at the organizational level to understand the ecological, institutional, and social movement processes affecting the emergence of new *forms* of organizations (Romanelli, 1991; Ruef, 2000). Examples in the literature highlighted forms that were initially perceived as deviant cases, such as pro-choice groups (Staggenborg, 1988) and consumer watchdog organizations (Rao, 1998), but which were transformed by entrepreneurs to adapt to – and influence – the cognitive and normative expectations of their audiences. Such cultural entrepreneurship has become a key theme in research on institutional change and how entrepreneurs mobilize resources (e.g., Lounsbury & Glynn, 2001).

## IS ENTREPRENEURSHIP A MARGINAL SUBFIELD?

Given our brief history, how should entrepreneurship studies be viewed in broader intellectual context? Despite the trend toward an increasing prevalence of research on entrepreneurship in American sociology, there is also considerable quantitative and qualitative evidence that suggests it remains a parochial – and, perhaps, even marginal – area of scholarly interest. With respect to the issue of parochialism, the attention of European sociologists, serves as an instructive international contrast to the American case. As seen in [Table 2](#) (column 3), very little sociological scholarship in Europe has explicitly been coded as concerning “entrepreneurs”. We could not find any instances in the *British Journal of Sociology* before 1973, when [Javillonar and Peters](#) (incidentally, both American sociologists) published an article on entrepreneurs in Northern India. The more recent British scholarship treats the term “entrepreneur” with some suspicion (e.g. [Hickox, 1995](#)); and the *European Sociological Review* has yet to publish any article that refers to entrepreneurial processes in its title or abstract. European management journals that tackle sociological topics are fairly recent in their vintage and, thus, less amenable to historical analysis. But even these publications exhibit fairly limited explicit coverage of entrepreneurship by the standard shown in [Table 2](#) – for instance, the EGOS journal *Organization Studies* only published half-a-dozen articles with explicit reference to entrepreneurship in their title or abstract between 1980 and 1999.

Further evidence for the parochial status of the entrepreneurship subfield can be gleaned through a critical examination of the article counts in [Table 2](#). At first glance, the nearly 70 articles published in the top management and American sociology journals appear to represent a respectable total. But a more detailed examination suggests that entrepreneurship is a minor topic in a significant number of the articles included in this total, often appearing only as a broad metaphor for instances of organizational or societal change. More importantly, many of the articles that could topically be identified with the entrepreneurship area invoke alternative terminology, mentioning “founders” rather than “entrepreneurs”, for example; and “founding” or “economic development” rather than “entrepreneurial” processes (see last three columns of [Table 2](#)). While a small subset of these articles cover issues that are genuinely distinct from entrepreneurship (e.g. references to “founding” figures in sociology), many of the remaining number offer significant contributions to the sociology of entrepreneurship, especially those written from an ecological or institutional perspective (e.g.



Hannan & Freeman, 1987; Baum & Oliver, 1992). Notably, very few of the articles that discuss founding processes also include mentions of entrepreneurship in their titles or abstracts.

In recent years, the bulk of work on entrepreneurship that can broadly be defined as being sociological has appeared in specialty journals, including the *Journal of Business Venturing* (JBV), *Journal of Small Business Management* (JSBM), *Small Business Economics* (SBE), *Entrepreneurship and Regional Development* (E&RD), and *Entrepreneurship: Theory and Practice* (ET&P). For instance, an ABI search reveals that ET&P has featured 95 articles that make reference to “sociology”, since the inception of the journal in 1988. When one considers the proliferation of academic entrepreneurship journals – Katz (2003) identified nearly 50 – this suggests that there may be a fairly large amount of recent scholarship that is pertinent to the sociology of entrepreneurship. Unfortunately, quantity cannot be equated with quality. The sheer numbers of specialized journals contribute to the publication of many papers that do not meet the highest standards of methodological or conceptual rigor. The marginal status of such work is reflected in the meager citation impact of the specialty journals. Only four – JBV, JSBM, SBE, and E&RD – are listed in the Social Science Citation Index (SSCI) and their impact ratings (all under 0.6 in 2000) are low compared to those of the top American management and sociology journals (above 2.3 in 2000 for all those listed in Table 2).

What accounts for the limited impact of scholarship on entrepreneurs, considered in the context of American sociology and management theory, as well as the international field of sociology? While concrete instances of marginalization remain elusive, two mechanisms might suggest why entrepreneurship research lacks a mainstream presence in the discipline: (a) *theoretical dissension* concerning the definition and dynamics of entrepreneurship; and (b) *ideological opposition* to the study of entrepreneurial phenomena. Below, we argue that there is only weak evidence for the first of these mechanisms and very little for the latter.

### *Defining Entrepreneurship*

“Entrepreneur” is a vague and often elusive concept. According to Webster’s, the etymology of the term links it to the French verb “entreprendre” (to undertake) and leads to its common definition as someone “who organizes, manages, and assumes the risks of a business or enterprise”. While the concept was originated by the French political economists Cantillon and

Say, John Stuart Mill lamented the absence of an Anglo-American equivalent and helped introduce it into the English language.<sup>6</sup> Consistent with the term's French origins, Mill (1994[1848]) and a number of later economists (e.g. Knight 1921) emphasized the dual role of risk-taking and enterprise management as central features of entrepreneurship.

Not surprisingly, given its emphasis on rational, risk-calculating individuals, Mill's characterization of the entrepreneur has made only limited inroads among sociologists (see Xu & Ruef (2004) for an empirical critique). At least five other conceptualizations of entrepreneurship can be identified in the sociological literature.<sup>7</sup> First, as noted in Table 2, sociologists have maintained a sustained emphasis on entrepreneurship as a process of organizational founding (Carroll & Khessina, 2005). In an effort to focus on environmental conditions of founding, much of this literature makes only limited reference to individual entrepreneurs. Second, an even older literature on economic development (e.g. Hoselitz, 1952, 1960; Landes, 1965[1951]; Lipset, 1967) draws on the work of Parsons and McLelland to argue that entrepreneurship could be studied through regional variations in values, culture, and economic growth. This emphasis on culture and personality was advanced especially by Leland Jenks (e.g., 1965[1950]) through his affiliation with the Research Center on Entrepreneurial History at Harvard University, founded with a Rockefeller Foundation grant in 1948. With the aggregation of entrepreneurial outcomes to the regional or national level, the individual often fades into the background, perhaps even more so than in accounts of organizational founding processes (but see Inkeles & Smith (1974) for an important exception).

Two other conceptions of entrepreneurship – which might be designated the Schumpeterian and Weberian definitions – have already been discussed. The Schumpeterian definition focuses on the innovative capacity of the entrepreneur, while the Weberian version emphasizes the entrepreneurial role as a counterbalance to managerial bureaucracy (see also Burnham, 1941; Hartmann, 1959; Swedberg, 2005, pp. 87–88). Contemporary empirical work that relies on these definitions tends to be more micro-oriented than those that emphasize organizational founding or economic development. For instance, Ruef (2002) uses a Schumpeterian perspective to predict business innovation based on the interpersonal networks and enculturation of entrepreneurs. Lounsbury (2001) deploys a Weberian framing in distinguishing between status creation, where a new role is constructed from scratch by organizational entrepreneurs, and role accretion, where roles gradually accumulate rights and obligations as a function of bureaucratic development.

A final sociological conception of entrepreneurship jettisons both the methodological individualism of Mill's entrepreneur and his restricted emphasis on organizational and economic processes. As described by DiMaggio (1988), *institutional entrepreneurs* are self-interested and resource-endowed agents with the capacity to develop new social institutions. Inspired by the early work of DiMaggio and Bourdieu, empirical studies of institutional entrepreneurship have examined a wide range of outcomes (see review in Battilana, 2006). These include the adoption of new technological standards (Munir & Phillips, 2005), regulatory and normative changes in organizational fields (Scott, Ruef, Mendel, & Caronna, 2000), the establishment of new industries (Lounsbury, Ventresca, & Hirsch, 2003), and the emergence of the EU common market (Fligstein, 2001). A unifying theme in these studies is an emphasis on the cultural products of entrepreneurship, conceived as new concepts, meanings, cognitive frames, rules, and the like.

Considering the diversity exhibited by these five sociological conceptions of entrepreneurship, the lack of cohesiveness of the entrepreneurship subfield may seem unsurprising, and even expected. For example, in a citation analysis of the *Journal of Business Venturing* over the period of a year, Carroll and Khessina (2005) find that only 2.7% of all references are articles or books on organizational ecology, which emphasizes startup activity as an industry-level rather than individual-level phenomenon. More generally, our own qualitative impressions of the articles identified in Table 2 suggest limited patterns of co-citation among articles that emphasize either different theoretical *or* topical features of entrepreneurship. Given deep divisions in the level of analysis employed (individual, organization, industry, and region), the types of outcomes considered (e.g. material versus cultural), and underlying philosophical assumptions about human agency, one might well ask how a more unified sociology of entrepreneurship might be developed. We return to this issue shortly.

For the time being, however, it is worth emphasizing that the subfield of entrepreneurship is hardly alone in its lack of cohesiveness. The general balkanization of organizational sociology has long been recognized by both American (Pfeffer, 1993) and European (Burrell & Morgan, 1979) scholars. What is instructive about this parallel is that the proliferation of different theoretical conceptions of organizations has *not* led to the marginalization of this field of scholarship; to the contrary, organizational sociology remains one of the most vibrant and influential sections in the discipline. Consequently, the link between a lack of cohesiveness within the entrepreneurship field and its own marginalization appears somewhat tenuous. An alternate explanation for this outcome turns to external factors, considering the place of the sociology of entrepreneurship within the politics of the discipline.

*Ideology and the Sociology of Entrepreneurship*

The argument for ideological opposition to research on entrepreneurship hinges on the assumption that many sociologists have historically viewed such work as validating a paradigm of self-sufficient individualism, to the neglect of structural constraint or heritable skills and resources. Empirical evidence for such ideological opposition is hard to find in written form. One source we considered were book critiques that sociologists have written for monographs and edited volumes on entrepreneurship, including those undertaken by scholars outside the discipline. During the early postwar years, William Miller's (1952) edited volume, *Men in Business: Essays in the History of Entrepreneurship*, would have seemed a suitable lightning rod for sociological critique. Still, it was reviewed favorably by both C. Wright Mills and Everett C. Hughes. Mills (1952, p. 505), never one to shy from controversy, concluded that the "work displayed in this volume is everything that one could now expect of historical work on the sociology of business". Hughes acknowledged that, in the past, "there has been a tendency to obscure, by the opposition of enterprise and socialism, the problems in this field", but went on to stress the importance of sociological scholarship on entrepreneurship: "the enterprising [i.e. entrepreneurial] way of working is a thing to be understood, as in distinction from the ritualistic or bureaucratic, from the resistant and unwilling way ..." (1952, p. 517).

Arguably, the 1950s provide a less-than-ideal period for capturing opposition to entrepreneurship research. By the 1960s and 1970s, sociologists had a greater intellectual understanding of management ideologies, including the "entrepreneurial ideology" that emerged in Weber's work on the *Protestant Ethic* (Winter, 1974). Still, there is limited evidence that this translated into personal opposition to a sociology of entrepreneurship. *AJS* book reviews from the period were often glowing, with special praise for David McLelland's (1961) influential study of need-based achievement and economic development (Hoselitz, 1962). Evans (1971) offered a more critical perspective on a study of nearly 100 Midwestern independent entrepreneurs, which was largely aimed at a nonacademic audience. Nevertheless, his critique was entirely analytical, emphasizing the authors' failure to engage the extant literature on entrepreneurship more seriously, rather than the evident ideological biases of the monograph being reviewed. Even in an era of popular protest movements and misgivings about capitalism, sociologists saw value in "looking at the way in which small business and the entrepreneurial role fits into contemporary social structures" (Evans, 1971, p. 618).

## CONTEMPORARY RESEARCH ON ENTREPRENEURSHIP

Contemporary scholarship on the sociology of entrepreneurship is sufficiently eclectic and voluminous as to defy simple categorization. Rather than attempt a broad survey of the recent literature (see Aldrich & Ruef [2006] for an organizational perspective), we devote this section to the more modest task of presenting a neo-Weberian schema that helps to highlight sociological contributions at different levels of analysis and with different substantive foci. We then introduce the empirical articles found in this volume as exemplars of contemporary research streams and place them within the schematic framework.

The schema begins with four sociological perspectives on entrepreneurial behavior that were hinted at in Weber's work: (a) the *contextual* perspective, which emphasizes the role of the material and cultural environment in influencing individual orientations toward entrepreneurship; (b) the *behavioral* perspective, which examines the structure and process of entrepreneurial activity at a micro-level; (c) the *constructivist* perspective, which addresses the implications that entrepreneurial decisions have for an organizational startup, an industry, a community, or society as a whole; and (d) the *ecological* perspective, which analyzes the direct impact of the material and cultural environment on economic and institutional development, considered apart from the actions of individual entrepreneurs. Applied across levels of analysis, these perspectives yield the familiar macro-micro-macro linkages introduced by James Coleman (1990, Chapter 1) in his discussion of Weber's *Protestant Ethic*. For instance, a contextual perspective might consider the impact of Protestant religious doctrine and culture on the tendency of individuals to acquire values oriented toward calculation, predictability, efficiency, and self-control (macro-micro link). In turn, a behavioral perspective would analyze how these values translate into economic activity that is entrepreneurial in character, such as breaking rank with traditional guild organizations or founding mercantile partnerships (micro-micro link). A constructivist perspective could then address how such individual activities, in the aggregate, yield a systemic transformation in society, from feudal to capitalist organization (micro-macro link). Finally, scholars employing an ecological lens might argue that the reference to individual entrepreneurs in this account is superfluous; instead, Weber's thesis could be studied more directly as a correlation between the historical predominance of Protestantism within a society and its level of capitalist development (macro-macro link) (see Delacroix & Nielsen, 2001, for a critique).

A simple schema for the sociological analysis of entrepreneurship emerges if we differentiate the micro- versus macro-level distinction further into multiple levels of analysis and separate the material aspects of entrepreneurial activity from its cognitive and institutional dimensions (see Fig. 2). Within this rubric, a contextual perspective addresses the social antecedents of entrepreneurship in the left-hand side of the figure, with special emphasis on the question of who becomes an entrepreneur and why. A behavioral perspective follows the process of entrepreneurship highlighted at the bottom of the figure, answering questions concerning the social psychological basis of entrepreneurship, the timing of entrepreneurial activities, the formation of entrepreneurial teams, the definition of task and membership boundaries, and the like. A constructivist perspective follows the upward arrow in the right-hand side of the figure, attending to the effects of entrepreneurship on interpersonal networks, organizational populations, and communities, as well as the startup venture itself. An ecological perspective

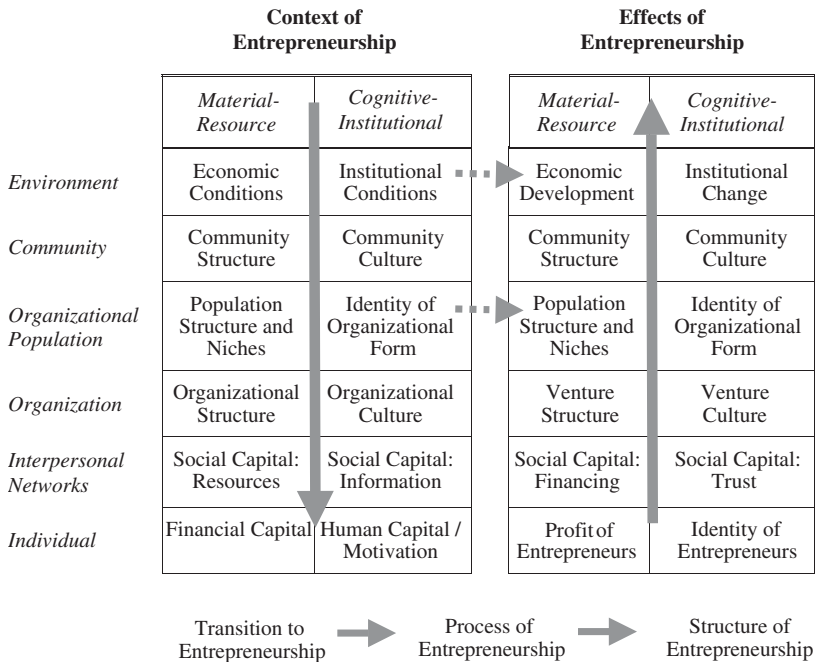


Fig. 2. A Multilevel Framework for the Sociology of Entrepreneurship.

may propose a causal association between the left- and right-hand sides of the figure at any level beyond the individual. As we noted in our review of definitions of entrepreneurship, a common tack in sociological theories of modernization has been to conceptualize entrepreneurship in terms of economic development at the highest level of aggregation. A second ecological approach, found in organizational ecology, is to explain organizational founding rates (aggregated at the population level) as a direct function of the size and distribution of organizational populations.

The papers represented in this volume exemplify the wide array of levels and causal mechanisms that can be invoked in the sociology of entrepreneurship. The articles in Part I employ a contextual perspective to understand entry into entrepreneurial activity. Aldrich and Kim couch their argument at the level of individual entrepreneurs and their distinct life course stages, addressing whether entrepreneurial parents pass on privilege – i.e. genetic inheritances, human capital, financial resources, social capital, and motivation – to their children. Surveying a range of empirical studies, as well as preliminary findings from the Panel Study of Entrepreneurial Dynamics (PSED), they argue that there is strong evidence for genetic inheritance and value socialization during childhood, moderate evidence for adolescent reinforcement of socialization patterns, and weak evidence for parental transfers of financial resources or social capital to adult entrepreneurs. Sorensen offers a complementary analysis of entry into self-employment, but draws instead on new Danish data set. Consistent with our distinction between material-resource and cognitive dimensions of entrepreneurship, he differentiates two mechanisms of intergenerational transmission: occupational closure, which emphasizes the propensity of parents to pass on tangible financial and social capital, and occupational exposure, which addresses the impact of parental socialization on aspirations and skills. Like Aldrich and Kim, he finds that there is little evidence for occupational closure, but considerable support for the intuition that exposure to entrepreneurial parental role models before adulthood can drive entry into self-employment.

While a contextual perspective is perhaps the most common framework for understanding entry into entrepreneurship, the topic can also be pursued from a behavioral approach. Xu and Ruef take the existence of a startup and focal entrepreneur as given, and then ask how members of that focal entrepreneur's support network are sorted into organizational "insider" roles (owner-managers) and "outsider" roles (passive investors and advisors). The question of transition into entrepreneurial activity is thus re-framed as a behavioral process of drawing membership boundaries within

an emergent organization. Analyzing the PSED data set, they find that the cognitive aspects of boundary formation, which emphasizes trust among startup participants, are generally more salient than material-resource aspects, which emphasize considerations such as asset specificity and non-redundancy of participant contributions.

Part II of the volume turns to the topics of immigrant entrepreneurship and enclaves. Portes and Shafer revisit a path-breaking study of Cuban immigrants 20 years later to understand changing patterns of ethnic enterprise in Miami. Contrary to common intuitions among labor economists, who often argue that ethnic enclaves represent mobility “traps” for immigrants, they find that self-employed Cubans in the Miami enclave enjoy considerable benefits from entrepreneurial activity, provided that they can tap into the extensive social networks and legitimacy of the pre-Mariel generation of Cuban immigrants. Portes and Shafer conclude that most of the previous arguments against upward mobility within ethnic enclaves have rested on faulty measures of the enclave concept or inadequate understanding of enclave history. In a parallel paper on ethnic commercial precincts, Pang and Rath employ a “multiple embeddedness” perspective to understand the effects of immigrant entrepreneurship, tracing entrepreneurial agency back to the community level. While much of the literature has focused on ethnic networks and culture, they argue that this focus must be balanced with attention to the regulatory environment. Considering Washington DC’s Chinatown as an illustrative case, Pang and Rath show how ethnic enclaves can persist – at least, symbolically – even when they are challenged by the encroachment of nonethnic enterprises.

The studies in the first two parts of the volume address both the micro-level context and process of entrepreneurship, as well as some macro-level consequences. Within these perspectives, formal organizations make a somewhat tentative appearance, either as emergent entities or as part of the infrastructural background of a community. In the next two parts of the volume, organizations assume center stage. Part III emphasizes academic entrepreneurship within specific organizational settings. Colyvas and Powell offer a detailed historical study of the emergence and spread of biomedical entrepreneurship at Stanford University between 1970 and 2000. They demonstrate that the engagement of faculty members in invention disclosure is not simply a matter of individual predilection, but depends to a considerable extent on the structure and content of organizational networks linking scientists into research clusters. The article by Evans provides a complementary constructivist perspective, showing how academics in molecular biology plant laboratories opened a new field of entrepreneurial



activity. Analyzing all scientific publications on *Arabidopsis thaliana*, a plant whose genetic structure has extensive commercial implications, Evans argues that the novelty of these scientific contributions was tied to the funding diversity of the laboratories where they were produced. Like Colyvas and Powell, he makes a case for the theoretical and methodological importance of studying entrepreneurs outside of traditional business organizations.

The last part of the volume addresses explicit entrepreneurial efforts to construct and defend new organizational forms. Caronna studies how the founders and members of Kaiser Permanente, an early prepaid group practice in the United States, sought to legitimate the organizational form that would come to be known as a health maintenance organization. She demonstrates how common concepts in the sociology of entrepreneurship – including those pertaining to individual entrepreneurs, organizational identities, and human capital – can be recast at a meso-level to understand the process whereby an organization “founds” a new organizational form. Following a detailed case study that applies these concepts to Kaiser Permanente, from its origins in the 1930s through the present, Caronna concludes that shifting the study of entrepreneurship upward in level of analysis offers both promises and pitfalls over individualist (i.e. behavioral) perspectives. Continuing this theme, Solari engages the literature on the legitimacy of new organizational forms, with specific attention to the informative case of gray and black market organizations. Noting that the existing literature on organizational legitimacy has overwhelmingly emphasized forms that are widely viewed as legal and appropriate, Solari calls our attention to a black market form (the mafia), a gray market form (“sexy” shops in Italy), and a form that has traversed the spectrum from black to white markets (Samba schools in Brazil). In all three cases, he finds that entrepreneurs could not draw on the general social legitimacy accorded to traditional businesses and, instead, negotiated a difficult path between endorsing audiences and opponents.

## CONCLUSION

Taken as a whole, the articles in this volume highlight how the sociological imagination emphasizes the need for a richly nuanced, multilevel perspective on entrepreneurship. Unlike economic and psychological approaches to the phenomenon, sociological analyses suggest that entrepreneurship cannot be adequately understood outside of its socio-cultural context. To make further progress in creating a unified framework for a sociology of entrepreneurship

subfield, we believe that it would be useful to acknowledge the Weberian foundations of much of the research to date and explicitly build upon those foundations to create a shared theoretical emphasis.

Following the implications of the chapters in this volume, three lines of inquiry may be especially fruitful. First, sociological studies can follow Weber's lead in attending to the social and institutional environment of entrepreneurship. Whether the environment is defined at a micro-level of analysis – as in Sorenson, Aldrich, Kim, and Xu's studies of parental and network influences on entry into entrepreneurial activity – or at a more macro-level – as in Portes, Shafer, Pang, and Rath's studies of ethnic enclaves, an emphasis on context leads sociologists away from the individualist emphases of other disciplines. Sociological studies of entrepreneurship can also reveal how the institutional environment is constructed in the first place, through the lens, for instance, of academic entrepreneurship (Colyas, Powell, and Evans) or the literature on the emergence of organizational forms (Caronna and Solari).

Second, sociological scholarship can draw attention to tensions between material and nonmaterial influences on entrepreneurial activity or the interplay of these factors. As revealed by this volume's chapters on entry into entrepreneurship, the predominant economic concern with material resources seems overdrawn and should be complemented by attention to factors such as social networks, human and social capital, trust, and even genetic inheritances. At the same time, sociologists should remain open to the possibility that the structure of material resources may affect the nature and novelty of nonmaterial entrepreneurial outcomes (Evans); and that some symbolic communities of entrepreneurs may only persist with the material and regulatory support of outsiders (Pang and Rath).

A third line of inquiry follows from an emphasis on entrepreneurship outside of a "traditional" business context. Examples in this volume are drawn from ethnic enterprise, academic entrepreneurship, and form development in healthcare. These cases demonstrate that mechanisms of organizational creation in these sectors share a number of features in common with traditional industry. At the same time, sociologists must confront important questions concerning the legitimacy of entrepreneurial activity outside of the business mainstream, especially when such activity falls under the auspices of the informal economy (Solari). As Weber recognized over a century ago, entrepreneurial activity undertaken in the absence of institutional support may generate severe liabilities for the organizational actors that emerge in the process. Of course, such "illicit" entrepreneurship also represents one of the most profound possibilities for institutional change.

## NOTES

1. It must be acknowledged, of course, that sociologists do not have a monopoly on a perspective that addresses contextual, nonmaterial, and nonmarket aspects of entrepreneurship. To name just two prominent counterexamples, consider the work of Annalee Saxenian (1994, 2006), an urban geographer, on regional advantage; or the interest displayed by institutional economists (e.g. Nelson, 1993) in national innovation systems.

2. Recent efforts at cross-national data collection, such as the Global Entrepreneurship Monitor (GEM) project initiated in 1999, may serve to mitigate the geographic fragmentation of scholarly work, though their sociological impact is not yet clear.

3. Notably, one major inspiration for the *Protestant Ethic* was an empirical study of the relationship between occupational stratification and religious denomination, conducted by his student, Martin Offenbacher. Examining statistics for the German state of Baden around 1900, the study revealed that a disproportionate number of Protestants owned capital and were involved in entrepreneurial activity (Käsler, 1988, p. 75).

4. Not surprisingly, the strongest imprint of this Weberian legacy can often be found in contemporary work on the sociology of law and entrepreneurial organizations (e.g. Suchman, Steward, & Westfall, 2001).

5. Naturally, medieval entrepreneurial organizations developed on this basis (such as the *societas maris*) did not have many of the features that we expect to see in modern corporate forms. They were intended to be run for particular mercantile and trading purposes, not as continuous enterprises. Moreover, they lacked the legal elements of joint liability (see Table 2). For a recent analysis of the origins of a more durable partnership form, see Padgett and McLean (2006).

6. Despite Mill's serious appropriation of the term "entrepreneur", it also had a somewhat light-hearted connotation in late-Victorian Britain, as someone who organizes "entertainments" or directs a "musical institution" (Cole, 1999).

7. By no means these are intended to be exhaustive. See Hoselitz (1952) for an early typology of definitions and Aldrich and Ruef (2006, Chapter 4) for a recent treatment of the entrepreneurship literature.

## REFERENCES

- Aldrich, H. (1979). *Organizations and environments*. Englewood Cliffs, NJ: Prentice-Hall.
- Aldrich, H. (2005). Entrepreneurship. In: N. Smelser & R. Swedberg (Eds), *Handbook of economic sociology* (pp. 451–477). Princeton, NJ: Princeton University Press.
- Aldrich, H., & Ruef, M. (2006). *Organizations evolving* (2nd ed.). London: Sage.
- Battilana, J. (2006). Agency and institutions: The enabling role of individuals' social position. *Organization*, 13, 653–676.
- Baum, J., & Oliver, C. (1992). Institutional embeddedness and the dynamics of organizational populations. *American Sociological Review*, 57, 540–559.
- Becker, S., & Gordon, G. (1966). An entrepreneurial theory of formal organizations, part I: Patterns of formal organizations. *Administrative Science Quarterly*, 11, 315–344.
- Berle, A., & Means, G. (1968). *The modern corporation and private property*. New York: Harcourt, Brace and World.

- Boeker, W. (1989). The development and institutionalization of subunit power in organizations. *Administrative Science Quarterly*, 34, 388–410.
- Bottomore, T. (Ed.) (1991). *A dictionary of Marxist thought* (2nd ed.). Oxford: Blackwell.
- Brockhaus, R. (1980). Risk taking propensity of entrepreneurs. *Academy of Management Journal*, 23, 509–520.
- Burgelman, R. (1983). A process model of internal corporate venturing in the diversified major firm. *Administrative Science Quarterly*, 28, 223–244.
- Burnham, J. (1941). *The managerial revolution*. New York: John Day.
- Burns, T., & Stalker, G. (1961). *The management of innovation*. London: Tavistock.
- Burrell, G., & Morgan, G. (1979). *Sociological paradigms and organizational analysis*. London: Heinemann.
- Carroll, G., & Khessina, O. (2005). The ecology of entrepreneurship. In: S. A. Alvarez, R. Agarwal & O. Sorenson (Eds), *Handbook of entrepreneurship research: Disciplinary perspectives*. New York: Springer.
- Cole, A. (1999). *Business enterprise in its social setting*. Lincoln, NE: iUniverse (originally published in 1959).
- Coleman, J. (1974). *Power and the structure of society*. New York: W.W. Norton.
- Coleman, J. (1990). *Foundations of social theory*. Cambridge, MA: Harvard University Press.
- Collins, R. (1980). Weber's last theory of capitalism: A systematization. *American Sociological Review*, 45, 925–942.
- Davis, S. (1968). Entrepreneurial succession. *Administrative Science Quarterly*, 13, 402–416.
- Delacroix, J., & Nielsen, F. (2001). The beloved myth: Protestantism and the rise of industrial capitalism in nineteenth-century Europe. *Social Forces*, 80, 509–553.
- DiMaggio, P. (1982a). Cultural entrepreneurship in nineteenth-century Boston, part I: The creation of an organizational base for high culture in America. *Media, Culture and Society*, 4, 33–50.
- DiMaggio, P. (1982b). Cultural entrepreneurship in nineteenth-century Boston, part II: The classification and framing of American art. *Media, Culture and Society*, 4, 303–322.
- DiMaggio, P. (1988). Interest and agency in institutional theory. In: L. Zucker (Ed.), *Institutional patterns and organizations: Culture and environment* (pp. 3–22). Cambridge, MA: Ballinger.
- Durkheim, E. (1984). *The division of labor*. New York: Free Press (translated by W. D. Halls, originally published in 1893).
- Eisenstadt, S. N. (1980). Cultural orientations, institutional entrepreneurs, and social change: comparative analysis of traditional civilizations. *American Journal of Sociology*, 85, 840–869.
- Evans, P. (1971). Review of *The Organization Makers*. *American Journal of Sociology*, 77, 617–619.
- Fligstein, N. (2001). Institutional entrepreneurs and cultural frames: The case of the European union's single market program. *European Societies*, 3, 261–287.
- Granovetter, M. (1984). Small is bountiful: Labor markets and establishment size. *American Sociological Review*, 49, 323–334.
- Hambrick, D., & Schecter, S. (1983). Turnaround strategies for mature industrial-product business units. *Academy of Management Journal*, 26, 231–248.
- Hannan, M., & Freeman, J. (1987). The ecology of organizational founding: American labor unions, 1836–1985. *American Journal of Sociology*, 92, 910–943.
- Hartmann, H. (1959). Managers and entrepreneurs: A useful distinction? *Administrative Science Quarterly*, 3, 429–451.

- Hickox, M. S. (1995). The English middle-class debate. *British Journal of Sociology*, 46, 311–324.
- Hirsch, P. (1972). Processing fads and fashions: An organization-set analysis of cultural industry systems. *American Journal of Sociology*, 77, 639–659.
- Hoselitz, B. (1952). Entrepreneurship and economic growth. *American Journal of Economic Sociology*, 12, 97–106.
- Hoselitz, B. (1960). *Sociological aspects of economic growth*. Glencoe, IL: Free Press.
- Hoselitz, B. (1962). Review of *The Achieving Society*. *American Journal of Sociology*, 68, 129–130.
- Hout, M. (1984). Status, autonomy, and training in occupational mobility. *American Journal of Sociology*, 89, 1379–1409.
- Hughes, E. C. (1934). Review of *Zur Sociologie des Unternehmertums*. *American Journal of Sociology*, 39, 548–549.
- Hughes, E. C. (1952). Review of *Men in Business*. *American Journal of Sociology*, 57, 516–517.
- Hwang, H., & Powell, W. (2005). Institutions and entrepreneurship. In: S. A. Alvarez, R. Agarwal & O. Sorenson (Eds), *Handbook of entrepreneurship research: Disciplinary perspectives* (pp. 179–210). New York: Springer.
- Inkeles, A., & Smith, D. H. (1974). *Becoming modern: individual change in six developing countries*. Cambridge, MA: Harvard University Press.
- Javillonar, G., & Peters, G. (1973). Sociological and social psychological aspects of Indian entrepreneurship. *British Journal of Sociology*, 24, 314–328.
- Jenks, L. (1965[1960]). Approaches to entrepreneurial personality. In: H. Aitken (Ed.), *Explorations in enterprise* (pp. 80–92). Cambridge, MA: Harvard University Press.
- Käsler, D. (1988). *Max Weber: An introduction to his life and work*. Chicago: University of Chicago Press.
- Katz, J. (2003). *Core publications in entrepreneurship and related fields: A guide to getting published*, Accessed January 4, 2006, URL: <http://eweb.slu.edu/booklist.html>
- Keister, L. (Ed.) (2005). *Research in the Sociology of work: Entrepreneurship* (Vol. 15). Oxford: Elsevier.
- Kim, P. H., Aldrich, H. E., & Keister, L. A. (2004). *Access (not) denied: The impact of financial, human capital, and cultural capital on becoming a nascent entrepreneur*. Working Paper, University of North Carolina at Chapel Hill.
- Knight, F. (1921). *Risk, uncertainty, and profit*. New York: Houghton Mifflin.
- Landes, D. (1965[1951]). French business and the businessman: A social and cultural analysis. In: H. Aitken (Ed.), *Explorations in enterprise* (pp. 184–200). Cambridge, MA: Harvard University Press.
- Lee, R. H. (1949). Occupational invasion, succession, and accommodation of the Chinese of Butte, Montana. *American Journal of Sociology*, 55, 50–58.
- Light, I. (2005). The ethnic economy. In: N. Smelser & R. Swedberg (Eds), *Handbook of economic sociology* (pp. 650–677). Princeton, NJ: Princeton University Press.
- Lippmann, S., Davis, A., Aldrich, H. (2005). Inequality and entrepreneurship. In: L. Keister (Ed.), *Research in the sociology of work: Entrepreneurship* (Vol. 14, pp. 3–31). Oxford: Elsevier.
- Lipset, S. M. (1967). Values, education and entrepreneurship. In: S. M. Lipset & A. Solari (Eds), *Elites in Latin America* (pp. 3–60). New York: Oxford University Press.
- Litzinger, W. (1963). Entrepreneurial prototype in bank management: A comparative study of branch bank managers. *Academy of Management Journal*, 6, 36–45.

- Louis, K., Blumenthal, D., Gluck, M., & Stoto, M. (1989). Entrepreneurs in academe: An exploration of behaviors among life scientists. *Administrative Science Quarterly*, *34*, 110–131.
- Lounsbury, M. (2001). Institutional sources of practice variation: Staffing college and university recycling programs. *Administrative Science Quarterly*, *46*, 29–56.
- Lounsbury, M., & Glynn, M. A. (2001). Cultural entrepreneurship: Stories, legitimacy and the acquisition of resources. *Strategic Management Journal*, *22*, 545–564.
- Lounsbury, M., Ventresca, M., & Hirsch, P. (2003). Social movements, field frames and industry emergence: A cultural-political perspective on U.S. recycling. *Socio-Economic Review*, *1*, 71–104.
- March, J. (2004). Parochialism in the evolution of a research community: The case of organization studies. *Management and Organization Review*, *1*, 5–22.
- McLelland, D. (1961). *The achieving society*. Princeton, NJ: D. Van Nostrand.
- Meyer, J., & Jepperson, R. (2000). The actors of modern society: The cultural construction of social agency. *Sociological Theory*, *18*, 100–120.
- Mill, J. S. (1994). *Principles of political economy*. Oxford: Oxford University Press (originally published in 1848).
- Miller, W. (1952). *Men in business: Essays in the history of entrepreneurship*. Cambridge, MA: Harvard University Press.
- Mills, C. W. (1952). Review of men in business. *American Sociological Review*, *17*, 504–505.
- Mizruchi, M. (2004). Berle and Means revisited: The governance and power of large U.S. corporations. *Theory and Society*, *33*, 579–617.
- Munir, K., & Phillips, N. (2005). The birth of the ‘Kodak Moment’: Institutional entrepreneurship and the adoption of new technologies. *Organization Studies*, *26*, 1665–1687.
- Nee, V. (1996). The emergence of a market society: Changing mechanisms of stratification in China. *American Journal of Sociology*, *101*, 908–949.
- Nelson, R. (Ed.) (1993). *National innovation systems: A comparative analysis*. Oxford: Oxford University Press.
- Ondrack, D. A. (1973). Emerging occupational values: A review and some findings. *Academy of Management Journal*, *16*, 423–432.
- Owen-Smith, J., Riccaboni, M., Pammolli, F., & Powell, W. W. (2002). A comparison of U.S. and European university–industry relations in the life sciences. *Management Science*, *48*, 24–43.
- Padgett, J., & McLean, P. (2006). Organizational invention and elite transformation: The birth of partnership systems in renaissance Florence. *American Journal of Sociology*, *111*, 1463–1568.
- Pennings, J. (1982). The urban quality of life and entrepreneurship. *Academy of Management Journal*, *25*, 63–79.
- Peterson, R., & Anand, N. (2004). The production of culture perspective. *Annual Review of Sociology*, *30*, 311–334.
- Peterson, R., & Berger, D. (1971). Entrepreneurship in organizations: Evidence from the popular music industry. *Administrative Science Quarterly*, *16*, 97–106.
- Pfeffer, J. (1993). Barriers to the advance of organizational science: Paradigm development as a dependent variable. *Academy of Management Review*, *18*, 599–620.
- Podolny, J. (2001). Networks as the pipes and prisms of the market. *American Journal of Sociology*, *107*, 33–60.

- Portes, A., & Jensen, L. (1989). The enclave and the entrants: Patterns of ethnic enterprise in Miami before and after Mariel. *American Sociological Review*, 54, 929–949.
- Portes, A., & Zhou, M. (1996). Self-employment and the earnings of immigrants. *American Sociological Review*, 61, 219–230.
- Rao, H. (1998). Caveat emptor: The construction of nonprofit consumer watchdog organizations. *American Journal of Sociology*, 103, 912–961.
- Romanelli, E. (1991). The evolution of new organizational forms. *Annual Review of Sociology*, 17, 79–103.
- Rona-Tas, A. (1994). The first shall be last? Entrepreneurship and communist cadres in the transition from socialism. *American Journal of Sociology*, 100, 40–69.
- Ruef, M. (2000). The emergence of organizational forms: A community ecology approach. *American Journal of Sociology*, 106, 658–714.
- Ruef, M. (2002). Strong ties, weak ties and islands: Structural and cultural predictors of organizational innovation. *Industrial and Corporate Change*, 11, 427–449.
- Sanders, J. M., & Nee, V. (1987). Limits of ethnic solidarity in the enclave economy. *American Sociological Review*, 52, 745–773.
- Sarapata, A., & Wesolowski, W. (1961). The evaluation of occupations by Warsaw inhabitants. *American Journal of Sociology*, 66, 581–591.
- Saxenian, A. (1994). *Regional advantage: Culture and competition in Silicon Valley and Route 128*. Cambridge, MA: Harvard University Press.
- Saxenian, A. (2006). *The new argonauts: Regional advantage in a global economy*. Cambridge, MA: Harvard University Press.
- Schumpeter, J. (1934). *The theory of economic development*. Cambridge, MA: Harvard Economic Studies (translated by R. Opie, originally published in 1911).
- Schwiedland, E. (1933). *Zur Sociologie des Unternehmertums*. Leipzig: C. L. Hirschfeld Verlag.
- Scott, W. R. (2003). *Organizations: Rational, natural, and open systems* (5th ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Scott, W. R., Ruef, M., Mendel, P., & Caronna, C. (2000). *Institutional change and healthcare organizations: From professional dominance to managed care*. Chicago: University of Chicago Press.
- Simmel, G. (1990[1907]). *The philosophy of money*. NY: Routledge.
- Sine, W., Mitsuhashi, H., & Kirsch, D. (2006). Revisiting Burns and Stalker: Formal structure and new venture performance in emerging economic sectors. *Academy of Management Journal*, 49, 121–132.
- Sorenson, O., & Stuart, T. (2001). Syndication networks and the spatial distribution of venture capital. *American Journal of Sociology*, 106, 1546–1588.
- Staggenborg, S. (1988). The consequences of professionalization and formalization in the pro-choice movement. *American Sociological Review*, 53, 585–605.
- Stinchcombe, A. (1965). Social structure and organizations. In: J. G. March (Ed.), *Handbook of organizations* (pp. 142–193). Chicago: Rand McNally.
- Stuart, T., Hoang, H., & Hybels, R. (1999). Interorganizational endorsements and the performance of entrepreneurial ventures. *Administrative Science Quarterly*, 44, 315–349.
- Suchman, M., Steward, D., & Westfall, C. (2001). The legal environment of entrepreneurship: Observations on the legitimization of venture finance in Silicon Valley. In: C. B. Schoonhoven & E. Romanelli (Eds), *The entrepreneurship dynamic: Origins of entrepreneurship and the evolution of industries* (pp. 349–382). Stanford, CA: Stanford University Press.

- Swedberg, R. (Ed.) (2000). *Entrepreneurship: The social science view*. Oxford: Oxford University Press.
- Swedberg, R. (2005). *The Max Weber dictionary: Key words and central concepts*. Stanford, CA: Stanford University Press.
- Thornton, P. (1999). The sociology of entrepreneurship. *Annual Review of Sociology*, 25, 19–46.
- Uzzi, B. (1997). Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative Science Quarterly*, 42, 35–67.
- Uzzi, B. (1999). Embeddedness in the making of financial capital: How social relations and networks benefit firms seeking financing. *American Sociological Review*, 64, 481–505.
- Weber, M. (1930). *The protestant ethic and the spirit of capitalism*. London: G. Allen and Unwin (translated by T. Parson, originally published in 1904–1905).
- Weber, M. (2003). *The history of commercial partnerships in the middle ages*. Lanham, MD: Rowman and Littlefield (translated by L. Kaelber, originally published in 1889).
- Winter, J. A. (1974). Elective affinities between religious beliefs and ideologies of management in two eras. *American Journal of Sociology*, 79, 1134–1150.
- Wolff, K. (1950). *The sociology of Georg Simmel*. New York: Free Press.
- Xu, H., & Ruef, M. (2004). The myth of the risk-tolerant entrepreneur. *Strategic Organization*, 2, 331–355.



# A LIFE COURSE PERSPECTIVE ON OCCUPATIONAL INHERITANCE: SELF-EMPLOYED PARENTS AND THEIR CHILDREN

Howard E. Aldrich and Phillip H. Kim

## ABSTRACT

*Using a life course perspective, we develop a theoretical model of how parents can influence their children's propensity to enter self-employment. We draw on the sociological, economic, psychological, and behavioral genetics literatures to develop a model in which parental influence occurs in different ways, depending on someone's stage in their life course. We review and summarize existing findings for parental influences on entrepreneurial entry using a three-part life course framework: childhood, adolescence, and adulthood. We also analyze new data from the Panel Study of Entrepreneurial Dynamics on the extent to which children were involved in their parents' businesses. From our review, we propose strong effects from genetic inheritances and parenting practice (during childhood); moderate effects from reinforcement of work values and vocational interests (during adolescence); and little influence from financial support but stronger effects from other tangible means of support (during adulthood).*

---

The Sociology of Entrepreneurship

Research in the Sociology of Organizations, Volume 25, 33–82

Copyright © 2007 by Elsevier Ltd.

All rights of reproduction in any form reserved

ISSN: 0733-558X/doi:10.1016/S0733-558X(06)25002-X

## INTRODUCTION

Since the pioneering studies of social mobility in the 1950s, sociologists have shown that children tend to choose occupations that are very similar to those of their parents. Self-employed professionals and farmers are especially likely to have children following in their footsteps, but that tendency has been muted by the disappearance of solo professional practices and family farms. The term “occupational inheritance” was coined as a general label for this tendency, although strictly speaking, most studies of the intergenerational transmission have been of occupational status, rather than of specific occupational choices. Some sociologists have focused on the intergenerational transmission of occupational status because of their interest in patterns of inequality and stratification in modern societies, whereas others have been more interested in what occupational inheritance tells us about the importance of human capital and wealth for occupational choices.<sup>1</sup>

In this paper, we focus on one occupational status, self-employment, rather than the more general issue of occupational inheritance. We are especially interested in exploring why the sons and daughters of self-employed parents have a heightened tendency to attempt entrepreneurship (and thus become self-employed) at some point during their working careers.<sup>2</sup> Cross-national studies of self-employment and entrepreneurship over the past few decades have repeatedly confirmed this pattern, and theorists have offered a broad variety of explanations for it. Although we focus narrowly on self-employment, we draw on the larger literature on social mobility and occupational attainment to provide concepts and principles for our review.

We offer a life course perspective on work careers as a framework for integrating previous work and generating new propositions, building on previous researchers who used diverse data sets, different analytic approaches, and focused mainly on adults. By taking a long-term view of parent–child relationships, we turn a critical eye on the literature, revealing gaps and suggesting new research directions. A unique feature of our review is the addition of the concept of nascent entrepreneur to a model of occupational inheritance among the self-employed. A great deal of research on self-employment conflates selection into entrepreneurship with survival in that status, whereas we propose keeping the two states analytically distinct.

Our paper is organized as follows. First, we suggest a number of reasons for why sociologists ought to care about such links. Second, we propose a life course model of occupational attainment and apply it to work careers and self-employment. We break the life course into semi-discrete segments

for ease of exposition: childhood and adolescence, adolescence into adulthood, and adult status. Because no single data set covers all the points we wish to make, we draw on many. We rely heavily on two data sets in which we were principal investigators: a study of business owners in British Columbia and a national study of nascent entrepreneurs in the United States.

## **WHY CARE ABOUT PARENT TO CHILD OCCUPATIONAL TRANSMISSION?**

Analysts have offered at least three reasons for focusing our attention on the issue of occupational inheritance among the self-employed. First, *inequalities* of income, wealth, and power characterize nearly all societies, but the extent to which such inequalities persist over generations varies widely across societies and over time. Sociologists refer to the persistence of intergenerational inequality as *stratification*: the degree to which families and their offspring remain in roughly the same level in the hierarchy of inequality over time. In a highly stratified society, origins and destinations are highly correlated. Thus, studying occupational inheritance can provide a clue to the open or closed nature of a society's reward structure (Rytina, 1992).

Second, studies of occupational inheritance can identify the sources of values and attitudes that prove useful to entrepreneurs. Miller and Swanson's (1958) oft-cited Detroit Area Study of parenting practices claimed that youth could be imprinted with "bureaucratic" or "entrepreneurial" values by their parents. Kohn and Schooler's (1983) research on parenting practices associated with different parental work environments also suggested that parents could induce predispositions in youth toward particular kinds of occupations, such as a desire for autonomous work. The child's preference for such work might then be fulfilled by self-employment. We thus have theoretical as well as policy-oriented reasons for studying occupational inheritance among the self-employed.

Third, a society's degree of occupational inheritance sheds light on the importance of educational experiences during childhood, adolescence, and even adulthood. Questions arise such as, "does more training in entrepreneurial practices among youth result in more adults becoming self-employed?" To the extent that education and work experience moderate the link between parental origins and children's occupational destinations, it is human capital, rather than parental self-employment, that enables people to enter self-employment.<sup>3</sup> Thus, governments wishing to sever the link

between social origins and destinations would be advised to invest heavily in the education and training of young people.

## A LIFE COURSE PERSPECTIVE ON WORK AND CAREERS

In this section we present a brief overview of the life course perspective. We draw on the work of [Elder \(1999\)](#) to explain the principles underlying the perspective. Full application of the life course perspective to an empirical question makes huge demands on research designs and analyses. Accordingly, in our subsequent review we note the strengths and weaknesses of the data sets that we draw upon.

### *The Life Course Perspective*

The life course perspective has developed across a variety of social and behavioral sciences during the past three decades and refers to the social patterning of events and roles over a person's life span, a process shaped by the interaction of individuals' behaviors and changing historical contexts. Temporal organization is embedded in people's work histories and careers, making the life course perspective especially suited for examining the cumulative effects of career choices. In all capitalist societies, work constitutes an important context for the expression and further development of people's identities and knowledge. [Elder \(1999\)](#) noted four primary principles of the life course perspective.

First, individuals' life trajectories are embedded in and shaped by their experiences of historical times and places over their entire lifetimes. Individuals develop their skills and expectations within specific socio-historical contexts. Second, the developmental impacts of life transitions and events are contingent on when they occur in people's lives. For example, the impact of an event such as parents starting their own businesses may depend on whether it occurs in someone's childhood, adolescence, or adulthood. Third, social and historical influences on individual development are mediated through networks of shared relationships and linked lives. As they move through their lives, individuals take on some roles and give up others, and these roles link them to important others in their environments. Fourth, the life course perspective assumes that people play a constructionist role in shaping their own life course. Within the constraints and opportunities of

changing historical circumstances, individuals make choices and take actions that shape their own life courses. The perspective is not deterministic.

The life course perspective on work careers focuses on issues of timing, and the interaction between individuals' actions and environmental contingencies. It assumes no single fixed cycle or set of stages. For example, workers' likelihood of entering self-employment depends in part on the strength of their attachment to current employers. Individuals enter the labor force in their late teens or early twenties and remain until reaching retirement age or becoming disabled. Depending on their occupations, they may thus be in the labor force for three to five decades, although many interrupt their work careers with spells of other activities, such as unemployment, child rearing, or further education and training (Rosenfeld, 1992). Younger individuals change jobs frequently: in January 2004, median employee tenure was only 2.9 years for workers 25–34, whereas it was approximately four years for the entire workforce. Older workers averaged much longer tenure: 9.6 years for workers ages 55–64 (US Department of Labor, 2004). Age and job tenure are thus quite strongly associated, partially accounting for the curvilinear relationship between aging and becoming an entrepreneur.

### *Importance of History*

In his various reviews of the perspective, Elder has painted an expansive and complex view of life course explanations for individual development. Understanding how and why adults select entrepreneurship as a career option thus requires that we take account of workers' social origins, family environment, early work experiences, proximate opportunities, and many other factors. Historical influences lie at the heart of the perspective and may be classified into three types of effects: cohort, period, and maturation or aging.

*Aging or maturation effects*, as defined by demographers, describe the secular process of aging. For example, Evans and Leighton (1989) found that transition probabilities into self-employment did not increase as people aged, but people did remain in self-employment for longer spells. Consequently, the likelihood of being self-employed rose with age. Uusitalo (2001) found a similar pattern for Finland, with the entry rate constant at 1.5 percent up until age 40 and then declining slightly to a constant 1 percent until retirement. A *cohort effect* occurs when historical events have a differential impact on younger versus older persons. For example, younger children, especially boys, were most adversely affected by the economic strains of the Great

Depression in the 1930s and this might have colored their view of the importance of security in choosing an occupation such as shopkeeper (Elder, 1999). Similarly, desperation led many middle-aged men to attempt self-employment during the Great Depression. A *period effect* occurs when historical events have similar consequences on different age cohorts. For example, the Internet explosion of the 1990s and the resulting dot com bubble were so intense that they affected all age groups' propensities to enter entrepreneurship, raising the entrepreneurial entry rate to an historic high.

Fig. 1 shows a simple schematic diagram of the time span encompassed by a comprehensive application of the life course perspective to entrepreneurial entry. For ease of exposition, we have broken the life course into three segments: childhood, adolescence, and adulthood. In subsequent sections, we will discuss each segment separately. In keeping with the life course view, the picture we have drawn is contextual, contingent, and social, and posits that individuals play an active role in constructing their work careers. For example, for children, their family determines their genetic makeup and their nurturing environment. As adolescents, children begin playing an

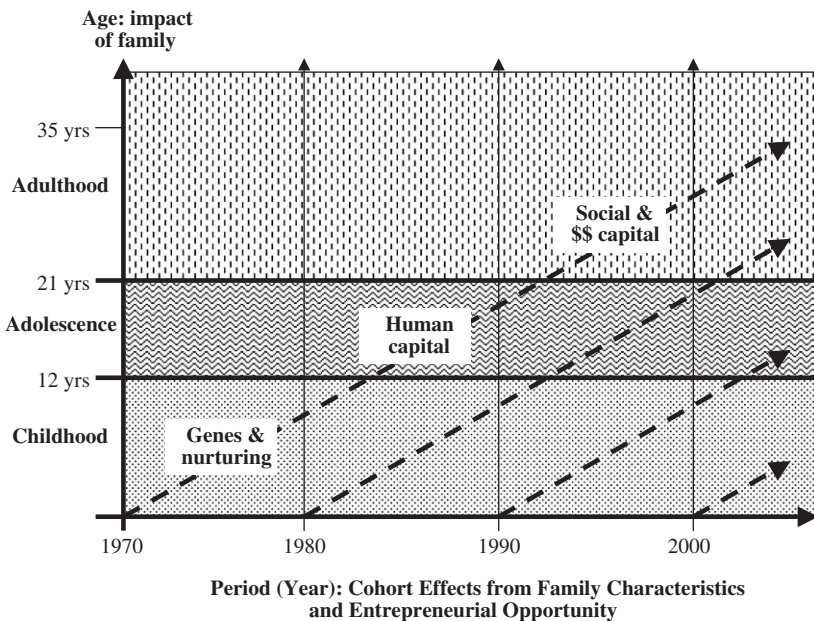


Fig. 1. Lifecourse Perspective of Parental Influences on Entrepreneurial Entry.

active role in choosing compatible peer groups and those peers often become influential enough to displace the primary role of the family.

### *Relevant Research: Data Sets*

A life course perspective on work careers poses severe demands on analysts. To fully capture the historically contingent nature of person–environment interactions, a research design should follow people over a lengthy period and include a substantial number of individual, family, and contextual measures. We uncovered no data sets that met the most stringent tests for a life course analysis of entry into entrepreneurship. Instead, we found data sets that did reasonably well on a few key criteria: (1) clear sampling design, (2) truly longitudinal or very good retrospective information, and (3) adequate controls for important covariates. [Table 1](#) presents a few details on the studies we will mention often in our review. Although [Arum and Mueller’s \(2004\)](#) edited book is a compendium of multiple studies, we also include it in [Table 1](#) because we make several references to it in our review.

In the second column we show the years in which data collection occurred, and in parentheses the years covered by the retrospective questions asked in the surveys. We only searched for English language publications and thus most of the studies cover North America and Great Britain, with a few others from Northern European nations. In the fourth column, we note whether the study used a retrospective or a prospective panel design. Seven of the 15 studies are based on surveys in which people were asked retrospective questions about their lives, and seven are panel studies in which information was collected prospectively about people, either through interviews or archival information. People were then followed over time for at least one more wave of data collection.

From a life course perspective, five of the seven longitudinal studies have the undesirable property of being based on one or more fixed cohorts of individuals. Accordingly, researchers have difficulty separating out “aging” from “cohort/period” effects in most of the longitudinal studies. However, they do permit investigators to calculate transition probabilities. In retrospective designs, investigators are usually limited to analyze inflow/outflow statistics, although if complete work histories are collected, the information can be broken into spells ([Carroll & Mosakowski, 1987](#)). The studies also differ widely in how they defined “self-employment.” For example, [Sørensen \(2006\)](#) only counted employers with employees, whereas [Uusitalo \(2001\)](#) included anyone reporting self-employment earnings or membership in a self-employment pension scheme and [Burke et al. \(2000\)](#) included all

**Table 1.** Studies of Occupational Inheritance.

Authors	Dates of Data Collection and (Years Covered, if Retrospective)	Nation(s)	Data Description
<i>Retrospective</i>			
Aldrich et al. (1998)	1994 (1921–1994)	Canada	229 small business owners in Vancouver area
Carroll and Mosakowski (1987)	1979 (1929–1979)	West Germany	West German life history study: Three birth cohorts
Fairlie (1999)	1996	USA	Characteristics of business owners
Gartner et al. (2004)	1999 (?–1999)	USA	PSED
Hout and Rosen (2000)	1973–1996 (1909–1996)	USA	Pooled GSS surveys
Lentz and Laband (1990)	1979 (?–1979)	USA	Study of 514 NFIB members
Taylor (1996)	1991 (?–1991)	Great Britain	British Household Panel Study
<i>Panel</i>			
Arum and Mueller (2004)	1980–2002	12 industrialized countries	Time period studied varies with country
Burke, FitzRoy, and Nolan (2000)	1958–1991 (1958–1991)	Great Britain	National Child Development Study: single cohort born in 1958
De Wit and Van Winden (1989)	1952–1983 (1939–1983)	The Netherlands	Single Cohort of 6th grade students
Dunn and Holtz-Eakin	1966–1982 (1942–1982)	USA (men only)	National longitudinal surveys of labor market experience: 10-year cohort (born in 1942–1952)
Evans and Leighton (1989)	1966–1981 (1942–1981)	USA (men only)	Series of 2-wave panels from national longitudinal survey of men
Sorensen (this volume)	1980–1997	Denmark	Three cohorts of Danish-born children (born 1966–1968)
Van Praag and Cramer (2001)	1952–1993 (1939–1993)	The Netherlands	Single cohort of 6th graders

those reporting self-employment as their main economic activity plus those who reported self-employment income and met three other conditions.

Most of these studies include very few measures from a respondent's pre-adult years, and thus we found only limited data from their childhood and



adolescence. All of them, however, do report at least the occupation of the father when the child was growing up. Some also report the occupation of the mother. The panels begun when children were born or in primary school contain some personality and educational achievement information, but none provide information on parenting styles. Accordingly, in our review we draw upon other studies of children and adolescents that contain occupationally relevant information.

Most of these studies also include very few respondents from ethnic minority groups. Blacks, Hispanics, and other racial and ethnic groups were overlooked until the 1970s, when their growing political presence led to special government surveys of their business ownership. In their analysis of data from the 1973 through 1996 General Social Surveys, [Hout and Rosen \(2000\)](#) noted that Latino and black men were at a triple disadvantage with regard to self-employment. First, their fathers were less likely than others to be self-employed. Second, even if their fathers were self-employed, they were significantly less likely to become self-employed themselves, compared to other groups. Third, if their fathers were not self-employed, they were less likely than others with similar backgrounds to become self-employed. Using the Panel Study of Income Dynamics (PSID), [Fairlie \(1999\)](#) estimated that African-American men are one-third as likely to be self-employed as white men. He found that much of the gap arose from lower asset levels among blacks and their lower likelihood of having self-employed fathers. Thus, future research projects collecting data on occupational inheritance should over-sample African-Americans and Hispanics so that these issues can be pursued in more depth.

#### *Parent to Child Transmission: Previous Findings*

The data sets we have reviewed, plus several others, permit us to offer three empirical generalizations about the connection between parental self-employment and the self-employment of their sons and daughters. We break the research into three groups: inflow, outflow, and transition probability studies.

#### *Inflow Studies*

Inflow studies start with destination statuses – people who are currently self-employed – and examine the social origins of their parents. Most studies ask about parents' occupations when the respondent was 15 or 16 years of age or when they were growing up, but some simply ask if parents were ever self-employed. Studies consistently find that a high proportion of self-employed

people report that their parents were also self-employed. For example, [Lentz and Laband \(1990\)](#) found that 52 percent of the proprietors they surveyed had parents who were proprietors. Studies in Canada, the United States, and many European nations have found that about one-quarter to one-third of self-employed men report that their fathers were self-employed when they were growing up. When we expand the time frame to parental self-employment over their entire life course, we find that about half of the self-employed people report self-employed parents ([Fairlie & Robb, 2005](#); [Kim, Aldrich, & Keister, 2006](#)).

### *Outflow Studies*

In contrast to inflow studies, outflow studies start with origin statuses – people who are self-employed – and examine the destination occupations of their children. Beginning with occupational mobility studies in the 1950s, sociologists have found that children of self-employed parents are much more likely to be self-employed themselves than would be expected if destinations and origins were independent. For example, [Blau and Duncan \(1967\)](#) showed that the sons of self-employed professionals in the United States were about 12 times more likely to be self-employed professionals than expected by chance. The ratios of observed-to-expected frequencies were much lower for self-employed proprietors and farmers (2.3 and 3.2). Even when not choosing their father's exact occupation, they chose occupations that were similar in autonomy and prestige. For example, sons of proprietors were more likely to become managers and about as likely to become salaried professionals as they were to follow in their fathers' footsteps.

### *Transition Probability Studies*

Inflow and outflow studies usually work with cross-sectional data, whereas transition probability studies examine the likelihood of someone transitioning to self-employment from another status and are based on repeated observations of the same individuals over time. Some examine only a worker's first transition into self-employment, whereas others examine all transitions and include measures of whether the worker previously had a spell of self-employment. The shortest time intervals typically involve panel studies over two years, whereas studies such as the PSID follow respondents over several decades.

Labor force and life course studies, using dynamic data, have consistently found that having self-employed parents (fathers more than mothers) raises the likelihood of transitioning into self-employment by two or three times the baseline rate, e.g. [Dunn and Holtz-Eakin \(2000\)](#). For example, in a

collection of papers from 11 nations assembled by Arum and Mueller (2004), having a self-employed father typically raised the odds of a son transitioning to self-employment by a factor of between 1.3 and 2.2. Thus, at the extreme, the odds of entering self-employment were slightly more than doubled if a respondent's father had been self-employed. However, a more conservative estimate would be that father's self-employment raises the odds by about one-third. The effects differ substantially by whether the father was a professional, skilled, or unskilled self-employed person.

*The Meaning of "Occupation" in Occupational Inheritance Research*

We note several issues that cloud our ability to generalize from these data sets. First, occupational inheritance is not unique to the self-employed. Many other occupations show a disproportionate number of children taking up the same or a very similar occupation as their parents. Even though occupational inheritance can thus occur across the entire occupational spectrum, most researchers have been interested in intergenerational continuity among the more prestigious and highly rewarded occupations, such as lawyers, doctors, and business owners.

Second, most of the transition probability studies – the most technically sophisticated approach – measure only whether the parents were self-employed and do not include other aspects of the parents' occupations or social and psychological environments. For example, Hout and Rosen (2000) included a seven-category occupational coding scheme in their analysis, but nothing else. Thus, any features that "self-employment" shares as a social status with other occupations will be encompassed by this simple coding scheme. "Parent self-employed" could be tapping characteristics of the parents' socialization practices, social networks, financial assets, genomes, and so forth.<sup>4</sup>

Third, almost all sociological studies of intergenerational social mobility use measures of occupational status or prestige, not actual occupational destinations, as independent and dependent variables.<sup>5</sup> Fourth, the virtue of using occupational prestige as a marker of destination, rather than "self-employment," is that "prestige" captures the possibility that children will transition to similar but not identical occupations as their parents'. Even though the children of self-employed parents may never serve a spell as self-employed themselves, they may choose an occupation that is similar in many respects on dimensions such as autonomy, prestige, income potential, and so forth. If we are interested in the advantages self-employed parents provide for their children, then studies focusing only on a limited set of

occupations may cause us to miss many of the benefits derived from particular parental origins.

## LIFE COURSE MODEL OF PARENTAL INFLUENCES ON CHILDREN'S SELF-EMPLOYMENT

From a life course perspective, processes of socialization and control extend from childhood until old age, with succeeding generations linked by age-graded role sequences and social roles. A person's experience within a particular sequence is historically and contextually dependent, with cumulative advantages and disadvantages seamlessly altering transitions and turning points. However, for presentation purposes, we have broken the life course into three discrete segments: childhood, adolescence, and adulthood. For each segment, we review the theoretical principles relevant to understanding which causal forces affect someone's eventual transition into attempting self-employment and succeeding in it. Because genetic endowments interact with environmental factors to influence outcomes across all segments, we begin with a discussion of genetic influences on childhood development. We note that previous studies have tended to err on the side of attributing most of the similarity in parents' and children's occupations to socio-economic factors. We believe that research taking account of genetic influences establishes a more balanced context within which to assess all forms of parental influence.

### *Genetic Influences: General Considerations*

Over the past few decades, research on various developmental outcomes has confirmed the importance of taking account of the genetic inheritance as well as the social origins of people attaining certain occupations (Bouchard & McGue, 2003; Shanahan & Hofer, 2005). When born, children possess a plethora of innate genetic potentials as a result of genes inherited from their parents.<sup>6</sup> Whether these potentials will be fully expressed depends upon the environment in which children are raised. For example, in supportive environments, a child's potential with regard to intelligence has a high probability of being expressed, whereas in impoverished environments, the potential may be suppressed. Thus, the realization of a child's genetic potential is highly conditioned by environmental factors. Note that this implies an interaction between genes and environment, and not a simple additive process, substantially complicating research on the topic.<sup>7</sup> Maccoby (2000),

in her critical review of behavioral genetics research on parents and children, made this point forcefully, arguing that we should not assume a zero-sum game between genes and environment.

To the extent that children and adolescents actively seek out conditions allowing them to develop their potential, the impact of favorable environments is amplified.<sup>8</sup> Poorer environments suppress the potential range of genetic variation of children raised in them. Thus, somewhat paradoxically, genes have a greater impact in highly resource-rich environments than in poor ones. For example, [Guo and Stearns \(2002\)](#) found that heredity not only had a major effect on children's intellectual development but that parental unemployment and ethnicity had a significant effect on the extent to which a child's genetic potential was realized.

Genetic variation has significant effects on a surprising range of behaviors. In their early summary of results from the Minnesota Study of Twins Raised Apart, [Bouchard, Lykken, McGue, Segal, and Tellegen \(1990\)](#) reported that "for almost every behavioral trait so far investigated, from reaction time to religiosity, an important fraction of the variation among people turns out to be associated with genetic variation."<sup>9</sup> For example, current research suggests that from 50 to 70 percent of the observed variation in general intelligence may be due to genetic variation. [Bouchard et al. \(1990\)](#) argued for many characteristics, being reared by the same parents in the same physical environment does not, on average, make siblings any more alike as adults than they would have been if reared separately in adoptive homes.<sup>10</sup> They noted that people might select environments, e.g. peer groups, under pressure from their genomes, thus finding situations that allow a greater expression of their genetic potential.<sup>11</sup>

In a more recent review of the heritability of five major personality characteristics, [Bouchard and McGue \(2003\)](#) found broad consensus on the significant effects of heritability across a range of twin studies and research reviews. One trait with high heritability is "conscientiousness," which [Bouchard and Loehlin \(2001\)](#) suggested encompassed sub-traits such as self-discipline, control, and locus of control. Rotter's measure of locus of control is the only such trait we found measured in any of the occupational inheritance studies we reviewed. In the NLS panel used by [Evans and Leighton \(1989\)](#), Rotter's scale was administered in 1976, when the men were between the ages of 24 and 34. When included in a cross-sectional analysis of the 1981 panel data, individuals who had a more internal locus of control were more likely to be self-employed than wageworkers, as psychological theory would predict. Controlling for locus of control reduced the coefficient for education to statistical insignificance.

Uusitalo (2001) examined three personality characteristics in his panel study of Finnish army recruits from 1982 until 1993: leadership, dynamism, and cautiousness. Psychological theory assumes that these are stable personality traits with roots in genetic differences and childhood rearing environments. All were measured with paper and pencil tests given by the army in 1982, when the men were aged 20. In a cross-sectional logit analysis of the likelihood of being self-employed in 1993, all three variables were statistically significant, with “leadership” and “dynamism” positive and “cautiousness” negative in their effects. Without the original questions, we had difficulty in interpreting precisely what these three really mean, but Uusitalo (2001, p. 1636) argued that “the cautiousness score appears to be very close to what the economists mean by risk aversion.” The significance of the three personality characteristics was maintained when parents’ income and self-employment were included in the equation.

Burke et al. (2000) included four personality traits in their probit analysis of the probability of self-employment in their sample in 1991. The National Child Development Study began in 1958, and respondents were tested in 1965 at age seven to assess their “creativity,” “unforthcomingness,” “anxiety acceptance,” and “hostility toward others.” No details were given on the origins of these tests or their interpretation, and we have no way to assess the magnitude of their effects. Two of the four had statistically significant positive effects on self-employment probability: creativity and anxiety acceptance. We report Uusitalo’s and Burke et al.’s results because they represent one of the very few longitudinal studies of entry into self-employment that assessed psychological characteristics years before measuring entrepreneurial entry. Their results suggest that childhood experience within particular environmental contexts may well affect personality development in ways that lead some people to prefer self-employment later in life.

### *Genetic Influences on Occupational Attainment and Work Values*

Sociologists have historically been skeptical of arguments positing a major role for genes in explaining between-family variance in social behavior, but accumulating evidence from studies of twins reared together and apart suggests an important role for genetic endowments. Discerning readers may recall that Aldrich and Wiedenmayer (1993) argued strongly against a simple “traits” explanation for entrepreneurial behavior. However, we also note that Aldrich and Ruef (2006, p. 61) argued against a sociological alternative of moving to a very macro-level of analysis. Instead, they suggested a multi-level

approach, connecting individuals, organizations, and social contexts. We believe the emerging cooperation between social scientists and genetics researchers will help us avoid the simplistic “traits” approaches of the past.<sup>12</sup>

Lichtenstein, Herschberger, and Pedersen (1995) pointed out that sociologists have observed similarities in occupational position and status between fathers and sons and have taken them as evidence for the influence of socio-economic factors, e.g. childhood socialization. However, they noted that “fathers and sons share 50 percent of their genes” (Lichtenstein et al., 1995) and the observed similarities could thus be due to genetic as well as environmental influences. They used the Swedish Adoption/Twin Study of Aging (SATSA), a sample of twins (monozygotic and dizygotic) similar to the Minnesota study mentioned earlier, to study father/son similarities in occupational status. Lichtenstein and his colleagues used a simple six-part occupational category scheme to classify the 308 men and 288 women in their sample, and showed that two dimensions adequately described the multi-dimensional occupational space: “status” and “farm.” The same solution worked for twins reared together and apart. Then, they performed a second analysis, using the occupational status dimension, to assess the relative importance of genetic and environmental influences on the similarities in twins’ occupational statuses. For men, genetic effects accounted for slightly more than 50 percent of the between-family variation in occupational status, whereas the effects were not significant for women.<sup>13</sup>

In a follow-up study of a smaller sample, they investigated the extent to which the genetic variation in occupational status could be attributed to genetic variation in cognitive abilities (Lichtenstein & Pedersen, 1997). They found significant genetic variance in educational attainment and occupational status that was not due to genetic variance in cognitive abilities, suggesting other genetic-based factors were at work. For example, standard cognitive tests might not capture genetic variance in personality, interests, or talents. In reviewing this research, Pedersen, Spotts, and Kato (2005, p. 81) noted that the “rearing environmental effects of educational achievement and occupational status were completely overlapping, which implied that the same factors in the rearing home made family members similar to each other for both education and occupation.” Thus, the social context in which the twins were raised also contributed to their similarity as adults.

Cognitive abilities can be assessed through IQ tests, which can then be linked to occupational attainment. The heritability of IQ is generally estimated at between 50 and 70 percent and thus if IQ were strongly related to self-employment, it could be offered as partial explanation for occupational inheritance among the self-employed. However, only two of the studies

listed in Table 1 included the IQ of respondents, as measured when they were children aged 12. Both studies used the same data set from the Netherlands, which suffers from severe missing data problems. De Wit and Van Winden (1989) found that IQ had a positive effect on children being self-employed in 1983. Using an updated data set, Van Praag and Cramer (2001) used as their dependent variable whether a person had ever been self-employed from the time they entered the labor force until 1993, and also found that IQ had a positive effect. Although provocative, without replications in other nations and across different periods these two Dutch studies remain mostly an interesting curiosity.

In summarizing research over the past several decades on genetic influences on vocational values and interests, Bouchard and McGue (2003) noted “while the number of studies of interests and work measures is much more limited than in the domains of abilities and personality, there are enough studies using different kinships to make a convincing case that reliable measures in this domain are significantly influenced by genetic factors.” In particular, we speculate that genetic influences may account for between 25 and 50 percent of the association between fathers’ and sons’ self-employment status, in part because of the strong genetic component in the sons’ work values.

### *Childhood: Parental Practices and Work Values*

Whiston and Keller (2004) identified 77 high-quality studies published between 1980 and 2002 in 29 different journals related to the impact of family of origin on career development and occupational choice. Although directed primarily at child psychologists, their comprehensive review provided a valuable summary of the available evidence. They were surprised to find few studies that examined family influences on the career development of young children, given other research on family and parental influence. Bouchard et al. (1990, p. 223) noted that the “evidence for the strong heritability of most psychological traits, sensibly construed, does not detract from the value or importance of parenting, education, and other propaedeutic interventions.”

### *Parenting*

Parents have many routes through which they can influence the occupational attainment of their children: serving as role models, choosing particular child-rearing practices, acting as vocational advisors, and so on. They may also try to shape the educational experiences and peer-group



choices of their children by sending them to private schools and otherwise controlling their rearing environments. Later, in adolescence, they may invest in vocational or college educations for their children, and eventually, provide capital for business ventures.

Early sociological and economic research on occupational inheritance implicitly posited within-family homogeneity in the experiences of children and adolescents. Family characteristics such as parental self-employment and educational attainment were often treated as having the same meaning for all children, but subsequent research has shown substantial within-family variability in outcomes such as educational attainment and occupational status (Conley, 2004). For example, Conley and Glauber (2005) noted that in the PSID, the sibling correlation between occupational prestige scores in 2001 was only 0.225 for sisters, 0.302 for brothers, and 0.233 for all siblings. When controls were introduced for educational attainment, the correlations dropped substantially. Thus, any model of occupational inheritance must allow for substantial within-family diversity in occupational outcomes. In short, we should not expect that all children of self-employed parents themselves become self-employed.

We focus on two types of parental influence in this section. First, we examine research on the sources of *work values* among children. Second, we examine research on children's choices of *specific vocational interests and occupations*. Note that parental influences may lead children to value entrepreneurship, but children may nevertheless fail to develop the other values and skills needed to succeed. They may also simply encounter an unfavorable environment or a run of bad luck.

#### *Parenting: Work Values*

With regard to *work values*, Kohn and his colleagues (1983, 1986) produced perhaps the best-known model linking parents' occupations to their values, and their values, in turn, to the orientation toward work they foster in their children. In this respect, they followed the lead of Miller and Swanson (1958) in arguing that parents' work environments influence how they socialize their children. However, whereas Miller and Swanson used the Detroit area study survey to argue that the entrepreneurial or bureaucratic nature of a father's work setting affected maternal child rearing values and practices, Kohn emphasized the distinction between intrinsic versus extrinsic work values.<sup>14</sup> Kohn argued that work high in substantive complexity, intellectual flexibility, and autonomy led workers to value self-direction and internal standards, whereas work lacking these features led workers to value conformity to externally imposed rules.

In their studies in the United States and Poland, [Kohn, Slomczynski, and Schoenbach \(1986\)](#) confirmed these predictions and found that work values influenced parents' child-rearing practices. Workers with high occupational self-direction tended to value self-direction in their children, whereas workers low in occupational self-direction favored conformity. Most important, interviews with their children confirmed that parental values did, in fact, affect children's values. A follow-up study collected a third wave of interviews for the American sample in 1994–1995 and showed that the effects of occupational self-direction on intellectual functioning and valuing self-directedness persisted over many decades ([Schooler, Mulatu, & Oates, 2004](#)).

#### *Parenting: Gender Differences*

In a portion of the study conducted only in the United States, [Kohn et al. \(1986\)](#) found that the effect of mothers' values on what their children valued was mediated by whether the children accurately perceived their mothers' values. However, for fathers, they found that the effect of values was largely direct and mostly independent of whether the children accurately perceived what their father valued. Thus, fathers had apparently used child-rearing practices that conveyed what they expected, without necessarily articulating those values in a form that children consciously understood.

Kohn's research on gender differences suggests that an inquiry into childhood socialization, parenting, and institutional constraints might prove fruitful in understanding the gender gap in business start-up rates. Until the early 1970s, self-employment rates were very low among women, and women-owned businesses constituted only about 5 percent of all businesses in 1970. As women's proportional representation among business owners began to grow, researchers paid more attention to the phenomenon and a subfield of "women's entrepreneurship" was created. Our review suggests several promising lines of inquiry into forces that created, sustained, and are now closing the gap.

First, women entered the labor market in growing numbers in the 1950s and then began starting their own businesses in greater numbers in the 1970s. Accordingly, daughters of the self-employed raised in the 1970s and later might have experienced a different rearing environment than those raised earlier. Second, parenting styles for boys versus girls may have differed more widely in the 1950s than later ([Buldroft, Carmody, & Buldroft, 1996](#)), as mothers' participation in the labor force grew and self-employment rates increased. Period and cohort differences in the transmission of occupational

values may thus account for the growing number of women-owned businesses.

*Parenting: Ethnographic Studies*

Most studies of parenting practices rely on self-reports from parents and children, or reports from third parties, such as teachers. Ethnographic studies that observe parents and children as they go about their daily activities are clearly needed, such as the research carried out by Lareau (2002). She argued that childrearing practices differ between middle class and working class parents because of resource differences, as well as differences in occupational experiences. In her ethnographic study of black and white 10-year-old children, she contrasted the “concerted cultivation” of middle class children with the “accomplishment of natural growth” of working class children. Among middle class children, the complex organization of their every day activities, parental encouragement of sophisticated language use, and extensive social connectedness beyond kin led to an emerging sense of entitlement. In contrast, working class children gained an emerging sense of constraint in their lives because of their less structured daily lives, little monitoring or coaching of their verbal fluency, and heavy reliance on kin-based socializing.

Lareau’s analysis has ambiguous implications for class-based explanations of entrepreneurial entry, as her findings do not map neatly onto an extrinsic/intrinsic divide in work values. Middle class children’s greater sense of entitlement could give them the confidence they need to mobilize needed resources, whereas working class children’s greater experience with more autonomy in their personal lives could stimulate a similar desire for occupational autonomy. Lareau’s findings suggest that we must look to occupation-specific family experiences to understand the genesis of work values. Unfortunately, her study did not include any self-employed people or employers. Nonetheless, her detailed account of parenting illustrates the potential value of ethnographic studies of self-employed parents and their children.

Returning to our theme of the neglected role of genetic influences, we note that Keller, Bouchard, Arvey, Segal, and Dawis (1992) examined the heritability of work values. Specifically, they examined the aspects of work that people came to value as they grew up with their genetic potential within a particular environment, using a subset of cases from the Minnesota study of twins. After adjusting for age and sex effects, they found that about 40 percent of the variance in measured work values (achievement, comfort, status, safety, and autonomy) was associated with genetic differences within

the sample. Keller et al. (1992) pointed out that people are not “born” with work values, but rather learn or acquire them as they mature. Nonetheless, this small study reinforces findings from others that showed genetic factors influence vocational interests, e.g. Waller, Lykken, and Tellegen (1995), and reminds us that parental influence operates along several dimensions.

*Parenting: Vocational Interests and Occupational Choices*

With regard to *specific vocational interests and occupational choices*, a number of studies have suggested that parents play at least a moderately significant role, although the design of most has not matched parents’ to children’s reports of influence (Whiston & Keller, 2004). Such influences are already apparent in pre-school children (Barak, Feldman, & Noy, 1991). We have already noted that Kohn and his colleagues found strong evidence that parents’ occupations shape their children’s orientations to general conditions of work via the parents’ value systems. In addition, children’s attitudes and aspirations toward specific occupations might be sensitive to parental influence.

However, Trice’s research raises questions about children’s awareness of parents’ jobs. In a study of 11 to 14-year-olds, Trice and Knapp (1992) found that only 68 percent accurately reported their father’s occupations, and children’s occupational aspirations had no statistically significant relationship to their fathers’ occupations (occupations were classified into a six-category scheme). Fewer than 10 percent of the children they interviewed reported that “parental suggestions” in response to the question, “has anyone ever told you that when you grow up you should be something or that you would be good at something?” Even when parental suggestions were reported, the children had *not* described these suggestions as either their own first or second choices. In their study of 949 elementary school children from 11 different schools, Trice, Hughes, Odum, Woods, and McClellan (1995) again found no significant relationship between boys’ or girl’s expressed job preferences and their fathers’ jobs, whereas mothers’ jobs were associated with children’s preferences (jobs were coded in a six-category scheme).

Longitudinal studies of children’s occupational aspirations and expectations are rare, and one of the few to observe how parental influence wanes over time was Helwig’s (2004) decade-long study of 208 second graders from the Denver area interviewed six times.<sup>15</sup> He also obtained information about their parents. Students were asked what job they would like to have as an adult, what job they think they will actually have, and what job do they think their mothers and fathers would like them to have.<sup>16</sup> Although sample attrition makes interpretation of the results ambiguous, there was a clear

tendency for older children to report that parents became less directive and more willing to let them choose “anything I want.”

In addition to their role as socializing agents, parents also affect their children’s interests via their genetic endowments. A number of studies using the Minnesota twins study has concluded that variation in occupational interests is strongly correlated with genetic factors. For example, [Lykken, Bouchard, McGue, and Tellegen \(1993\)](#) found that genetic influences accounted for as much as half of the variance across a battery of specific vocational and recreational interests. They noted that individuals learn such interests and precursor traits of aptitude and personality shape people’s interests in specific aspects of occupations. [Whiston and Keller’s \(2004\)](#) review of child development research over the previous two decades suggests that parental contributions to their children’s aptitude and personality constitute probably a bigger influence on their occupational interests and expectations than either directive suggestions or specific occupational role modeling.

### *Adolescence into Adulthood*

Children move into their teenage years with very general vocational interests and expectations, having shed their “fantasy occupations” of childhood, e.g. an astronaut or doctor, but without fixing on any specific future jobs. Their families have influenced their development through the structuring of home environments, such as through discipline practices that encourage self-direction. Adolescence is a time of vocational exploration and identity development, as aptitude and personality traits interact with environmental contingencies to open some avenues and close others. During this period of transition, studies have documented that children perceive their families, including siblings as well as parents, as influencing their career choices ([Whiston & Keller, 2004](#)). Not surprisingly, children who perceive their families as supportive and having high expectations tend to have higher occupational aspirations than others.

Given the obvious importance of the transition from school and adolescence to begin a career, we might expect adolescents to be quite active in exploring possible jobs and occupations. However, as [Mortimer, Zimmer-Gembeck, Holmes, and Shanahan \(2002, p. 221\)](#) noted in summarizing the work of [Schneider and Stevenson \(1999\)](#) “only a minority of high school students seriously considers potential career paths by seeking information or by engaging in appropriate activities, even though almost all report

occupational aspirations when asked. In addition, many lack basic information about how much education they need for the occupations they are considering.” Thus, we should not think of youth as preparing for future self-employment in any systematic way, but instead consider ways in which family contexts shape work values, preferences, and possibly skills.

*Difficulties in Linking Adolescent Experience to Entrepreneurship*

We face several difficulties in establishing a connection between family contexts during children’s adolescent years and their becoming entrepreneurs sometime in the future. First, very few people choose self-employment as their first job. For example, [Dunn and Holtz-Eakin \(2000\)](#), using the PSID, found that the average age of first self-employment for sons ranged from 26 to 27.5 years, depending on whether their father or mother was self-employed. Standard deviations for these estimates were about 5 years, indicating that many sons were in their early 30s before becoming self-employed for the first time. Also using the PSID, [Williams \(2004\)](#) noted that less than 1 percent of youth aged 16 and 17 were self-employed, with the percent increasing to about 3.5 percent by age 24.<sup>17</sup> [Lentz and Laband \(1990\)](#), using National Federation of Independent Businesses (NFIB) survey data, reported that the age of first-time owners following in their parents’ footsteps, but not inheriting their business, was about 36, compared to 37 for non-followers and 32 for people who inherited their business. In their national survey of 692 youth aged 14–19, [Kourilsky and Walstad \(2000\)](#) asked their respondents how many years they would wait before they acted on their desire to start a business, and 53 percent said at least seven years or more.

Most entrepreneurs therefore gain work experience at other jobs before tackling self-employment. Thus, from a life course perspective, the time lag between children leaving home and actually becoming self-employed as adults provides many opportunities for intervening events to dampen or amplify earlier experiences. Over such long durations, disruptive or catalytic events might disrupt the linked lives of parents and children. Additionally, conditions promoting entrepreneurial aspirations might not be the same as those promoting entrepreneurial success. Young adults may come to value an entrepreneurial career but not have the skills and resources to succeed in their attempts. Accordingly, we must be cautious in searching for direct links between adolescent experiences and adult occupational attainment.

Second, most studies of occupational inheritance, whether retrospective or panel, obtain very little information about processes of socialization and control that are household specific. Most questions about parents’ occupations ask about fathers’ and mothers’ jobs when the respondent was 16 years

old. However, they typically obtain no information about business size, years of self-employment, whether the respondent worked in the business, and so forth. For example, we found no studies asking the kinds of questions necessary to examine the family dynamics studied by Kohn and his colleagues.

### *Adolescent Occupational Aspirations*

In this section, we review studies that link occupational inheritance to parental activities that occurred in late adolescence, such as Mortimer's study of students at the University of Michigan. We examine findings from previous studies that included age-specific information on self-employment durations in families, such as Aldrich, Renzulli, and Langton's (1998) Vancouver study and Sørensen's (2006) Danish study. In the next section, we offer some new findings from the Panel Study of Entrepreneurial Dynamics (PSED).

Beginning in the 1950s, researchers found that many adults expressed an interest in being self-employed or owning their businesses (Chinoy, 1955; Lipset & Bendix, 1959). For example, Steinmetz and Wright (1989, p. 973) noted that in 1980, "57% of all people in the American working class and two-thirds of all male workers say that they would like to be self-employed someday." Adult interest in self-employment has apparently been matched by that of high school students. For example, a Gallup Organization poll of 967 youths aged 14–19 in 1995 (Kourilsky & Walstad, 1998) found that 72 percent of the young men and 62 percent of the young women answered yes to a question, "Would you like to start a business of your own?" The study was repeated in 1999 with a national sample of 1,148 youth and 1,104 young adults, aged 21–30 (Kourilsky & Walstad, 2000). At that time, 61 percent of the youth and 58 percent of the young adults said "yes."<sup>18</sup> Very similar results were obtained 10 years later in a Junior Achievement online poll of 1,155 students in 2005, which found that 69 percent of students expressed an interest in starting their own business (Bell, 2005).

Over the past few decades, many non-profit organizations have been founded around the world to promote "entrepreneurship education" among young people, e.g. the Consortium for Entrepreneurship Education in the USA, the National Collegiate Entrepreneurship Organization in the USA, the Junior Achievement Worldwide Association, the Enterprise and Industry Initiative of the European Union, the Enterprise Insight program sponsored by the British government, and many state-level initiatives in the United States, such as those sponsored by the Appalachian Regional Commission. In the United States, the Ewing Marion Kauffman Foundation of

Kansas City has funded many initiatives aimed at age groups ranging from elementary school children to high school youth. Thus, today's youth live and work in an environment saturated with information and images about entrepreneurship and business ownership. Indeed, the period beginning with the 1980s revival of interest in entrepreneurship changed American culture to such an extent that research results on occupational inheritance from earlier decades may no longer be valid. As we noted in [Table 1](#), most of the occupational inheritance data sets cover a period in which the respondents were growing up in the 1940s through the 1960s, and their findings may thus be historically specific.<sup>19</sup>

Even though their environments are rich in information about entrepreneurship, research on occupational aspirations and expectations suggests that interest in specific occupations develops from more immediate experiences. Extrapolating from the research we have reviewed on career development and counseling among youth, we assume that most adolescents lack knowledge of the occupationally specific skills needed for self-employment. Moreover, because typically a decade or more passes before high school graduates attempt to become self-employed, we doubt that learning entrepreneurially specific skills during adolescence makes much of a difference. By contrast, youth can develop a preference for occupational autonomy, flexibility, and substantive complexity while observing parents engaged in their chosen occupations. If their parents are self-employed, of course, youth may acquire some specific skills in business management.

Jumping too quickly into self-employment during adolescence actually seems to harm the future career prospects of youth. In his analysis of youth over the period 1979–1989, using the PSID, [Williams \(2004, p. 334\)](#) found that “time spent in self-employment as a youth is negatively correlated with current earnings, and ... the returns to self-employment experience are significantly lower than the returns to wage employment experience in the wage and salary sector.” Most of the youth were self-employed for very short periods, but apparently they lost out on the development of new skills and also lost opportunities to gain experience about the functioning of the labor market.

#### *Acquiring and Reinforcing Occupational Values*

Does growing up in an entrepreneurial household make a difference in young adults' career choices? We found no large-scale research projects focusing on family dynamics involving adolescents within self-employed households. However, we found a well-designed study that explored the effect of fathers' occupational attributes on their 18–22-year-old sons' career



choices. Mortimer (1974, 1976) used survey data collected from two entering classes of University of Michigan students in 1962 and 1963, who were also interviewed in 1967 and 1968. Within this group of middle and upper middle class sons, most had fathers in high prestige occupations: 12 percent had fathers who were doctors, dentists, or lawyers, and 27 percent had fathers who were self-employed businessmen. Sons were asked which occupation they were planning on entering after graduation.

Mortimer used the Dictionary of Occupational Titles to code occupations on the basis of “interests,” and applied them to all 129 occupational groups. The groups were collapsed into a 13-category code, using “interests,” for a smallest space analysis of fathers’ and sons’ occupational choices. Results showed that the students had a strong tendency to choose their father’s occupations. If they did not choose exactly the same occupation, they chose ones with similar levels of autonomy, reward structure, and work activities (as inferred from the smallest space analysis). Parental values – as revealed in the occupations they held – regarding work thus seemed to strongly influence their sons’ expectations for careers. The closeness of father and son relationships had a mediating effect on whether sons adopted their fathers’ occupations.

Several studies have tried to assess the extent to which adults were exposed to an entrepreneurial lifestyle when they were children. Aldrich et al. (1998) studied 229 business owners in the Greater Vancouver area of British Columbia, Canada, in 1995, collecting a great deal of information about parents who were self-employed. About 55 percent reported that one or both of their parents were self-employed at some point in their lives, with the average duration of business ownership almost 28 years (standard deviation of 16 years). Under these conditions, children could gain experience with an entrepreneurial lifestyle in two ways: by simply living in the same household and by actually working in the business. About 83 percent of the children of owners had some experience with an entrepreneurial household prior to age 22, with 63 percent having 10 or more years of exposure.

Exposure to an entrepreneurial lifestyle while living at home, however, may not be sufficient to gain the tacit knowledge and values involved in being self-employed. Children may gain an appreciation of an entrepreneurial life but not learn enough to succeed at it. Working in the business involves a commitment that can drive home the lesson of the demands of self-employment. With regard to working, only 61 percent had actually worked for their parents. About 25 percent began working before the age of 10, 27 percent between 10 and 14, 33 percent between 15 and 19, and 15 percent at 20 years of age or older.

Even though most of the employed children started at a young age, very few stayed in the business as a career. About 30 percent worked less than five years, and 51 percent left after their ninth year. Only 12 percent reported working 20 or more years for their parents, even though many more had the opportunity to do so.<sup>20</sup> Thus, although a majority of the children whose parents owned a business began working in it at a fairly young age, the work was short term and most left for other jobs before they turned 21. These results show that a minority of the children of self-employed parents is not only actively exposed to the lifestyle but also have first hand working knowledge of the activities involved. Unfortunately, none of the transition probability studies we reviewed contained any information about the duration of exposure of children to self-employment or the extent of their involvement.

#### *Occupational Inheritance in Denmark*

Sørensen (2006) tried to separate the consequences of being exposed to an entrepreneurial lifestyle during adolescence from exposure during adulthood in his study of the transition to being a self-employed person with employees in Denmark. He found that parental self-employment had a significantly positive effect on the transition to self-employment, regardless of whether it occurred during adolescence or adulthood. For exposure during adolescence only, the odds increased by 1.52, for adulthood only by 1.68, and for exposure during both periods by 1.84. (The last coefficient is statistically significantly different from the first but not the second.) Sørensen explained the adolescent exposure effect by invoking role modeling, noting that even if parents were no longer self-employed by the time their children entered the labor force, the consequences of early exposure persisted.

He explained the effects of adult-only exposure by positing that self-employed parents provide social capital to their children, who aspire to become entrepreneurs, such as by passing on knowledge of opportunities to them and making introductions for them. This interpretation is strengthened by his finding that prior work experience in a parent's industry and having a self-employed parent increases the likelihood that a child will enter the same industry. Returning to our theme of including a consideration of genetic endowments in a life course model, we note that if children inherit aptitudes and abilities that require a favorable environment for their development, parental self-employment during adolescence could enhance their effects. However, if the aptitudes and abilities can be developed in *any* familial context that fosters occupational self-direction and autonomy, then parental self-employment during a child's adulthood is not a sign of the

transfer of social capital. Instead, it could be taken as a marker of a genetic endowment combined with a child-rearing environment that allowed the development of entrepreneurial-relevant skills and values during childhood.

Assessing the extent to which Sørensen's results reflect children acquiring their parents' occupational values and skills versus social capital would require additional information: for how many years was the child exposed to parental self-employment and did the child actually work in the business, and what occupational route did siblings follow. (If a disproportionate share of siblings also enters self-employment, then we have added grounds for investigating between-family variance.) Sørensen's analysis ended when the children were aged 29–31, and thus he captured only the entrepreneurial early movers, as well as those with enough capital and organizing ability to hire employees for their first venture. We will return to Sørensen's analysis when we consider what resources self-employed parents can offer their adult children.

*Evidence from the PSED: Parental Influences on becoming a Nascent Entrepreneur*

Many of the studies we have drawn upon include information about parental occupations but not about the businesses they owned. In contrast, the PSED includes extensive information on business characteristics. In this section, we use the PSED to look more closely at possible sources of parental influence.

*PSED Study Design*

The PSED's module on parents built on some of the findings from the Greater Vancouver project, as more information was obtained about the businesses owned by respondents' parents. Unlike the other studies we have reviewed, the PSED focused on *nascent entrepreneurs* rather than people already in business. Respondents were included in the nascent entrepreneur sample if they reported that had taken action, alone or with others, within the past year to start a new business. It thus identified people very early in the start-up process, before the outcome of their efforts was known, and is arguably an indicator of people's interest in self-employment without regard to their abilities to actually succeed.

The PSED also included a comparison sample of people who identified themselves as not trying to start a business. Thus, by combing the two samples, properly weighted, we can assess the extent to which parental

characteristics differentiate people who say they were trying to start a business from those who were not during the period 1999–2000. Because the respondents were not followed over their working careers, we cannot use the PSED to calculate transition probabilities. Instead, we have a snapshot of who was attempting a start-up at a moment in time when entrepreneurial interest was rising rapidly in the United States. The combination of a different period and a different outcome means that we expect differences in our results, compared to other projects that examined people reporting their occupations as “self-employed.”

We can use the PSED to gain an understanding of how much experience respondents have had with parental self-employment. The survey asked people if their parents were ever self-employed, how many businesses they owned, for how many years their parents ran their businesses, the size of the largest business they owned, and whether the respondent had ever worked for any of the parents’ businesses. Using this information, we can assess the 1999–2000 PSED sample’s experience of parental self-employment, but we must note two limitations that prevent us from generalizing about the past. First, differential parental fertility means that our sample of sons and daughters may not accurately capture the historical distribution of self-employment by parents. Second, differential cohort survival and other factors affecting sample selectivity mean that we must be careful in inferring cohort differences, using the age of our respondents.

All respondents were asked, “Did your parents ever work for themselves or run their own businesses, alone or together?” and 50 percent answered “yes.” Although this percent may seem high, note that it refers to any spell of self-employment, however brief, over the parents’ entire life span.<sup>21</sup> For comparison purposes, we note that [Dunn and Holtz-Eakin \(2000\)](#) reported that 30 percent of the fathers and 16 percent of the mothers of sons in the PSID experienced at least one year of self-employment over the 15 years of their study (1966–1981). [Fairlie and Robb \(2005\)](#) reported that about 52 percent of all business owners in the 1992 CBO had at least one self-employed family member prior to starting their firm. [Reynolds and White \(1997\)](#) showed that between 40 and 42 percent of the respondents aged 40–69 in the 1968–1988 waves of the PSID reported at least one spell of self-employment. In the national longitudinal survey of youth (NLSY) 1979–1998 waves, at least 25 percent of the original respondents experienced at least one self-employment spell ([Budig, 2006](#)). This percentage is biased downward as it is unadjusted for sample attrition and only covers the early working years (aged 18–41). Thus, at the outset, we know that a large

proportion of the American work force has parents who have experienced at least one spell of self-employment over their working careers.

Many occupational inheritance studies either include only whether a respondent's father was self-employed, or include both parents into a single indicator of parental self-employment. As we will show, however, family employment dynamics are different for children, depending upon the business's ownership structure. Among the PSED respondents reporting parental self-employment, 43 percent reported that only their father had owned a business, 11 percent said only their mother, 31 percent reported joint ownership by both parents, and in 8 percent of the cases, each parent owned a separate business. (Some other combination of activity was reported by 3 percent of the respondents.) In the following tables, we will report statistics separately for three types of ownership: father's business, mother's business, and joint ownership.

Because the startup propensity for women has historically been lower than that for men, we also report statistics separately for men and women respondents so that we can observe any possible contributions of parental self-employment to the sex differential. To anticipate our results, the only significant sex difference we found concerned whether respondents had worked in their parents' businesses, with men much more likely to work for their fathers' or a jointly owned business than were women.

### *Results from the PSED Analysis*

In most cases, respondents reported that their parents owned only one business, as shown in [Table 2](#). Some of the parents were serial entrepreneurs, owning more than one business, with 11 percent of the fathers owning three or more and 15 percent of the joint owners having three or more. Only 9 percent reported that their mothers had owned more than one business. As might be expected from variation in numbers of businesses owned, parents varied greatly in the total number of years they had owned businesses, as shown in the second panel of [Table 2](#). Women and men were equally likely to report their parents were serial entrepreneurs and the duration of ownership was also the same.

Comparison data are rare, but we note that [Reynolds and White \(1997\)](#), summarizing PSID data, reported that about half of the self-employed heads of households aged 40–59 had self-employment spells lasting six or more years. [Dunn and Holtz-Eakin \(2000\)](#) reported that over the 15 years covered by their study, father who were self-employed at any time during the study years spent nearly three-quarters of their working time in that state. Interestingly, in their equations predicting the transition to self-employment, they

**Table 2.** Parental Ownership Characteristics: PSED.

	Father ( <i>N</i> = 252)			Mother ( <i>N</i> = 111)			Joint ( <i>N</i> = 159)		
	Female (%)	Male (%)	Total (%)	Female (%)	Male (%)	Total (%)	Female (%)	Male (%)	Total (%)
Number of businesses owned									
1	70	59	64	89	95	92	68	66	67
2	22	28	25	2	1	2	12	23	18
3 or more	8	14	11	9	3	7	19	11	15
Total percent	100	100	100	100	100	100	100	100	100
<i>p</i>			0.55			0.44			0.49
Number of years as owner									
1–9	26	24	25	47	58	51	8	20	14
10–19	25	20	22	37	39	38	39	20	30
20–39	37	39	38	16	3	11	33	40	36
40 or more	12	18	15	0	0	0	20	20	20
Total percent	100	100	100	100	100	100	100	100	100
<i>p</i>			0.86			0.35			0.42
Largest business owned									
No employees	23	12	17	50	30	43	14	2	9
1–4 employees	50	34	41	24	45	32	35	35	35
5–9 employees	17	24	21	9	1	6	23	32	27
10 or more employees	10	30	21	17	24	19	28	31	30
Total percent	100	100	100	100	100	100	100	100	100
<i>p</i>			0.11			0.35			0.36
Worked for parent									
Full-time	1	24	14	8	1	6	8	53	30
Part-time	36	32	34	18	26	21	50	26	38
Did not work at all	63	44	52	73	73	73	42	20	32
Total percent	100	100	100	100	100	100	100	100	100
<i>P</i>			0.01			0.39			0.00

found that taking account of the duration of parents' self-employment, as well as its timing, produced coefficients of the same magnitude as their preferred "any parental exposure over the sample period" indicator of parental involvement in self-employment. Thus, the duration of exposure to parental self-employment was no more significant than the exposure itself.

Most of the businesses owned by parents were quite small, as shown in panel 3 of Table 2. About 58 percent of businesses owned by fathers and 75 percent of businesses owned by mothers had fewer than five employees. In contrast, jointly owned businesses were larger: only 44 percent had fewer than five employees, and 30 percent had more than 10. Again, we found no significant sex differences. Unfortunately, none of the studies listed in Table 1 reported information on the number of businesses owned or their size, and so we cannot offer comparative statements. However, the size distribution of the parents' businesses compares quite closely with that of all US businesses (Aldrich & Ruef, 2006).

#### *Children's Employment in Parents' Businesses*

Very few of the PSED respondents worked full time for their parents' businesses, as shown in the last panel of Table 2. Indeed, for those without jointly owned business in their families, most did *not work at all* in the business owned by one parent.<sup>22</sup> Full time work was also rare. Just 14 percent of the children whose fathers owned businesses and only 6 percent of those whose mothers owned businesses worked full time. By contrast, children whose parents jointly owned businesses were much more involved in the running of the business: 30 percent worked full time and 38 percent worked part time. These jointly owned businesses fit the classic notion of a family-owned and managed business: father and mother as joint owners and at least one child working in the business. Thus, about one in twenty respondents in our sample spent at least part of their working lives in a traditional family-owned business.<sup>23</sup>

Sons and daughters differed significantly in their involvement in parents' businesses. For businesses owned by respondents' fathers, about 24 percent of the men reported working full time and 32 percent part time, compared to only 1 percent of the women who worked full time and 36 percent who worked part time. Thus, almost two out of three women reported not working at all in their father's business. Even more dramatically, 53 percent of the men reported working full time in a jointly owned family business, compared to only 8 percent of women. Only 20 percent of the men reported not working at all in a jointly owned business, compared to 42 percent of the women. In contrast, neither men nor women were very involved in their

mothers' businesses, as 73 percent of each sex reported not working at all in the business. From this information, we were tempted to infer that part of the gender gap in business start-up and persistence rates stems from the differential involvement of sons and daughters in their parents' businesses. However, as we shown in Table 3, controlling for involvement in parental businesses does not significantly reduce the gender gap in the entrepreneurial entry rate.<sup>24</sup>

To determine whether parental business ownership might have influenced the likelihood of their children attempting entrepreneurial entry, we conducted a logistic regression analysis. We coded our dependent variable (nascent entrepreneur = 1, otherwise = 0) if the respondent qualified based on whether he/she answered yes to the following question: "Are you, alone or with others, now trying to start a new business?" In addition, we only included individuals as nascent if they expected to be owners or part owners in the firm, reported being active in trying to start the new business in the past 12 months, and were still in the start-up phase. The fully specified model included measures of education, work experience, current employment status, age, marital status, sex, race and ethnicity, and region, as developed in Kim et al. (2006). In this analysis, we report only the coefficients relevant to parental business ownership, as shown in Table 3.

In Model 1 of Table 3 we show the simple additive effects of three forms of parental business ownership on a respondent being a nascent entrepreneur in 1999–2000, and none of the three is statistically significant. Prior to entering these three variables into the equation, the coefficient for sex (male = 1, female = 0) in an equation with all the control variables was 0.515 ( $p < .01$ ), and adding the three ownership variables does not change the coefficient significantly. Men were about 1.67 times as likely as women to be nascent entrepreneurs, a result similar to that observed in the studies in Table 1 where the dependent variable was actual self-employment rather than attempting to start a business.

We checked to see whether the effects of parental self-employment might differ by sex, as shown in Model 2 of Table 3, by including an interaction term for ownership by sex. None of the interaction terms is statistically significant and the coefficient for sex increases slightly. Thus, the gender gap in becoming a nascent probably does not result from a differential effect of parental ownership on men and women.

As a final check on the impact of parental ownership, we included two of the four parental ownership characteristics in another logistic regression, as shown in Model 3. We included number of businesses owned by fathers, mothers, and joint parental ownership because we felt it was the best



**Table 3.** Logistic Regression Results of Entrepreneurial Entry on Parental Ownership Characteristics: PSED.

Independent Variables	Model		
	1	2	3
Male	0.535** [0.176]	0.640** [0.219]	0.592** [0.180]
Father owns business	-0.331 [0.225]	0.058 [0.328]	
Father owns business × male		-0.592 [0.430]	
Mother owns business	0.075 [0.291]	-0.332 [0.417]	
Mother owns business × male		0.771 [0.589]	
Joint parental ownership	-0.056 [0.245]	0.085 [0.360]	
Joint parental ownership × male		-0.251 [0.475]	
Number owned: father's bus			-0.114* [0.069]
Work for father's bus			-0.116 [0.312]
Number owned: mother's bus			-0.131 [0.287]
Work for mother's bus			0.928 [0.586]
Number owned: joint bus			0.224* [0.118]
Work for joint bus			-0.422 [0.350]
Constant	1.354 [1.512]	1.294 [1.535]	1.537 [1.561]
Observations	1021	1021	1021
-2LL	-212.25	-211.6	-210.93
DF	26	29	29
$\chi^2$	119.12	122.36	120.6

Robust standard errors in brackets, two-tailed test. The fully specified model included measures of education, work experience, current employment status, age, marital status, sex, race and ethnicity, and region, as developed in Kim et al. (2006).

\*Significant at 10%;

\*\*Significant at 5%.

indicator of whether parents had pursued an entrepreneurial lifestyle. We did not use number of years owned or the size of the largest business because without measures of business profitability, we were uncertain of how to interpret them. We included whether respondents had worked for their parents' businesses not only because it indicated that children might have acquired some entrepreneurial skills, but also because it might have affected their occupational aspirations and values. None of the parental ownership variables in Model 2 were significant and so we left them out of Model 3.

Our results paint an equivocal picture of parental influence on children's desires for self-employment. Of the two indicators for fathers' businesses, having worked for their father had no effect, and the number of father-owned businesses actually *decreased* the likelihood of becoming a nascent entrepreneur. The effect is fairly small: each additional business owned, compared to never owning a business, reduced the odds of nascency by about 10 percent. If some fathers owned a succession of marginal or failed businesses, the negative effect we uncovered might reflect children's reactions to observing unsuccessful serial entrepreneurship. Alternatively, parental serial entrepreneurship might actually have been successful, thus allowing sons and daughters to pursue the managerial or professional careers that most small business owners desire for their children. Without more information about the performance of parents' firms, we cannot adjudicate between these interpretations.

Neither of the indicators for mothers' businesses had a statistically significant effect, although we note that again, the coefficient for number of businesses owned is negative. Finally, for jointly owned businesses, the number of businesses owned has a positive effect, whereas working in the business has no significant effect. Having a jointly owned parental business in the family increased the odds of being a nascent entrepreneur by a factor of 1.25, an effect apparently not due to the son or daughter having worked in the business. Indeed, the non-significant coefficient for having worked in a jointly owned parental business is negative.

Our findings in Tables 2 and 3 show that parental ownership has little discernable effect on children's propensities toward becoming nascent entrepreneurs. About half of all adults in our 1999–2000 survey reported at least some parental business ownership in their families, with most reporting the business was either their father's or was jointly owned by their mothers and fathers. Of those reporting some ownership, most said their parents experienced only one spell of ownership, especially for their mothers. However, the spells were of fairly long duration. Most businesses were very small, in keeping with the overall organizational landscape of the American

economy. Very few worked full time for their parents' businesses, with only those reporting jointly owned parental businesses seemingly very much involved in the business.

Strikingly, people's involvement in nascent entrepreneurship in 1999–2000 was not associated with whether their parents were owners, or how much they were involved in the businesses. Indeed, experience with father-owned businesses appears to have decreased people's interest in starting a business themselves. The only hint of a possible lingering effect of family-run businesses is the positive effect of a family being heavily involved in ownership through multiple firms. Finally, the gender gap in becoming a nascent entrepreneur was not narrowed, regardless of which indicators of parental ownership we included in our models.

#### *Adulthood: Do Parents Assist Adult Children with Business Ventures?*

To this point, we have examined two of the three life course segments portrayed in Fig. 1: childhood and adolescence. We turn now to the issue of the extent to which parents help their adult children enter self-employment. By the time they enter adulthood, institutional environments have imposed constraints on children and have interacted with family, school, and genetic endowments to establish most of the human capital people bring to their working careers. In the previous sections, we noted that occupational values and expectations have been established by late adolescence. Nonetheless, self-employed parents could potentially serve as role models and provide some on-the-job training that affects the likelihood of adult children taking up self-employment.

In the life course perspective, the principle of linked lives reminds us that parental assistance depends upon the conjuncture of parents' and children's stages in the life course. Are parents in a position to help? Do children time their pursuit of business opportunities to coincide with parental abilities to make resources available? Or, do children pursue their occupational goals, regardless of current parental resources, meaning that parents may not be in a position to help at a critical juncture. In this section, we focus on two types of resources that have been discussed in the literature on occupational inheritance of self-employment: financial support and social capital.

#### *Financial Support*

Three types of financial support have been studied: the inheritance of a business, the inheritance of capital such as through an estate settlement, and

the availability of parental assets through loans or grants. Most studies indicate that very few self-employed people *inherited* their businesses from their parents or other family members. The 1994 Vancouver study found that 5.2 percent of all owners inherited their business directly from parents (Aldrich et al., 1998), and another 3.1 percent bought their businesses from their parents. Using the 1993 CBO data, Fairlie and Robb (2005) estimated that only 1.6 percent of all the small businesses were inherited.

Fairlie and Robb (2005) noted that two federal reserve surveys have provided information on businesses obtained through “inheritances and gifts,” thus commingling possible parents’ passing on businesses to children with other people’s “gifts” to new owners. The Survey of Small Business Finances estimated that 4 percent of firms were inherited or acquired as gifts, and the Survey of Consumer Finances estimated that 3.5 percent of businesses were similarly obtained. The 1992 CBO survey reported that 6.6 percent of owners acquired their businesses through a transfer of ownership or a gift. However, after removing owners who did not have self-employed family members prior to starting their businesses, Fairlie and Robb (2005) estimated that only about 4 percent obtained their firms through transfers of ownership or gifts.<sup>25</sup>

*Inherited wealth* can facilitate the founding of a business, and a number of studies have found that sudden increase in wealth appear to increase the likelihood that someone will enter self-employment (Blanchflower & Oswald, 1998). Other studies have investigated the association between parental wealth and becoming self-employed (Dunn & Holtz-Eakin, 2000), and we will consider them shortly. Fewer studies, however, have looked specifically at wealth inherited from parents and applied to entrepreneurial entry. In their Vancouver study, Aldrich et al. (1998) explicitly asked owners if they had used capital from an inheritance to fund their business, and all *denied* receiving such a bequest.

In Finland, Uusitalo (2001) was able to obtain information about who had received an inheritance or other unusual income during the previous five years, but this information was available only in the first wave of the panel study. About 18 percent had received an inheritance, but when a dummy variable for inheritance was included in an equation for the transition into self-employment, the coefficient was not significant. In contrast, Burke and Oswald (2000), using an updated version of the data set employed by Blanchflower et al. (1998) found a sizeable effect for inheritances. However, as with Uusitalo’s study, the actual source of the inheritance was not given.

Loans and grants comprise a more likely way for parents to make resources available to the children, who are contemplating becoming

self-employed. However, most studies have found that parents seldom provide startup capital to their children. In Vancouver, Aldrich et al. (1998) found that only 8.7 percent of the owners obtained any capital from self-employed parents, comparable to the 7.8 percent who obtained capital from their non-self-employed parents. In the 1992 CBO study, only 6.4 percent of the owners borrowed capital from their family (Fairlie & Robb, 2005).

Several studies have investigated the possible effects of parental assets on children's transitions to self-employment, although they cannot provide evidence that the children actually *received* any assets from the parents. Dunn and Holtz-Eakin (2000, p. 298) found that parents' total assets had a statistically significant but quite small effect on sons' self-employment: "a \$10,000 increase in parents' total assets raises the probability of a son's transition into self-employment by 0.0009, which is small relative to both the sample transition probability of 0.031 and the impact of the son's own assets." By contrast, they noted that the parents' self-employment experience – measured simply as whether father, mother, or both were ever self-employed during the survey years – had a powerful effect, almost doubling the probability of a son's entering self-employment. Similarly, in his study of transitions to self-employment in Denmark, Sørensen (2006) concluded that parental wealth did not have a substantively significant effect on entrepreneurial entry. Based on our review, we believe that direct financial transfers are not a major cause of the association between the self-employment of parents and their children.

### *Social Capital*

If parents do not provide much in the way of financial resources to their children who are considering entrepreneurial entry, what else might they provide? Aldrich and Zimmer (1986) proposed that resources obtained via social networks could supplement or even supplant the financial and physical resources that a nascent entrepreneur controlled. Subsequently, other theorists have offered the more general concept of "social capital" to refer to advantages that people obtain through direct and indirect ties with resource providers.

For example, Sørensen (2006) argued that parents with work experience in a particular industry can pass on their valuable knowledge to their children, thus saving them some of the costs of trial and error learning. Dunn and Holtz-Eakin (2000) drew a similar inference from their finding that transition probabilities to self-employment were increased to the extent that parents were self-employed longer, had higher business assets, or higher business income than others. They also noted that because the majority of

sons entered different industries and occupations than their fathers, the expertise being passed within families was not entirely specific to a particular job or industry. Because of their research designs, neither study could directly assess the extent to which children obtained skills and knowledge from their parents as opposed to other sources during their adolescence and adulthood, such as via work experience or networks of advisors. For example, Fairlie and Robb (2005) reported that slightly more than half of all the small business owners reported that, before starting their own business, they worked in a similar business for someone else. Sørensen noted that children whose parents were self-employed were much more likely to choose their parents' industry than children whose parents were not self-employed, but having parents in the same industry made no difference in their degree of business success, as measured by exit rates or self-employment income. Thus, whereas parents might have influenced their children's choice of industry, they did not seem to actually pass on any valuable industry-specific skills to them.

Unfortunately, the studies of occupational inheritance in Table 1 contain no information about actual interactions between parents and adult children, nor have many studies – survey or ethnographic – examined parental contributions to children's businesses. The literature on family business mostly focuses on families already in business, rather than the conditions that may lead to children launching their own ventures (Aldrich & Cliff, 2003). Thus, in this section, we draw on a few selected studies to convey a sense of what might be included in future research projects.

In a series of reports on business owners in the Research Triangle Park Area of North Carolina, Aldrich and his colleagues investigated the extent to which social ties within families affected business start-ups and business practices. They found that networks spanning multiple domains of social life, beyond kinship networks, provided nascent entrepreneurs with greater access to multiple sources of information than homogeneous networks (Renzulli, Aldrich, & Moody, 2000). The higher the proportion kin in nascent entrepreneurs' networks, the *less* likely they were to subsequently start a business. They argued that the increased social support provided by family ties did not offset information lost because people relied so heavily on insiders. After nascent started their businesses, however, having a high proportion of kin in their core networks did not affect owners' abilities to mobilize resources from the core (Renzulli & Aldrich, 2005). Nonetheless, most did *not* turn to family members for help.

When owners were asked who they relied on for legal, financial, business loan, and industry expert advice, they rarely mentioned family members.

Instead, they turned to people with the necessary qualifications, rather than their parents or other kin (Aldrich, Elam, & Reese, 1996). For legal and financial assistance, and help with business loans, almost no one turned to family members (less than 5 percent of men and women for any of these three resources). Most relied on accountants for financial assistance and lawyers for legal advice, rather than kin. Finally, for expert advice, owners turned to business associates and friends, rather than family.

We have noted that rates of entrepreneurial entry remain fairly constant over most of the life courses (Evans & Leighton, 1989). For most people, years and even decades elapse between leaving home and attempting self-employment. During that time, they accumulate work experiences that both build on and supersede what they learned from their parents, such as by taking a job with managerial responsibility. Using the PSED, Kim et al. (2006) found that years of managerial experience was a strong predictor of being a nascent entrepreneur in 1999–2000. In the Vancouver study, about 53 percent of all owners' held jobs as managers before their current ownership, and parental self-employment made no difference in who held such jobs. The next-to-last job for about 54 percent of the owners also involved managerial responsibility, and again, parental self-employment made no difference. We suspect that over the life course, workers' career trajectories gradually attenuate the advantages they might have gained as children and adolescents with self-employed parents.

From a life course perspective, what happens to the children of self-employed parents during adulthood, before they attempt entrepreneurial entry, is the key to sorting out the influences we portrayed in Fig. 1. To the extent that people gain experiences during childhood and adolescence that change their career aspirations and identities, they may subsequently seek self-affirming environments that accentuate parental influences. However, adults' environments may not permit them to select contexts that ensure continuity. Instead, disruptive events may attenuate parental influences, making adult entrepreneurial entry much more contingent on proximate environmental factors. The research designs we have reviewed are not complex enough to allow us to decide between alternative interpretations.

## CONCLUSIONS AND IMPLICATIONS

Previous research has tried to explain why the sons and daughters of self-employed parents have a heightened tendency to attempt entrepreneurship (and thus become self-employed) at some point during their working

careers. As such, investigators might have framed the problem too narrowly, looking for patterns of association between two discrete states: parents' and children's self-employment. We have noted that the issue actually requires a more general consideration of the forces producing occupational inheritance. We suggested following the lead of social mobility researchers and conceptualizing the problem as one of the intergenerational continuity in classes of occupational attainment, rather than simply thinking of specific occupations. Framing the problem more generally helps us see that instead of looking only for occupation-specific skills, resources, and training, we should look for more general factors, such as occupational self-directedness, which predisposes children not only to self-employment but also to other occupations with high autonomy.

### *The Life Course Perspective*

We offered a life course perspective on work careers as a framework for integrating previous work and generating new propositions, taking a long-term view of parent-child relationships. The life course view brings two benefits to the study of occupational inheritance. First, it makes salient the role of timing, duration, and historical change in the patterning of events and roles over a person's life span. Understanding how temporal organization affects people's careers requires that we build theories about the cumulative effects of career decisions as shaped by changing historical contexts. We noted that historical influences can be classified into three types of effects (age, period, and cohort) and that untangling their separate and joint effects requires dynamic research designs.

Second, the life course perspective makes salient the special methodological requirements of studying transitions and turning points in people's lives. Inflow and outflow studies provide snapshots of linked lives across generations, but they fail to capture duration-dependent processes that are sensitive to changing historical circumstances. In [Table 1](#), we noted that investigators have used panel studies to estimate the effects of various contingencies on the likelihood of switching into self-employment from other states. However, few of these studies contained more than a handful of time-varying covariates and most used fixed cohorts, rather than adding new cohorts as the sample aged. Given the limited historical period covered by these studies, it would be advantageous to begin new transition probability studies to ascertain whether factors leading to self-employment have changed in the new millennium.



*Explaining Occupational Inheritance among the Self-Employed*

In our review, we have summarized dozens of empirical projects and offered empirical generalizations about parents’ affects on their children’s occupational attainment in three discrete life course segments: childhood, adolescence, and adulthood. For each segment, we noted the conceptual principles relevant to developments during that segment and reviewed selected studies. For our summary, we revisit those segments in reverse order, starting with adults. **Table 4** contains a list of the key propositions emerging from our review.

*Adulthood*

We think it is unlikely that more than a small portion of the association between parents’ and children’s self-employment can be accounted for by activities undertaken by adults for their adult children. We noted that few children inherit their parents’ businesses or receive any startup capital from

**Table 4.** Propositions of a Life Course Model of Parental Influences on Children’s Self-Employment.

---

Impact throughout life course
P1: Parents influence their children’s propensity to share similar occupational status, generally, and to enter self-employment, specifically, through the interaction of genetic inheritances and environmental conditions (within and between families)
Childhood
P2: Parents strongly influence their children’s propensity to enter self-employment through parenting practices that affect their children’s work values
P3: Parents moderately influence their children’s propensity to enter self-employment through parenting practices that affect their children’s awareness of vocational interests and occupational choices
Adolescence
P4: Parents moderately influence their children’s propensity to enter self-employment by reinforcing work values during adolescence developed through childhood
P5: Exposure to their parents’ occupational environment during adolescence has little effect on the likelihood that individuals with self-employed parents will enter self-employment themselves, except in traditionally structured family businesses
Adulthood
P6: Parents rarely influence their children’s propensity to enter self-employment by providing financial support during adulthood
P7: Parents slightly influence their children’s propensity to succeed in self-employment by providing advice and other tangible means of support

---

them. Extraordinary examples of second and third generation family firms have probably misled family business theorists into thinking that parent-to-adult child capital assistance is routine. We found no evidence to support such a belief. With regard to *social capital*, we found mostly speculation rather than empirical confirmation regarding the extent to which parents offer their adult children valuable entrepreneurial assistance. Indeed, parents and family seem to play a minor role in business operations by other family members.

### *Adolescence*

With regard to adolescence, we began with an assumption that much of what parents can provide to their children that might eventually affect their decisions to enter entrepreneurship lies outside the realm of parental “investment” strategies. In the section on childhood, we laid out the foundations of this argument: genetic endowments cannot be altered, child-raising practices are driven by dynamics much more short-term oriented than concern for future occupations, adolescent children are mostly consumers of financial resources rather than investment vehicles, and investments in educational attainment after high school probably represent the most forward-looking human capital planning carried out by parents. Thus, we looked for parental activities that could happen in the natural course of events, when parents interact with their adolescent and college-age children.

We noted that most adolescents’ and young adults’ first jobs are as employees, and most will spend years in this state before attempting self-employment. Studies show adolescents lack basic information about the occupations they report considering, and few engage in robust search activities to learn more. Thus, they are susceptible not only to parental influence but also to peer and media influence. For example, we reported that very high proportions of youth and young adults in recent years have expressed an interest in starting their own business, although they are quite cautious about actually following through on their interests. Perhaps self-employed parents who employ their children during their teenage years reinforce values such as occupational self-directedness that eventually lead to entrepreneurship.<sup>26</sup> However, our analysis of the PSED found little indication that parental business ownership directly influenced their children’s decisions regarding nascent entrepreneurship, at least in 1999–2000. Nonetheless, [Mortimer’s \(1976\)](#) research showed that college students’ career choices are apparently influenced, to some extent, by how they perceive their parents’ occupations. Perhaps collecting such information from survey respondents, in addition to occupational data, will shed light on this conundrum.

Our review of the few studies with information on the involvement of the children of self-employed parents in their families' businesses found that as many as half of all the adults had parents who had experienced at least one spell of self-employment in their careers. Thus, potential exposure to an entrepreneurial lifestyle has been widespread in the United States over the past half century. However, most did not work in their parents' businesses, and if they did, they worked only part time. The one exception was businesses jointly owned by both parents, which seemed to involve their children much more actively than other businesses. However, we found that involvement in parents' businesses made no difference in who became a nascent entrepreneur in 1999–2000 (Kim et al., 2006).

### *Childhood*

With regard to childhood, few studies of occupational inheritance among the self-employed have considered either studies of parental genetic contributions to their children's aptitude and personality or studies of parental socializing influence on children's occupational values and interests. A great deal of research on parenting has attributed most outcomes to the effects of parenting per se, without regard to the possibility that the outcomes result from a much more complex mix of genetic, environmental, and genetic/environmental interaction (Maccoby, 2000). Studies show that genetic variation has significant effects on a wide range of behaviors and personality traits, including occupational interests and values. We noted that the new thrust in genetic research views genetic endowments as developing within environmental contexts that can allow their full expression or hinder their development, and moreover may interact with them magnify or dampen their effects. Moreover, children and adolescents can be active in seeking out environments that allow the development of their potential. Although few genetics researchers have tackled the issue of occupational inheritance, one tantalizing study argued that about half of the between-family variance in occupational status could be attributed to genetic effects. We envision future research on occupational inheritance among the self-employed as paying much more attention to genetic endowments, to parental values as reflected in occupational working conditions, and to the interaction between genetic endowments and child-rearing environments.

## NOTES

1. One reader of our paper, Philip Cohen, noted that we make heavy use of the term "choice" in this paper, even though self-employment might better be thought of

as an “achievement” rather than a preference. We note that the literature on occupational attainment, work values, career mobility, and work more generally also uses the term “choice,” even when an analysis clearly implies that few options might have been open to people. We have tried to restrain ourselves, but trying to avoid the agentic implications of “choice” by substituting more cumbersome circumlocutions can bog a paper down. Thus, readers should be aware that we often use “choice” while being fully aware that the available options in the process we are analyzing might have been heavily constrained.

2. The occupational inheritance literature uses the term “self-employment” to characterize people who fall into the broad category of earning a living that does not depend on being employed by someone else. Within the category of the self-employed, researchers have made distinctions between self-employed professionals, skilled workers, and unskilled workers as well as between farmers and other self-employed persons (Arum & Mueller, 2004). Blau and Duncan (1967) distinguished between self-employed professionals, proprietors, and self-employed farmers. By contrast, the entrepreneurship literature typically distinguishes between entrepreneurs with employees and those without, between low- and high-growth businesses, and between innovative and non-innovative businesses (Aldrich & Ruef, 2006; Kim et al., 2006). For our purposes, we will use the generic terms “self-employed” and “entrepreneur” to cover all persons who attempt to start a business, regardless of its size or growth orientation, by themselves or with others, and who are not engaging in such activity for a third party.

3. To the extent that self-employed parents have more resources to invest in their children’s education, their children will benefit from greater acquired human capital.

4. Despite the potential for serious specification error, most empirical projects include only a few characteristics of a self-employed person’s social origins, such as parents’ occupations and education.

5. Rytina (2000) argued that prestige scores are constructed strictly on empirical grounds from two components, education and income, and it is therefore improper to equate prestige scale scores with “occupation.”

6. Although the term “potential” seems to convey an inherent positive connotation, we use it in a neutral way to mean an ability, skill, or trait that can be put to any ultimate use, positive or negative.

7. As our colleague Glen Elder reminded us, genetic effects involve not only main effects but also gene/environment interaction effects. Shanahan and Hofer (2005, p. 65) noted the complex relationship between genotypes and phenotypes introduced by gene–environment interactions, “which occur when genes alter the organism’s sensitivity to specific environmental features or environmental features exert differential control over genetic influences.”

8. Behavioral geneticists do not argue that individuals necessarily consciously seek out environments favorable to their genetic endowments. Rather, the pressures are mostly pre-conscious.

9. For a review of the strengths and weaknesses of twin studies, see Bouchard and McGue (2003). We note that twin studies give an upper bound estimate of possible genetic effects, rather than an unequivocal point estimate.

10. Researchers are able to estimate the heritable fraction of behaviors and traits using studies of monozygotic (“identical”) and dizygotic (“fraternal”) twins raised

together and apart. By making comparisons of the extent of similarity in monozygotic twins raised apart versus together, versus dizygotic twins raised apart versus together, researchers can estimate the relative impact of environmental and genetic influences.

11. Similarly, within a family, children in competition with siblings might move to a niche that best suits their temperament and skills. Some psychologists have referred to this as the principle of “niche picking.”

12. For a critique of studies of gene–context interactions, see Shanahan and Hofer (2005).

13. One explanation Lichtenstein offered for the sex difference in heritability is that the women were raised in much more constraining environments in which the tendencies latent in their genomes were suppressed. By contrast, men were raised in less constrained environments in which they were more free to pursue their natural tendencies.

14. Halaby (2003) used the 1993 follow-up study of Wisconsin high school seniors, who were first interviewed in 1957 to argue that a basic entrepreneurial versus bureaucratic dimension of work values is the key to understanding workers’ achievement motivation and mobility. However, Johnson, Mortimer, and Lee (2006) challenged his argument. In re-analysis of the Wisconsin data, plus four other data sets, Johnson et al. found that the extrinsic–intrinsic schema was a better fit to the data in samples that included men and women, as well as among women in the samples where such comparisons were possible. They expressed appreciation for Halaby’s inclusion of one specific aspect of extrinsic rewards – security – into studies of work values and suggested that researchers include “risk tolerance” in future studies of occupational choice.

15. Attrition reduced the sample size to half by the time of the final interviews. Tests for possible sample bias indicated no significant difference by sex or parental education or age for those interviewed at all six points and those who missed one or more interviews.

16. Unfortunately, parents were not asked what occupations they would like their children to have, and so Helwig could only report children’s perceptions of parents’ goals for them.

17. Youth self-employed in their teenage years were about three times as likely as others to be self-employed at age 27.

18. Six percent of the young adults said that they had already started a business.

19. In a personal communication, Jeylan Mortimer suggested that in an earlier era, self-employment conveyed more tolerance of risk than it does today, as well as a more distinctive organizational environment. With the media painting a picture of increasing turbulence among large firms – outsourcing, downsizing, mergers, temporary employment, and so forth – many workers may perceive self-employment as no more risky than being employed by a large firm.

20. Of the 62 children whose parents owned a business for 20 years or more, only nine worked in the business for 20 years or more.

21. As we might expect, the percent reporting parental self-employment is moderately related to a respondent’s age. By age cohort, the percent reporting parental self-employment is 20–29, 50 percent; 30–39, 54 percent; 40–49, 64 percent; 50–59, 58 percent, and 60–69, 56 percent.

22. We do not have information on how many years they worked in their parents' businesses.

23. In this small subset of businesses, we might expect some grown children to return to their parents' businesses to take them over, after their parents have retired. Elder and Conger (2000) noted this pattern in the case of farm families.

24. For comparison purposes, recall that in the Vancouver study, 61 percent of the respondents whose parents owned businesses worked in them, with most beginning to work in their teenage years. Very few began work in their parents' businesses in their 20s. In Fairlie and Robb's (2005) analysis of the 1992 CBO data, about 44 percent of the owners worked in a family member's business, but they did not report data on whether the work was full or part time. None of the studies listed in Table 1 provided any information on children's involvement with a parent-owned business, making it difficult for us to assess the extent to which the occupational inheritance they observed was due, in part, to on-the-job training.

25. Lentz and Laband (1990) reported that 14.2 percent of their NFIB sample had inherited their businesses, but the sample was not representative of the business population of the United States, as it contained many large firms.

26. Kourilsky and Walstad (1998) noted that a "desire to be my own boss" was the most often-mentioned reason youth and young adults wanted to start their own businesses.

## ACKNOWLEDGMENTS

We thank Philip Cohen, Amy Davis, Glen Elder, Lisa Keister, Steve Lippmann, Michael Lounsbury, Jeylan Mortimer, Linda Renzulli, Martin Ruef, Michael Shanahan, and the editors for their helpful comments on earlier drafts of this paper.

## REFERENCES

- Aldrich, H., Renzulli, L. A., & Langton, N. (1998). Passing on privilege: Resources provided by self-employed parents to their self-employed children. In K. Leicht (Ed.), *Research in stratification and mobility*, (Vol. 16, pp. 291–317). Greenwich, CT: JAI Press.
- Aldrich, H. E., & Cliff, J. E. (2003). The pervasive effects of family on entrepreneurship: Toward a family embeddedness perspective. *Journal of Business Venturing*, 18(5), 573–596.
- Aldrich, H. E., Elam, A., & Reese, P. R. (1996). Strong ties, weak ties, and strangers: Do women business owners differ from men in their use of networking to obtain assistance? In: S. Birley & I. C. MacMillan (Eds), *Entrepreneurship in a global context* (pp. 1–25). London: Routledge.
- Aldrich, H. E., & Ruef, M. (2006). *Organizations evolving*. London: Sage.
- Aldrich, H. E., & Wiedenmayer, G. (1993). From traits to rates: An ecological perspective on organizational foundings. In: J. Katz & R. H. Brockhaus (Eds), *Advances in entrepreneurship, firm emergence, and growth*, (Vol. 1, pp. 145–195). Greenwich, CT: JAI Press.

- Aldrich, H. E., & Zimmer, C. (1986). Entrepreneurship through social networks. In: D. Sexton & R. Smilor (Eds), *The art and science of entrepreneurship* (pp. 3–23). New York: Ballinger.
- Arum, R., & Mueller, W. (2004). *The reemergence of self-employment: A comparative study of self-employment dynamics and social inequality*. Princeton, NJ: Princeton University Press.
- Barak, A., Feldman, S., & Noy, A. (1991). Traditionality of children's interests as related to their parents' gender stereotypes and traditionality of occupations. *Sex Roles, 24*, 511–524.
- Bell, S. (2005). *JA enterprise poll on teens and entrepreneurship*. Colorado Springs, CO: Junior Achievement Worldwide.
- Blanchflower, D. G., & Oswald, A. J. (1998). What makes an entrepreneur? *Journal of Labor Economics, 16*(1), 26–60.
- Blau, P. M., & Duncan, O. D. (1967). *The American occupational structure*. New York: Wiley.
- Bouchard, T. J., Jr., & McGue, M. (2003). Genetic and environmental influences on human psychological differences. *Journal of Neurobiology, 54*(1), 4–45.
- Bouchard, T. J., Jr., & Loehlin, J. C. (2001). Genes, personality, and evolution. *Behavioral Genetics, 31*, 243–273.
- Bouchard, T. J., Jr., Lykken, D. T., McGue, M., Segal, N. L., & Tellegen, A. (1990). Sources of human psychological differences: The Minnesota study of twins reared apart. *Science, 250*(4978), 223–228.
- Budig, M. (2006). Intersections on the road to self-employment: Gender, family, and occupational class. *Social Forces, 84*, 2223–2239.
- Buldroft, R. A., Carmody, D. C., & Buldroft, K. A. (1996). Patterns of parental independence giving to adolescents: Variations by race, age, and gender of child. *Journal of Marriage and Family, 58*(4), 866–883.
- Burke, A. E., FitzRoy, F. R., & Nolan, M. A. (2000). When less is more: Distinguishing between entrepreneurial choice and performance. *Oxford Bulletin of Economics & Statistics, 62*(5), 565–587.
- Carroll, G. R., & Mosakowski, E. (1987). The career dynamics of self-employment. *Administrative Science Quarterly, 32*(4), 570–589.
- Chinoy, E. (1955). *Automobile workers and the American dream*. Garden City, NY: Doubleday.
- Conley, D. (2004). *The pecking order: Which siblings succeed and why?* New York: Pantheon Books.
- Conley, D., & Glauber, R. (2005). *Sibling similarity and difference in socioeconomic status: Life course and family resource effects*. New York: Center for Advanced Social Science Research.
- de Wit, G., & Van Winden, F. A. A. M. (1989). An empirical analysis of self-employment in the Netherlands. *Small Business Economics, 1*, 263–272.
- Dunn, T., & Holtz-Eakin, D. (2000). Financial capital, human capital, and the transition to self-employment: Evidence from intergenerational links. *Journal of Labor Economics, 18*(2), 282–305.
- Elder, G. H. (1999). *Children of the great depression: Social change in life experience*. Boulder, CO: Westview Press.
- Elder, G. H., & Conger, R. D. (2000). *Children of the land: Adversity and success in rural America*. Chicago: University of Chicago Press.
- Evans, D. S., & Leighton, L. S. (1989). Some empirical aspects of entrepreneurship. *American Economic Review, 79*(3), 519–535.

- Fairlie, R. W. (1999). The absence of the African-American owned business: An analysis of the dynamics of self-employment. *Journal of Labor Economics*, 17(1), 80–108.
- Fairlie, R. W., & Robb, A. (2005). *Families, human capital, and small business: Evidence from the characteristics of business owners survey*. Santa Cruz, CA: University of California.
- Gartner, W. B., Kelly, G. S., Nancy, M. C., & Paul, D. R. (Eds) (2004). *Handbook of entrepreneurial dynamics: The process of business creation in Contemporary America*. Thousand Oaks, CA: Sage.
- Guo, G., & Stearns, E. (2002). The social influences on the realization of genetic potential for intellectual development. *Social Forces*, 80(3), 881–910.
- Halaby, C. N. (2003). Where job values come from: Family and schooling background, cognitive ability, and gender. *American Sociological Review*, 63(2), 251–278.
- Helwig, A. A. (2004). A ten-year longitudinal study of the career development of students: Summary findings. *Journal of Counseling & Development*, 82(1), 49–57.
- Hout, M., & Rosen, H. S. (2000). Self-employment, family background, and race. *Journal of Human Resources*, 35(4), 671–694.
- Johnson, M. K., Mortimer, J. T., & Lee, J. C. (2006). *Dimensions of work values revisited*. Pullman, WA: Washington State University.
- Keller, L. M., Bouchard, T. J., Arvey, R. D., Segal, N. L., & Dawis, R. V. (1992). Work values: Genetic and environmental influences. *Journal of Applied Psychology*, 77(1), 79–88.
- Kim, P. H., Aldrich, H. E., & Lisa, A. K. (2006). Access (not) denied: The impact of financial, human, and cultural capital on entrepreneurial entry in the United States. *Small Business Economics*, 27, 5–22.
- Kohn, M. L., & Schooler, C. (1983). *Work and personality: An inquiry into the impact of social stratification*. Norwood, NJ: Ablex Publishing Corporation.
- Kohn, M. L., Slomczynski, K. M., & Schoenbach, C. (1986). Social stratification and the transmission of values in the family: A cross-national assessment. *Sociological Forum*, 1(1), 73–102.
- Kourilsky, M. L., & Walstad, W. B. (1998). Entrepreneurship and female youth: Knowledge, attitudes, gender differences, and educational practices. *Journal of Business Venturing*, 13(1), 77–88.
- Kourilsky, M. L., & Walstad, W. B. (2000). *The e generation: Prepared for the entrepreneurial economy?* Dubuque, IA: Kendall/Hunt Publishing.
- Lareau, A. (2002). Invisible inequality: Social class and childrearing in black families and white families. *American Sociological Review*, 67(5), 747–776.
- Lentz, B. F., & Laband, D. N. (1990). Entrepreneurial success and occupational inheritance among proprietors. *Canadian Journal of Economics*, 23(3), 563–579.
- Lichtenstein, P., Herschberger, S. L., & Pedersen, N. L. (1995). Dimensions of occupations: Genetic and environmental influences. *Journal of Biosocial Science*, 27, 193–206.
- Lichtenstein, P., & Pedersen, N. L. (1997). Does genetic variance for cognitive abilities account for genetic variance in educational achievement and occupational status? A study of twins reared apart and twins reared together. *Social Biology*, 44, 77–90.
- Lipset, S. M., & Bendix, R. (1959). *Social mobility in industrial society*. Berkeley: University of California Press.
- Lykken, D. T., Bouchard, T. J., McGue, M., & Tellegen, A. (1993). Heritability of interests: A twin study. *Journal of Applied Psychology August*, 78(4), 649–661.
- Maccoby, E. E. (2000). Parenting and its effects on children: On reading and misreading behavior genetics. *Annual Review of Psychology*, 51(1), 1–27.



- Miller, D. R., & Swanson, G. E. (1958). *The changing American parent: A study in the Detroit area*. New York: Wiley.
- Mortimer, J. T. (1974). Patterns of intergenerational occupational movements: A smallest-space analysis. *American Journal of Sociology*, 79(5), 1278–1299.
- Mortimer, J. T. (1976). Social class, work and the family: Some implications of the father's occupation for familial relationships and sons' career decisions. *Journal of Marriage and the Family*, 38(2), 241–256.
- Mortimer, J. T., Zimmer-Gembeck, M. J., Holmes, M., & Shanahan, M. J. (2002). The process of occupational decision making: Patterns during the transition to adulthood. *Journal of Vocational Behavior*, 61(3), 439–465.
- Pedersen, N. L., Spotts, E., & Kato, K. (2005). Genetic influences on midlife functioning. In: S. L. Willis & M. Martin (Eds), *Middle adulthood: A lifespan perspective*. Newbury Park, CA: Sage.
- Renzulli, L. A., & Aldrich, H. E. (2005). Who can you turn to? Tie activation within core business discussion networks. *Social Forces*, 84(1), 323–342.
- Renzulli, L., Aldrich, H. E., & Moody, J. (2000). Family matters: Consequences of personal networks for business startup and survival. *Social Forces*, 79(2), 523–546.
- Reynolds, P. D., & White, S. B. (1997). *The entrepreneurial process: Economic growth, men, women, and minorities*. Westport, CN: Quorum Books.
- Rosenfeld, R. A. (1992). Job mobility and career processes. In: J. Blake & J. Hagan (Eds), *Annual review of sociology*, (Vol. 18, pp. 39–61). Palo Alto, CA: Annual Reviews, Inc.
- Rytina, S. (1992). Scaling the intergenerational continuity of occupation: Is occupational inheritance ascriptive after all? *American Journal of Sociology*, 97(6), 1658–1688.
- Rytina, S. (2000). Is occupational mobility declining in the U.S.? *Social Forces*, 78(4), 1227–1276.
- Schneider, B. L., & Stevenson, D. (1999). *The ambitious generation: America's teenagers, motivated but directionless*. New Haven, CT: Yale University Press.
- Schooler, C., Mulatu, M. S., & Oates, G. (2004). Occupational self-direction, intellectual functioning, and self-directed orientation in older workers: Findings and implications for individuals and societies. *The American Journal of Sociology*, 110(1), 161–197.
- Shanahan, M. J., & Hofer, S. M. (2005). Social context in gene–environment interactions: Retrospect and prospect. *Journal of Gerontology*, 60B(Special Issue 1), 65–76.
- Sørensen, J. (2006). Closure vs. exposure: Mechanisms in the intergenerational transmission of self-employment. In: M. Ruef & M. Lounsbury (Eds), *Research in the sociology of organizations*. Elsevier: JAI.
- Steinmetz, G., & Wright, E. O. (1989). The fall and rise of the petty bourgeoisie: Changing patterns of self-employment in the Postwar United States. *American Journal of Sociology*, 94(5), 973–1018.
- Trice, A. D., Hughes, M. A., Odum, C., Woods, K., & McClellan, N. C. (1995). The origins of children's career aspirations: IV. Testing hypotheses from four theories. *The Career Development Quarterly*, 43(4), 307–322.
- Trice, A. D., & Knapp, L. (1992). Relationship of children's career aspirations to parents' occupations. *Journal of Genetic Psychology*, 153(3), 355–357.
- Uusitalo, R. (2001). Homo entrepreneurus? *Applied Economics*, 33(13), 1631–1638.
- Van Praag, C. M., & Cramer, J. S. (2001). The roots of entrepreneurship and labour demand: Individual ability and low risk aversion. *Economica*, 68(269), 45–62.

- Waller, N. G., Lykken, D. T., & Tellegen, A. (1995). Occupational interests, leisure time interests, and personality: Three domains or one? Findings from the Minnesota twins registry. In: R. Dawis & D. Lubinski (Eds), *Assessing individual differences in human behavior* (pp. 233–259). Palo Alto, CA: Davies-Black.
- Whiston, S. C., & Keller, B. K. (2004). The influences of the family of origin on career development: A review and analysis. *The Counseling Psychologist, 32*(4), 493–568.
- Williams, D. (2004). Youth self employment: Its nature and consequences. *Small Business Economics, 23*(4), 323–336.
- US Department of Labor. (2004). *Employee tenure summary*. Washington, DC: United States Department of Labor.

# CLOSURE AND EXPOSURE: MECHANISMS IN THE INTERGENERATIONAL TRANSMISSION OF SELF-EMPLOYMENT

Jesper B. Sørensen

## ABSTRACT

*Insights into the origins of entrepreneurial activity are gained through a study of alternative mechanisms implicated in the tendency for children of the self-employed to be substantially more likely than other children to enter into self-employment themselves. I use unique life history data to examine the impact of parental self-employment on the transition to self-employment in Denmark and assess the different mechanisms identified in the literature. The results suggest that parental role modeling is an important source of the transmission of self-employment. However, there is little evidence to suggest that children of the self-employed enter self-employment because they have privileged access to their parent's financial or social capital, or because their parents' self-employment allows them to develop superior entrepreneurial abilities.*

---

The Sociology of Entrepreneurship

Research in the Sociology of Organizations, Volume 25, 83–124

Copyright © 2007 by Elsevier Ltd.

All rights of reproduction in any form reserved

ISSN: 0733-558X/doi:10.1016/S0733-558X(06)25003-1

Family background is a strong and reliable predictor of the likelihood of engaging in entrepreneurial activity. Most importantly, numerous studies find that having self-employed parents makes children substantially more likely to enter into self-employment themselves<sup>1</sup> (Blau & Duncan, 1967; Featherman & Hauser, 1978; Hout, 1984; Wong, 1992; Western & Wright, 1994; Aldrich, Renzulli, & Langton, 1998; Hout & Rosen, 2000; Roberts, 1991). The magnitude of this effect is striking: using data on young men in the United States, for example, Dunn and Holtz-Eakin (2000) estimate that parental self-employment doubles the probability of entry into self-employment during the early career. Yet despite the voluminous evidence of the intergenerational transmission of self-employment, there remains substantial ambiguity concerning the precise mechanisms behind it (Aldrich et al., 1998).

Unpacking the mechanisms driving the intergenerational transmission of self-employment provides a unique lens on a fundamental question in the study of entrepreneurship, namely whether differences in individual propensities to engage in entrepreneurial activity reflect differences in access to entrepreneurial opportunities and resources, or differences in the desire of individuals to pursue entrepreneurial activity (Thornton, 1999). In organizational sociology, the dominant thrust in recent years has been to utilize opportunity-driven theoretical frameworks that see entrepreneurial entry as a response to environmental variability in the availability of the resources needed for entrepreneurial activity (e.g., Hannan & Freeman, 1989; Romanelli & Schoonhoven, 2001). From this perspective, whether or not someone becomes an entrepreneur is a function of their position in the organizational environment and the flow of resources to that position. Yet the rate of entrepreneurial activity may not correspond closely to the availability of entrepreneurial resources. Entry into entrepreneurship may, for some individuals, be driven more by the desire to occupy the social role of an entrepreneur than by the availability of an economically viable opportunity (Chinoy, 1955; Hamilton, 2000; Xu & Ruef, 2004). Faced with the same opportunity structure, individuals with different backgrounds and career histories may differ in their likelihood of becoming entrepreneurs. In this case, theoretical models and social policies that restrict their focus to the availability of entrepreneurial resources are inadequate.

A similar debate between the importance of resources and role aspirations in the intergenerational transmission of self-employment can be found in the literature on occupational mobility and status attainment. Broadly speaking, scholars in this literature have suggested two different types of reasons why children of the self-employed are more likely to become self-employed

themselves, which I term “exposure” and “closure” arguments, respectively. Exposure arguments draw on a large body of research highlighting the influence of parental status on socialization processes, focusing in particular on how the parents’ social position exposes children to experiences and normative expectations that have a lasting impact on their subsequent career choices (e.g., Kohn, Slomczynski, & Schoenbach, 1986; Sewell & Hauser, 1975). Exposure to and familiarity with self-employment in the family of origin may raise self-employment rates by increasing the perceived viability of self-employment as a career option (Carroll & Mosakowski, 1987) and through its influence on job values, particularly the preference for autonomous working conditions (Halaby, 2003; Benz, 2005). A second line of argument emphasizes social closure processes (Parkin, 1979), or the resources and privileges that derive from social position. Authors in this tradition attribute the transmission of self-employment to the ability of some parents to take advantage of resources derived from their social position in order to secure their children’s status (Robinson, 1984; Western & Wright, 1994). Children of the self-employed may not necessarily be more likely than other children to *try* to enter self-employment (as in exposure arguments); rather, they may be more likely to *succeed* in entering self-employment, once they do try, because their status as children of the self-employed gives them unique access to financial capital, social capital, and relevant work experiences. In other words, self-employed parents may help their children overcome the barriers to entry into self-employment, and different rates of entry into self-employment would therefore reflect unequal access to valuable resources.

Our limited understanding of the causes of transmission of self-employment, like our limited understanding of the entrepreneurial process more generally, reflects an identification problem in existing research: multiple theoretical accounts are consistent with the available empirical evidence. This identification problem grows out of the research designs typically employed in prior research. Past studies generally rely on data that measure parental self-employment status at a single point in time during the child’s life course, usually at age 16 (e.g., Blau & Duncan, 1967; Featherman & Hauser, 1978; Erikson & Goldthorpe, 1992; Sewell & Hauser, 1975). An association between parental self-employment at this time point and the subsequent likelihood of entering self-employment can be interpreted as a result of both socialization and resource transfer; the data allow no adjudication. Thus, if one holds that aspirations and work values are shaped in late adolescence and early adulthood and remain largely fixed thereafter (Johnson, 2002; Halaby, 2003), then higher rates of self-employment can be seen as a

manifestation of these early influences on children's career choices. However, parents who were self-employed when the children were 16 are also more likely to be self-employed when the children are older; similarly, parents who were not self-employed at one stage of the life course are less likely to be self-employed at a later stage. The higher rates of self-employment among children of the self-employed may therefore reflect unmeasured differences in the ability of parents to facilitate their adult child's entry into self-employment.

Even if the self-employment status of parents and children were measured after the child reached adulthood, the interpretation of any association would remain ambiguous. For example, a self-employed parent may possess the resources needed to facilitate his or her child's entry into self-employment. But even if the parent does not intervene on the child's behalf, he or she will – by virtue of occupying the position – expose the child to self-employment as a viable career option, and thereby potentially influence the child's aspirations and human capital investments.<sup>2</sup> In other words, parents cannot be in a position to transfer social resources to their children without at the same time potentially changing their children's aspirations or job values.

As this discussion suggests, disentangling the closure and exposure accounts empirically poses a substantial challenge and demands a different research design than found in prior research. In this paper, I use an unusually rich and comprehensive dataset on the Danish population to deploy a unique longitudinal research design. This design allows me to assess the empirical support for the mechanisms underlying the closure and exposure accounts. I make several analytic advances over prior research. First, and most importantly, I exploit information on the timing of parental self-employment to determine whether exposure to parental self-employment during adolescence alone is sufficient to generate increased rates of entry into self-employment later in life.<sup>3</sup> Individuals who only experience parental self-employment during adolescence and subsequently become self-employed themselves are less likely to have done so because their parents used their positional advantages to ease their entry. I am therefore able to present a test of the exposure argument that is not confounded with closure processes. Second, by enriching the set of parental characteristics considered, I am able to examine the extent to which children of the self-employed take advantage of access to their parents' financial and social capital to enter self-employment. Finally, by studying the post-entry performance of people who become self-employed, I investigate whether there is evidence consistent with the idea that the children of the self-employed acquire skills relevant to self-employment.

To anticipate the findings, my analyses suggest that closure processes play a limited role at best in the transmission of self-employment in Denmark. There is no empirical support for the claim that self-employed parents facilitate their children's entry into self-employment through wealth transfers, and only limited evidence consistent with the idea that children of the self-employed draw advantages from their parents' industry knowledge and contacts. The evidence is more generally consistent with the operation of exposure processes. Most dramatically, self-employment rates among children whose parents were only self-employed during the child's adolescence are almost equal to the rates of children whose parents were continually self-employed. Furthermore, there is little evidence to suggest that this exposure effect operates through greater entrepreneurial abilities, since children of the self-employed do not have superior performance once they enter self-employment relative to other children. Rather, the results suggest that parental role modeling plays a crucial role in generating the transmission of self-employment.

## MECHANISMS IN THE TRANSMISSION OF SELF-EMPLOYMENT

A review of the literature suggests that four distinct mechanisms are most commonly invoked as explanations for the transmission of self-employment. These mechanisms can usefully be seen as drawing directly or indirectly from two classic approaches to theorizing about intergenerational mobility and stratification processes in modern societies, although these mechanisms do not exhaust the possible arguments that could be advanced from each perspective. The exposure tradition, exemplified by status attainment research (Sewell & Hauser, 1975; Hauser, Tsai, & Sewell, 1983) and research on work values (Kohn, 1969; Kohn et al., 1986) holds that the effects of social origins are mediated through their impacts on the aspirations and job values of class incumbents. In such accounts, position in the social structure has no independent causal effect on occupational attainment; instead, class reproduction results as a by-product of class-determined life conditions and their impact, through socialization processes, on children's aspirations and values. The forces that influence aspirations and values (such as parenting styles) are not necessarily unique to particular classes, but their uneven distribution through social structure may lead to class reproduction.

The second major explanatory tradition in mobility research, exemplified by a variety of class perspectives, is more explicitly structural (e.g., Western & Wright, 1994; Robinson, 1984). Such closure perspectives (Parking, 1979)

encompass mechanisms whereby class reproduction results from the exclusion of non-class members from the positions that generate advantage. These accounts locate the source of inequality in the characteristics of the positions that people occupy in the social structure and the advantages those positions convey, and not in individual preferences and abilities. The three closure mechanisms trace the transmission of self-employment to the fact that the children of the self-employed benefit from superior access to the financial capital and social capital needed for entry into self-employment, respectively (e.g., Blau & Duncan, 1967, p. 41; Western & Wright, 1994; Aldrich et al., 1998), and from greater opportunities to develop entrepreneurial skills<sup>4</sup> (e.g., Lentz & Laband, 1990; Aldrich et al., 1998).

I discuss each of these mechanisms in turn.

### *Aspirations*

A substantial portion of the intergenerational correlation in social status can be attributed to the impact of parental characteristics on children's aspirations and values (e.g., Sewell & Hauser, 1975). Self-employed parents may serve as role models for their children (Carroll & Mosakowski, 1987; Hout, 1984). This role modeling may simply serve to legitimate or increase the child's awareness of self-employment "as a realistic alternative to conventional employment" (Carroll & Mosakowski, 1987, p. 576), and thus increase the possibility that the child will consider self-employment a viable career option later in life.<sup>5</sup> Parental self-employment may also affect children's values by shaping their basic orientation toward "what makes up 'earning a good living'" (Hout, 1984, p. 1384), leading to a preference for self-employment among children of the self-employed (Western & Wright, 1994; Aldrich et al., 1998). Furthermore, the child-rearing practices and values of self-employed parents may lead their children to value self-employment more highly than other forms of employment. The extensive work by Kohn and others (Kohn, 1969; Spenner, 1988) suggests that people whose work is characterized by high levels of self-direction develop values that emphasize the importance of self-direction and autonomy, and that parental values in turn affect children's values (Kohn et al., 1986). This suggests that children of the self-employed will value the autonomy and self-direction of self-employment more highly than children of parents with lower levels of occupational self-direction. Such a taste for autonomy may be an important factor in the decision to become self-employed. For example, the self-employed in the United States suffer a wage penalty relative



to what they could earn in paid employment (Hamilton, 2000) and accept lower risk-adjusted returns on their entrepreneurial investments (Moskowitz & Vissing-Jørgensen, 2002). Both of these financial penalties can be viewed as a price that people pay for the autonomy of self-employment. Furthermore, Bemz and Frey (2003), using cross-national surveys, find that autonomous work conditions explain the greater job satisfaction of the self-employed relative to the employed.

Halaby's (2003) analysis of data from the Wisconsin Longitudinal Survey (WLS) provides some support for this mechanism. The WLS has tracked a sample from the 1957 cohort of Wisconsin high school seniors, including a survey of job values performed in 1993. Halaby classifies some of these job values as indicating a more entrepreneurial orientation and others as indicating a more bureaucratic orientation (Miller & Swanson, 1958). His analyses suggest that the children of fathers who were self-employed in 1957 are more likely to express entrepreneurial job values in 1993 than other children. For example, grown children of the self-employed are significantly more likely to express a preference for jobs offering autonomy relative to jobs offering a pension. However, while Halaby (2003) focuses on the determinants of entrepreneurial job values, he does not focus explicitly on the process of entry into self-employment. It is therefore unclear whether the inheritance of self-employment is mediated by job values. Furthermore, the fact that the WLS did not collect information on job values prior to 1993 calls the causal ordering of the relationship between father's self-employment and child's job values into question; parental self-employment may have led children to enter self-employment and then subsequently develop entrepreneurial job values.

### *Entrepreneurial Skills*

More so than sociologists, economists have emphasized the consequences of exposure to parental self-employment during childhood and adolescence for the development of human capital, particularly the broad portfolio of skills relevant to self-employment. Because the potential returns to an entrepreneurial opportunity depend in part on the entrepreneur's ability, those with greater entrepreneurial skills can expect higher returns on average and should therefore be more likely to enter into self-employment (Dunn & Holtz-Eakin, 2000). Some of these skills may be acquired through observation of the parent engaged in self-employment, but the more commonly emphasized channel is through work experience in the family firm

(Carroll & Mosakowski, 1987). For example, in their study of Canadian entrepreneurs, Aldrich et al. (1998) found that 61% of those whose parents had been self-employed had worked in their parents' business; over half of these had started working for their parents before they were 15.<sup>6</sup>

Lentz and Laband (1990, p. 564) argue that children of the self-employed acquire valuable human capital about running a business

as a by-product of growing up. By the time he reaches the age of, say, eighteen, when most other youths his age are just starting to acquire job-specific skills ... or more general occupational skills ... the son of a proprietor normally has already had an opportunity to acquire the equivalent of an integrated, managerial education.

Similarly, Dunn and Holtz-Eakin (2000) suggest that because the children of the self-employed benefit from the transfer of "family-specific capital" and skills, they have greater entrepreneurial abilities than children whose parents are not self-employed. As a result of these differences in skill acquisition, the higher rates of self-employment among children of the self-employed may be explained by their superior expected returns from entrepreneurial opportunities.

### *Financial Capital*

Many sociologists have argued that the transfer of property or capital lies at the root of the inheritance of self-employment, either through the direct transfer of an ongoing business from one generation to the next, or through the ability of self-employed parents to fund their children's ventures. For example, Western and Wright (1994, p. 611) argue that "parental ownership of property is ... 'insurance' against downward mobility into wage labor for the offspring of capitalists, and the requirement of capital ownership is a barrier to entry for the children of most employees." Similarly, in seeking to explain the intergenerational reproduction of class advantage, Robinson (1984, p. 183) argues that "parents can simply give or will their business to their children or provide them with investment capital to start their own business." Hout (1984, p. 1385) argues that "although most men would like to help their sons, ready access to cash and credit is one of the fruits of autonomy," making it easier for self-employed parents to help their children enter self-employment<sup>7</sup> (see also Dunn & Holtz-Eakin, 2000).

The simplest means by which parents can facilitate their children's entry into self-employment is by allowing their children to take over their ongoing ventures. However, this accounts for only a very small proportion of

the intergenerational inheritance of self-employment; for evidence, see below as well as Aldrich et al. (1998). In the absence of this type of transfer, the ability of parents to provide capital to their children should depend on their wealth. The financial transfer hypothesis therefore implies that self-employment rates should be a positive function of parental wealth.

### *Social Capital*

While some analyses of entry into self-employment demonstrate that potential entrepreneurs face capital constraints (Evans & Leighton, 1989; Holtz-Eakin, Joulfaian, & Rosen, 1994; Dunn & Holtz-Eakin, 2000), these constraints are not sufficiently binding to make access to financial capital a necessary condition for entry into self-employment (Hurst & Lusardi, 2004; Aldrich et al., 1998; Kim, Aldrich, & Keister, 2006). More importantly, access to capital is by no means a sufficient condition for entry into self-employment. Successful entrepreneurial entry also requires the recognition of valuable entrepreneurial opportunities, and the ability to mobilize resources other than money to take advantage of them (Stinchcombe, 1965; Aldrich & Zimmer, 1986). Herein lies an additional potential source of closure fostering the intergenerational transmission of self-employment, namely the social capital of self-employed parents. The children of self-employed parents may enter self-employment at higher rates because they have better access – through their parents – to knowledge about entrepreneurial opportunities. By virtue of the social networks developed through their own entrepreneurial activity, self-employed parents may be exposed to more and/or better information about new market opportunities. Even if self-employed parents are not exposed to more or better entrepreneurial opportunities than conventionally employed parents, self-employed parents may be better able to assess the attractiveness of the opportunities, and pass that knowledge along to their children. Furthermore, self-employed parents may be better able to provide their children with referrals to suppliers of critical resources as well as potential customers, thereby easing the transition to entrepreneurship.

All four of these mechanisms imply a positive correlation between rates of self-employment and parental self-employment status. Progress in assessing the merits of the different mechanisms in a given empirical context therefore requires deriving and testing additional implications of the different mechanisms. I discuss the analytic strategy next, along with the implied hypotheses.

## HYPOTHESES

I begin with a strategy for identifying an effect of parental self-employment that does not confound the potential effects of exposure processes with the closure mechanisms. The key to doing so is to exploit data on the timing of parental self-employment relative to the child’s life course, because the range of parental influences depends on the timing of parental self-employment. Conventional mobility studies measure parental status at a single point in time (typically late adolescence) and its correlation with the child’s status at some later point in time. This measurement strategy obscures heterogeneity in parental career trajectories subsequent to the child’s adolescence; propelled by the imperatives of their own careers, parents may move in and out of different social positions while their children build their own careers. If the primary source of intergenerational immobility lies in the lasting effects of parental status on children’s aspirations and skills, and parental influence diminishes rapidly after adolescence, then the fact that parents may change status later in life is largely irrelevant. If, by contrast, intergenerational immobility is a result of parents exploiting their positional resources to benefit their children, then whether or not parents occupy a position at a given point in time is a critical issue. In particular, parents must occupy the position in question when their children are seeking entry.

Information on the timing of parental self-employment can therefore be used to gain a better understanding of the nature of intergenerational influence. Fig. 1 identifies four different simplified trajectories of parental self-employment during the child’s life course.<sup>8</sup> The rows of Fig. 1 differentiate parents according to their self-employment status before the child completed compulsory schooling, and hence before the child was at risk of entering self-employment (termed “adolescence” for convenience). During this period, children are socialized by their parents and might work in their parents’ business. The columns classify parents by their self-employment status after the child has completed compulsory schooling and is capable of entering self-employment (termed “adulthood”). Parental self-employment during this period may also influence children’s aspirations and skills,

		Child’s adulthood	
		Yes	No
Child’s adolescence	Parent Self-employed? Yes	<i>A</i>	<i>B</i>
	No	<i>C</i>	<i>D</i>

Fig. 1. Potential Combinations of Parental Self-Employment during Different Life Stages of the Child.

although the parental impact on aspirations and work values is weaker than earlier in the life course (Vollebergh, Idema, & Raaijmakers, 2001). More importantly for analytic purposes, the fact that the children are at risk of entering self-employment during this stage of the life course opens up the possibility that self-employed parents can ease their entry by transferring financial and social capital, or by giving their children relevant work experience.<sup>9</sup>

The cells on the main diagonal of Fig. 1 represent the simplest parental trajectories, where parents have been either continuously self-employed or never self-employed during the child's life course. Higher rates of self-employment by children of continuously self-employed parents may reflect either the influence of exposure processes, or be the consequence of positional advantage, or both. The more analytically interesting trajectories are found in the off-diagonal cells. The lower-left-hand cell contains parents who were not self-employed while their children were growing up, but who entered self-employment subsequent to their children's completion of compulsory schooling. These parents could in principle facilitate their children's entry into self-employment through the transfer of financial and social capital. However, because their entry into self-employment may reflect a long-held ambition, these parents may also have caused their children to acquire entrepreneurial job values and aspirations. This cell therefore does not allow one to cleanly adjudicate between accounts emphasizing the role of exposure to parental self-employment and accounts emphasizing positional advantage.

Greater analytic traction can be gained from the upper-right-hand cell in Fig. 1, which contains parents who were self-employed as their children were maturing, but who subsequently exited self-employment. Children in this cell may acquire entrepreneurial aspirations and skills while growing up, but when they reach adulthood their parents, having left self-employment, can no longer exploit their positional resources to benefit their children.<sup>10</sup> If exposure to parental self-employment is not implicated in the transmission of self-employment, the entry rates of these children should be the same as those whose parents were never self-employed. If their rates of entry are higher, it suggests that exposure to parental self-employment leads to entrepreneurial aspirations and skills. Because the parents subsequently leave self-employment, any effect is less likely to be the result of the exercise of positional advantage by self-employed parents.

**Hypothesis 1.** Self-employment rates will be higher if parents were self-employed during the child's adolescence alone than if parents were never self-employed.

Support for Hypothesis 1 would suggest that exposure processes contribute to the transmission of self-employment, but would leave two questions unresolved. First, such evidence does not speak to the empirical merits of the closure mechanisms; resource transfers may complement exposure processes in generating the transmission of self-employment. Second, evidence consistent with Hypothesis 1 would not speak to whether the effect is due to the impact of exposure on aspirations and work values, or to the acquisition of entrepreneurial skills, or both.

The financial capital mechanism can be examined by estimating the effect of parental wealth on the transition to self-employment. Evidence for such an effect is mixed. Using a sample of entrepreneurs, Aldrich et al. (1998) find no evidence to suggest that parental wealth was an important determinant of the likelihood of entering entrepreneurship (see also Kim et al., 2006). In the most careful study of this mechanism, Dunn and Holtz-Eakin (2000) use data from the National Longitudinal Survey of Market Experience (NLS) and find that parental assets have a positive effect on the self-employment rates of sons. However, the effect is substantively small, with a \$10,000 increase in parental assets (measured in constant 1982–1984 dollars) raising the probability of entry into self-employment by 0.0009, relative to an average transition probability in the sample of 0.031 (Dunn & Holtz-Eakin, 2000, p. 298). The effect of parental assets is reduced slightly when a dummy variable for parental self-employment is included in the model. However, the effect of parental self-employment in Dunn and Holtz-Eakin's models is strong and positive after controlling for parental assets, suggesting that parental assets do not account for the intergenerational inheritance of self-employment.

Dunn and Holtz-Eakin's analyses can be considered a test of a weak-form version of the closure argument because they only estimate a main effect of parental assets.<sup>11</sup> This is consistent with the capital transfer hypothesis, provided that self-employed parents have greater wealth on average than non-self-employed parents. In their data, self-employed parents do have higher assets, particularly when business assets are included. However, a stronger test of exclusionary behavior by self-employed parents would be to interact parental assets with parental self-employment. A positive interaction effect would indicate that self-employed parents are more likely to use their assets to help their children enter self-employment than are non-self-employed parents.

This reasoning suggests two hypotheses:

**Hypothesis 2.** The rate of entry into self-employment will be a positive function of parental assets.

**Hypothesis 3.** The positive effect of parental assets on the rate of entry into self-employment will be stronger if the parents are self-employed.

Parental social capital is considerably more difficult to measure than financial assets, particularly when one considers that an appropriate measure would have to capture those aspects of parental social networks and reputation that are relevant to entrepreneurial activity. I therefore rely on an indirect test of the social capital mechanism. If it is the case that self-employed parents use their business knowledge and contacts to help their children enter self-employment, then the children of the self-employed should be more likely to start a new venture in the same industry as their parents than they would be to start a venture in a different industry. While it is unlikely that the parents' contacts are focused exclusively around their own industry, it is reasonable to assume that their social networks are more concentrated around their own industry and have greater potential value there. Particularly, the ability of parents to identify and evaluate new entrepreneurial opportunities should be greatest in the industry in which they themselves work. This reasoning suggests that if the social capital mechanism is operating, then the children of self-employed parents should be more likely to enter the same industry as their parents than they would be to enter self-employment in a different industry, where their parents' social capital will be less useful.

**Hypothesis 4.** Parental self-employment makes children more likely to choose self-employment in the same industry as their parent than self-employment in a different industry.

Finally, one way to differentiate between the effects of entrepreneurial aspirations and skills acquired in childhood – in the absence of detailed data on either – is to shift the analytic focus away from the transition to self-employment and instead focus on the relative performance of the new ventures. In particular, if it is true that the children of the self-employed are more likely to enter self-employment because of their greater entrepreneurial ability, it follows that their performance while self-employed should be superior, on average, to the performance of self-employed individuals whose parents were never self-employed.

**Hypothesis 5.** Individuals with self-employed parents will be more successful in self-employment than individuals whose non-self-employed parents.

Aspirations could also lead to superior performance, if children of the self-employed are more committed to self-employment and therefore work harder. However, if there is no performance difference between the ventures of children of the self-employed and other children, it is difficult to argue that the children of the self-employed have acquired any skills. Such a null result would not, however, be inconsistent with the aspiration mechanism.

## DATA AND METHODS

Disentangling the roles of the four mechanisms requires data not only on individual work histories during adulthood coupled with information on parental characteristics, but also data on parental characteristics prior to the child's entry into the labor market. Moreover, to avoid selecting on the dependent variable, individuals must be observed prior to entry into self-employment, and the data must allow the observation of transitions to self-employment from any of a variety of states, including employment, unemployment, schooling, etc. Finally, because the transition to self-employment is a rare event, a large sample size is needed to provide sufficient statistical power. Prior research on the transmission of self-employment in the United States has relied on matching data on parents and children from the NLS (Dunn & Holtz-Eakin, 2000), but the resulting sample sizes are too small to reliably estimate the effects of different parental trajectories of self-employment.

With these constraints in mind, I turn to an unusually rich dataset describing the population of Denmark between 1980 and 1997. The Integrated Database for Labor Market Research, referred to by its Danish acronym, IDA (Integreret Database for Arbejdsmarkedsforskning), is drawn from population registers maintained by the Danish government. IDA has several advantages. First, it is comprehensive: all people (legally) living in Denmark in a given year are included in the registers. Individual characteristics are recorded in IDA on an annual basis, based on each individual's status in the 48th week of a given calendar year. IDA therefore amounts to an annual census of the population of Denmark. Second, IDA is longitudinal, with annual observations starting in 1980; as a result, it contains panel data for individuals. Third, IDA covers a wide range of phenomena, especially with respect to labor market outcomes. Fourth, and most importantly, the design of IDA allows individuals to be linked according to a variety of relevant characteristics. For example, children can be linked to their parents; in addition, employees can be linked to their employers.



The comprehensive character of IDA makes confidentiality a primary concern, so Statistics Denmark (which maintains IDA) restricts access to IDA in a variety of ways. Most relevant in this context is the fact that researchers are not allowed unfettered access to IDA, but must instead request particular extracts from the larger database. The analyses reported in this paper come from a special extract commissioned as part of a larger research project designed to examine a variety of issues related to the dynamics of self-employment and entrepreneurship. This extract was created by identifying all individuals who were living in Denmark in 1994 and were between the ages of 15 and 74. This is a population of 3.9 million individuals. For all of these individuals, relevant information from IDA was collected for each year from 1980 until 1997. It should be noted that this design means that the extract only covers the entire population in 1994. In other years, the extract does not capture people who were not in the population in 1994, for example because they died or emigrated before 1994. This type of attrition is more serious for older cohorts, and does not pose concerns for the current analyses.

Data on individuals in IDA are truncated prior to 1980. This truncation is especially problematic with respect to analyzing entry into self-employment, since those who have entered self-employment once are substantially more likely to do so again. Left-truncation in 1980 means there is no information on prior self-employment experience for people who were already in the labor market in 1980. Again, this limitation poses the greatest problems for older cohorts, since it is more likely that relevant career history information will be unobserved. As will be apparent, this left-truncation does not pose a problem in the analyses presented here, since individuals are sampled prior to labor market entry.

IDA contains rich demographic information and a wealth of labor market variables measured on a yearly basis. Demographic variables include age, sex, marital status, number of children, number of siblings in family of origin, birth order in family of origin, current school enrollment, and a highest educational level achieved. Current occupation for employees is recorded in broad categories, coded here as white collar vs. blue collar. Annual salary is recorded in constant 1980 Danish kroner, as is non-salary income. All income information comes from tax records. Furthermore, because Denmark had a wealth tax until 1996, individual tax records contain information on personal assets and liabilities. From this information, I created a measure of family net wealth by adding together the focal individual's net wealth and the net wealth of his or her spouse, if present.

Employees in IDA are linked to their primary workplace in a given year; workplaces are in turn linked to firms. Workplaces and firms are assigned unique identification numbers that are constant over time. Firms are in turn classified by industry using a 111-category aggregation of the ISIC (rev. 2) classification scheme. This workplace information allows me to calculate variables measuring industry tenure and tenure with an employer. In addition, I calculated a labor force experience measure by determining the age at which an individual first had an annual income exceeding 15,000 Danish kroner (constant 1980 values).<sup>12</sup>

Each person's record in the data extracted from IDA contains limited time-varying information on their parents, specifically father's occupation, mother's occupation, father's wealth, and mother's wealth. Where both parents are present, I compute their combined wealth. The parental occupation variables are used to create measures of parental self-employment, as discussed below.

### *Sample*

I selected for analysis all Danish-born children born in the years 1966, 1967, and 1968, who were residing in Denmark in 1994.<sup>13</sup> These children were therefore adolescents aged 12–14 in 1980, the first year of data in the IDA database. This resulting sample contains 228,372 individuals, of which the slight majority (51.2%) are male. These individuals are observed annually (contingent on being alive and resident in Denmark in a given year) until 1997.<sup>14</sup> As a result, the dataset consists of life histories for three cohorts of adolescents observed until they are aged 29–31; in short, the data cover their early career histories. The choice of these birth cohorts was dictated by a tradeoff between wishing to observe children both when their values are susceptible to parental influence and prior to labor market entry, on the one hand, and the need to observe sufficiently long career histories to capture an appreciable number of transitions to self-employment. Research suggests that parents have their greatest influence on children's values during early-to-mid adolescence, and that the influence of parents declines after the age of 16 (Vollebergh et al., 2001). Furthermore, the fact that children are observed at age 16 is consistent with the design of the standard occupational mobility studies. While it would have been possible, by selecting from later birth cohorts, to observe children at younger ages, these children would have less labor market experience before the data are right-censored.

### *Definition of Self-Employment*

I measure transitions to self-employment to include only transitions in which the self-employed individual employed other workers. The primary way of identifying transitions to self-employment is through the occupational codes assigned by Statistics Denmark. These codes make a distinction between two kinds of owners of privately held firms: those that are employers and those that have no employees but pay value-added tax (VAT). This latter category is very heterogeneous, largely because paying VAT is the minimum requirement for starting or running a firm. But this category also includes many people that one might think of as independent contractors instead of entrepreneurs, and may capture individuals with marginal labor force attachments. Raising financial capital is less likely to be an issue for independent contractors, so the inclusion of these transitions would bias the analysis against the closure mechanisms. Furthermore, of the two categories of private self-employment, entry into VAT-payer status is more likely to be driven by “push” factors such as poor prospects for paid employment. I therefore did not define transitions to VAT-payer status as transitions to self-employment.<sup>15</sup> However, separate transition-rate analyses that included VAT-payer status yielded very similar results to the estimates presented here (results available on request).

Statistics Denmark only assigns self-employment codes to individuals whose ventures are privately held or unincorporated. Entrepreneurs who start new incorporated ventures can only be identified indirectly since bureaucratic restrictions prevent the identification in IDA of the individuals filing for incorporation. However, the founders of an incorporated venture are typically among the employees of the new venture in the initial years. To identify these transitions, I therefore proceeded as follows. First, I identified new employers by comparing sequential years of the employer and workplace files. Second, these workplace identification numbers are matched against the individual data to identify all people who worked for a firm in November of the first year of its existence. I then examined the occupational codes assigned to these employees and identify those employees who are top managers. These individuals were considered the founders of the new venture.<sup>16</sup> This means that one venture may include several people entering self-employment at the same time.

Finally, I have attempted to exclude cases of direct inheritance of ongoing ventures. This can only be done indirectly in IDA. For direct inheritance to occur, the child’s entry into self-employment has to occur at the same time as the parent’s exit, and the child’s industry of self-employment must be the

same as the parent's. I treat apparent transitions to self-employment that met these criteria as censored. This is a conservative standard, since these events could both occur without the parent transferring a venture to the child. By these criteria, direct inheritance accounts for only a small proportion of self-employment transitions. Among the transitions to self-employment by children of the self-employed, 7.9% may be due to direct inheritance.<sup>17</sup> Direct inheritance does not explain the inheritance of self-employment in Denmark.

### *Definition of Parental Variables*

Yearly information on the father's and mother's occupations was used to determine the self-employment status of the parents. In addition to the aforementioned distinction between employers and VAT payers, the occupational coding scheme identifies a variety of other types of self-employment, including assisting spouses, "time-limited" self-employed, self-employed with unemployment insurance, and miscellaneous self-employed. These additional categories of self-employment account for a small proportion total self-employment.<sup>18</sup> I coded parents as being self-employed in a given year if their occupational code indicated any of these forms of private self-employment.

This definition of parental self-employment does not capture parents who found and run incorporated ventures. As noted earlier, in IDA one can only identify the founders of incorporated ventures indirectly, by assuming that the founders are among the top managers of the venture in its first year of existence. Parents may have already been self-employed when the children first come under observation in 1980, and so it is impossible to identify these examples. This implies that some parents who are considered employees in these analyses are actually self-employed.

Table 1 contains information on the parental trajectories of self-employment experienced by the three cohorts. In this table, and the subsequent analyses, the analytic break between adolescence and adulthood corresponds to the end of compulsory schooling at age 16, when children become at risk of entering self-employment. Almost 66% of the sample was never exposed to parental self-employment between entering the sample in early adolescence and leaving the sample because of entry into self-employment or censoring. Twenty-one percent had parents who were self-employed during both adolescence and adulthood, while only 3% were exposed to parental self-employment during adolescence alone.

**Table 1.** Observed Trajectories of Parental Self-Employment.

		Child's Adulthood	
		Yes	No
Child's adolescence	Parent Self-Employed? Yes	48,595 <i>21.3%</i>	6,662 <i>2.9%</i>
	No	22,692 <i>9.9%</i>	150,423 <i>65.9%</i>

*Note:* Values in italics are cell percentages.

### *Definition of Risk Set*

For the analyses, I use a very simple definition of the risk set, i.e., of the population that could possibly enter into self-employment. The most important consideration is the structure of the educational system. In Denmark, schooling is compulsory until the completion of the 9th grade, at which point children are 16 years of age. After the completion of 9th grade, children can choose to continue their education in a variety of ways, or they may end their schooling.<sup>19</sup> I treat the end of compulsory schooling as the point at which children become formally at risk of entering into self-employment. By contrast, [Dunn and Holtz-Eakin \(2000\)](#) consider individuals at risk in the first year following their last reported enrollment in school. However, post-compulsory schooling is best seen as a competing risk to entry into self-employment; the attractiveness of continued schooling is presumably affected by the relative attractiveness of self-employment. Defining the risk set as all those individuals who have completed compulsory schooling avoids this endogeneity. In the models estimated below, I control for current employment status and for current school enrollment.<sup>20</sup>

People leave the risk set either at the time of censoring in 1997 or after their initial entry into non-agricultural self-employment. The dependent variable is therefore the rate of initial entry into self-employment. People who have entered and exited self-employment are substantially more likely to attempt to re-enter self-employment. Such serial entrepreneurship is likely due in part to a set of mechanisms caused by the experience of self-employment, such as the development of entrepreneurial networks, reputation and skills as well as the possible strengthening of entrepreneurial job values. This suggests that the analysis of initial entry into self-employment should not be confounded with the analysis of subsequent entry. While it

would be interesting to examine the potential moderating effects of parental self-employment on these factors, such as analysis is beyond the scope of this paper.

### *Methods*

I analyze the hazard rate of entry into non-agricultural self-employment using discrete-time event history methods. Specifically, I estimate logistic regression models where the dependent variable is a dummy variable indicating entry into self-employment. I include dummy variables for distinct two-year intervals to allow the baseline hazard rate to vary with duration at risk.<sup>21</sup> All time-varying covariates are lagged one year and updated on an annual basis.

## RESULTS

Fig. 2 presents Kaplan–Meier estimates of the survivor function for first entry into non-agricultural self-employment. Of the 228,372 individuals in the three birth cohorts, a total of 4,399, or 1.9%, became employers at least once during the observation period. Of these transitions, 475 were cases of entry



Fig. 2. Kaplan–Meier Survivor Plot of First Entry into Self-Employment by Parental Self-Employment History.

into self-employment in the primary sector (agriculture, fishing), so the proportion of the sample that ever entered into non-agricultural self-employment was 1.7%. Fig. 2 clearly demonstrates the impact of parental self-employment on the entry rate. The survivor function for children whose parents have never been self-employed is considerably flatter than for children whose parents were ever self-employed, reflecting a substantially lower rate of self-employment.<sup>22</sup> The clear separation of the confidence intervals reflects the fact that these differences are statistically significant. Separate analyses by sex (not shown here) show that while men are more likely to become self-employed, parental self-employment has a dramatic effect for both sexes. Among men, the rate more than doubles from 1.3 to 2.8 transitions per 1,000 person-years at risk; among women, the rate increases by over 80%.

Having established that the children of self-employed parents differ from other children in their propensity to enter into self-employment, I turn now to a consideration of the four hypothesized mechanisms.<sup>23</sup> Fig. 3 presents a simple test of Hypothesis 1 by contrasting individuals whose exposure to parental self-employment was limited to their adolescence with those whose parents were never self-employed. If exposure processes are implicated in the transmission of self-employment, then these children of the self-employed should have higher rates of entry (and steeper survival curves). Moreover, because these parents left self-employment before their children could try to



Fig. 3. Kaplan-Meier Survivor Plot of First Entry into Self-Employment by Parental Self-Employment.

enter, closure processes are unlikely to be implicated in such an effect. The evidence in Fig. 3 is consistent with Hypothesis 1: exposure to self-employment during adolescence alone is sufficient to generate higher rates of entry into self-employment.

Multivariate models of the transition to self-employment corroborate this finding. Table 2 contains estimates of the effects of parental self-employment from a series of logistic regression models of entry into non-agricultural self-employment that control for a host of demographic and labor market variables. To account for historical variations in entry rates, I control for the state of the economy (Gross National Product and GNP growth) and average consumer sentiment as measured by consumer surveys performed by Statistics Denmark. I also control for demographics characteristics of the family of origin (number of siblings in the family of origin, whether a person was the first-born child, and a dummy variable if both parents are deceased) and the family of procreation (marital status, number of children, spousal employment status); schooling (a dummy variable for current school enrollment, and time-varying measures of the highest education completed); labor market variables (occupation, a summary measure of the degree of unemployment experienced over the past year, wage experience, industry experience, firm tenure, and log annual salary), and non-salary income. Finally, I include (logged) three-year moving averages of family assets and family debts computed from information on the wealth of individuals and their spouses if present.<sup>24</sup>

Table 2 contains estimates of the effects of parental self-employment during different stages of the child's life course. As expected, parental self-employment at any stage has a substantial impact on the transition rate. The estimates in the last column of Table 2 capture the effects of the different trajectories of parental self-employment identified in Fig. 1 with three dummy variables representing the different trajectories of parental self-employment. Hypothesis 1 is again supported: net of a host of relevant control variables, exposure to parental self-employment during adolescence alone is sufficient to generate a substantial increase in the rate of self-employment. The point estimate suggests that individuals exposed to parental self-employment during adolescence alone have a 53% ( $\exp(0.422)$ ) higher rate of entry into self-employment than individuals whose parents were never self-employed.

The estimates in Table 2 also demonstrate that exposure to parental self-employment during adolescence is not a necessary condition for generating higher rates of entry. Individuals who were adults by the time their parents entered self-employment have an entry rate that is 67% ( $\exp(0.517)$ ) higher



**Table 2.** Logistic Regression Estimates of the Rate of First Entry into Non-Agricultural Self-Employment.

Variable	(1)	(2)	(3)	(4)
In school	0.002** (0.000)	0.002** (0.000)	0.002** (0.000)	0.002** (0.000)
Highest education: vocational	0.086** (0.011)	0.086** (0.011)	0.085** (0.011)	0.083** (0.011)
Highest education: academic	0.050** (0.008)	0.050** (0.008)	0.049** (0.008)	0.049** (0.008)
Highest education: university	-0.061** (0.010)	-0.061** (0.010)	-0.060** (0.010)	-0.060** (0.010)
White collar	-0.237** (0.072)	-0.238** (0.072)	-0.233** (0.071)	-0.235** (0.071)
Blue collar	-0.362* (0.161)	-0.360* (0.161)	-0.366* (0.161)	-0.364* (0.161)
Degree of unemployment	-0.347 (0.269)	-0.335 (0.269)	-0.366 (0.269)	-0.356 (0.269)
Log salary	0.356 (0.263)	0.355 (0.263)	0.360 (0.263)	0.354 (0.263)
Non-salary income	0.317 (0.369)	0.306 (0.369)	0.318 (0.369)	0.315 (0.369)
Wage experience	4.622* (1.987)	4.672* (1.987)	4.414* (1.986)	4.414* (1.986)
Industry experience	-0.027** (0.007)	-0.027** (0.007)	-0.027** (0.007)	-0.027** (0.007)
Firm tenure	0.036** (0.012)	0.036** (0.012)	0.036** (0.012)	0.036** (0.012)
No parents	0.460** (0.141)	0.476** (0.140)	0.499** (0.140)	0.508** (0.141)
Log family assets	0.117** (0.009)	0.117** (0.009)	0.104** (0.009)	0.103** (0.008)
Log family debts	0.062** (0.006)	0.061** (0.006)	0.067** (0.006)	0.066** (0.006)
Parental self-employment during child's				
Adolescence only	0.197* (0.088)			0.422** (0.089)
Adulthood only		0.320** (0.054)		0.517** (0.056)
Adolescence and adulthood			0.523** (0.036)	0.612** (0.038)
Log-likelihood	-25,971	-25,941	-25,934	-25,930

Note: Two-sided *t*-tests: \*  $p < 0.05$ ; \*\*  $p < 0.01$ .

All models include controls for GNP and GNP growth, consumer sentiment, age (dummies for two-year intervals), sex, and a host of demographic variables interacted with sex: characteristics of the family of origin (*N* of siblings and being a first-born child), marital status, number of children, and spousal labor market status. Results available on request.

than individuals whose parents were never self-employed. While this effect is somewhat larger than the effect for exposure during adolescence alone, this difference is not statistically reliable ( $\chi^2 = 0.89, p < 0.35$ ). Finally, having continuously self-employed parents results in the highest entry rate, increasing the probability of entering self-employment by 84% ( $\exp(0.612)$ ). The effect of continuous parental self-employment is statistically different from the effect of parental self-employment during adolescence alone ( $\chi^2 = 4.31, p < 0.04$ ), but not statistically different from self-employment during adulthood alone ( $\chi^2 = 2.63, p < 0.11$ ).

While this evidence seems generally consistent with Hypothesis 1, there is a simple objection from the standpoint of the closure arguments: it seems unreasonable to assume that any positional advantages that parents derive from self-employment should disappear immediately upon their departure from self-employment. For example, parents may leave self-employment because they have been very successful and accumulated substantial wealth; this would put them in a position to help their children start their own ventures. Similarly, the social capital that the parents accumulated during self-employment likely does not disappear immediately upon their exit.

I investigate this issue in several ways. First, Fig. 4 presents information on the wealth of parents in the sample according to the trajectory of

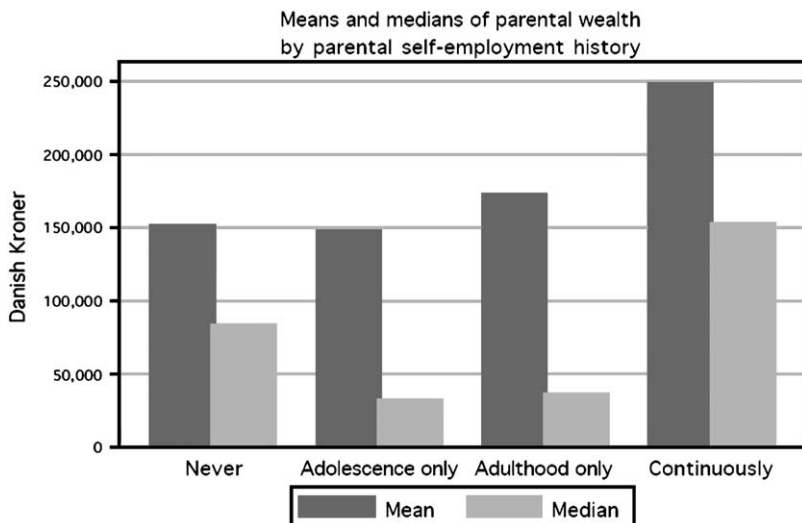


Fig. 4. Means and Medians of Parental Wealth by Parental Self-Employment History.

parental self-employment. I use the three-year moving average of the sum of the mother's and father's wealth as reported on their tax returns. The mean and median values of parental wealth are calculated over the time period when the children were in the risk set for entering into self-employment. There is little evidence to suggest that the parents who left self-employment after their child's adolescence did so because they had amassed substantial resources. Their mean wealth is approximately the same as parents who were never self-employed, while the median wealth is approximately half as large. Parents who are continuously self-employed do however appear to be more successful, with higher mean and median wealth. It is difficult to argue, based on these data, that the higher transition rate among people exposed to self-employment during adolescence alone can be attributed to the transfer of financial capital.

Table 3 further investigates whether the parents who left self-employment after the child's adolescence are likely to be in a position to pass on resources to their children. Table 3 characterizes these parents by their

**Table 3.** Destinations of Parents Who Left Self-Employment Before Child's Adulthood.

	Same Industry	Different Industry	Total
Upper white collar	216 2.9%	511 7.0%	727 9.9%
Lower white collar	305 4.2%	1,392 18.9%	1,697 23.1%
Skilled blue collar	131 1.8%	434 5.9%	565 7.7%
Unskilled blue collar	131 1.8%	1,290 17.6%	1,421 19.3%
Employed n.e.c.	101 1.4%	355 4.8%	456 6.2%
Unemployed			1,201 16.3%
NILF			947 12.9%
Retired			332 4.5%
	884 12.0%	3,982 88.0%	7,346 100%

*Note:* Percentage figures are percent of total parental transitions from self-employment (7,346). Number of parental transitions is higher than 6,662 in Table 1 due to double-counting of mothers and fathers who both left self-employment.

employment status in the year following their last observed year in self-employment. In general, their labor market behavior does not suggest that they left self-employment because they were very successful. The vast majority are still in the labor market: two-thirds of the parents who left self-employment were employed a year later, and an additional 16% were registered as unemployed. These parents generally did not take jobs that one would see as an attractive alternative to successful self-employment: of those who were employed, two-thirds found employment in lower white collar or unskilled blue-collar jobs. Moreover, the new jobs do not appear to take advantage of any parental knowledge and contacts in the industry of self-employment, as only 3% of the parents find upper white-collar positions in the same industry.

Table 4 presents estimates of the effect of parental self-employment from models that include a measure of parental wealth (divided by 100,000 for presentational purposes). Parental wealth generally does not have an effect on the rate of entry into self-employment. None of the models in Table 4, including the various interaction effects of parental wealth with parental self-employment history, improve significantly over the final model in Table 2. Furthermore, the inclusion of the parental wealth measure does not

**Table 4.** Effects of Parental Wealth.

Variable	(1)	(2)	(3)	(4)	(5)
Parental wealth	0.001 (0.001)	0.001 (0.001)	0.000 (0.001)	0.003* (0.001)	0.003* (0.001)
Parental self-employment during child's					
Adolescence only	0.421** (0.089)	0.446** (0.091)	0.422** (0.089)	0.417** (0.090)	0.449** (0.091)
Adulthood only	0.516** (0.056)	0.516** (0.056)	0.502** (0.056)	0.515** (0.056)	0.507** (0.056)
Adolescence and adulthood	0.611** (0.038)	0.610** (0.038)	0.612** (0.038)	0.618** (0.038)	0.619** (0.038)
Adolescence × parental wealth		-0.012 (0.013)			-0.014 (0.013)
Adulthood × parental wealth			0.005* (0.002)		0.002 (0.002)
Adolescence and adulthood × parental wealth				-0.003 (0.002)	-0.003 (0.002)
Log-likelihood	-25,929	-25,928	-25,927	-25,928	-25,925

*Note:* All models include the full set of covariates included in Table 4.

Two-sided *t*-tests: \* $p < 0.05$ ; \*\* $p < 0.01$ .

change the substantive conclusions regarding the effects of parental self-employment; while the magnitude of the impact of parental self-employment is reduced, the difference is trivial. We can therefore reject both Hypotheses 2 and 3: differences between self-employed and non-self-employed parents in wealth, or in the effect of parental wealth, cannot account for the inter-generational inheritance of self-employment. On the other hand, Hypothesis 1 still cannot be rejected after controlling for parental wealth.

The inheritance of self-employment may still be due to closure processes if self-employed parents ease their children's entry by transferring their social capital. If parents are transferring social capital – in the form of entrepreneurially relevant knowledge and contacts – then this should manifest itself in the child's choice of industry to enter. Specifically, children of the self-employed should be more likely to enter their parent's industry than other children (Hypothesis 4). I model industry choice as a function of two factors. The first is an individual's own work experience prior to entry into self-employment. If children worked in their parent's industry prior to becoming self-employed, their choice of industry may reflect knowledge and contacts derived from their own work experience. The second factor is parental self-employment status, which allows me to determine whether children of the self-employed have an excess tendency to enter the same industry as their parents. Employed parents serve as a baseline group in this case; they can also transmit knowledge about their industry of employment to their children, and lead their children to enter self-employment in their industry. If children of self-employed parents have an excess tendency to choose their parent's industry when entering self-employment, then it would suggest that self-employed parents facilitate their children's entry into self-employment.

Table 5 presents the results of two different modeling strategies for assessing whether children of the self-employed have a greater tendency to enter their parent's industry. Panel A contains estimates from a competing risk model, where I estimated the determinants of entry into self-employment into either the same industry as a parent's current employment or self-employment, or into a different industry. (These models contain the full set of covariates in the models in Table 2.) Panel B presents the results of logistic regression models of the likelihood that a new venture is in the same industry as one of the founder's parents.<sup>25</sup> The results across the two estimation strategies parallel each other, with the differences in the coefficient estimates for a given variable across the competing risks largely reflected in the logistic regression estimates. However, it is important to note in Panel A that parental self-employment has a substantial positive effect on the likelihood of entering self-employment in a different industry

**Table 5.** Models of Industry Choice.

(A) Competing Risk Logistic Regression Estimates of Rate of First Entry into Self-Employment into Same or Different Industry as Parent(s)

	(1)		(2)	
	Same Industry	Different Industry	Same Industry	Different Industry
Parental wealth	-0.003* (0.001)	0.002** (0.001)	-0.003* (0.001)	0.002** (0.001)
Employed in same industry as parent	2.140** (0.133)	-0.327** (0.059)	2.327** (0.167)	-0.418** (0.064)
Parent currently self-employed	0.971** (0.078)	0.428** (0.044)	1.350** (0.211)	0.255** (0.064)
Employed in same industry × parent currently self-employed			-0.435 (0.224)	0.331** (0.086)
Log-likelihood	-6,126	-21,374	-6,124	-21,367

(B) Logistic Regression Models of Choosing the Same Industry as Parent(s), Conditional on Entry into Self-Employment

	(1)	(2)
Parental wealth	-0.012* (0.006)	-0.012* (0.006)
Employed in same industry as parent	1.798** (0.111)	2.035** (0.149)
Parent currently self-employed	0.538** (0.087)	1.045** (0.206)
Employed in same industry × parent currently self-employed		-0.615** (0.227)
Log-likelihood	-1,815	-1,812

*Note:* All models contain the full set of covariates included in Table 3. Estimated standard errors are adjusted for the covariance of the estimators across the equations for the competing risks (Weesie, 1999).  $N = 3,934$ .

Two-sided  $t$ -tests: \* $p < 0.05$ ; \*\* $p < 0.01$ .

than the parents. Parental social capital, as operationalized here, therefore cannot explain away the transmission of self-employment; children of the self-employed are much more likely to enter self-employment than other children, even if they do not take advantage of their parent's industry knowledge.

**Table 6.** Net Effects on Industry Choice.

		Parent Currently Self-Employed	
		Yes	No
Employed in same	Yes	2.465	2.035
Industry as parent	No	1.420	0

*Note:* Net effects from Model 2 in Table 5, Panel B.

Nonetheless, parental self-employment does influence industry choice. The combined effects of prior industry of employment and parental self-employment status are summarized in Table 6, using estimates from the second logistic regression model in Panel B of Table 5. The baseline categories in this model are people without prior work experience in the parent’s industry and with employed parents. Comparatively, these individuals should have the least access to entrepreneurial knowledge and contacts in the parent’s industry. Indeed, we see from Table 6 that they are the least likely to enter into self-employment in their parent’s industry (conditional on entering self-employment). In the top right-hand cell of Table 6 we see that prior work experience in an employed parent’s industry makes it much more likely that an individual will choose that industry (multiplier of  $\exp(2.035) = 7.65$ ). This suggests that people are more likely to pursue entrepreneurial opportunities in industries where they have experience. For individuals who cannot draw on their own work experience, having a self-employed parent almost triples the likelihood that they will choose to enter the parent’s industry, relative to the baseline category ( $\exp(1.045) = 2.84$ ). This suggests that self-employed parents more effectively transmit knowledge about entrepreneurial opportunities and contacts to their children outside the industry than employed parents, consistent with Hypothesis 4. Finally, a comparison of the net effects in the top row of Table 6 suggests that a self-employed parent adds something above and beyond their employed child’s own knowledge of the common industry. Individuals with prior work experience in their parent’s industry are 54% more likely to choose that industry if their parent is self-employed than if their parent is employed, further supporting the idea that the social capital of self-employed parents influences the industry choice of their children when they enter self-employment.<sup>26</sup>

It is interesting to note that access to parental financial capital appears to lower the likelihood of entering the same industry as the parent, as evidenced by the negative effect of parental wealth in Table 5. In combination

with the results from Table 4, it appears that while parental wealth does not influence the rate of entry into self-employment in general, it does influence the choice of industry; as parental wealth increases, children who choose to become self-employed appear to venture further afield. However, this effect is substantively quite modest: a movement from the median of parental wealth to the 75th percentile (among those who entered self-employment) reduces the likelihood of entering the same industry as the parent by 2.4%.<sup>27</sup> Tests for interaction effects indicate that this effect does not vary by an individual's prior work experience or the self-employment status of their parents.

I turn now to consider how parental self-employment affects performance. If it is the case that exposure to parental self-employment during adolescence leads to the acquisition of entrepreneurial skills, then we should expect children of the self-employed to perform better once they enter self-employment (Hypothesis 5). I test Hypothesis 5 by looking at two measures of performance: the rate of exit from self-employment, and self-employment income in the first year of self-employment. Both measures are imperfect in certain respects. In the case of exit from self-employment, there are two measurement limitations. First, I am only able to track exit timing for individuals who entered into self-employment as unincorporated or private employers. However, these cases account for 83% of entrants in this sample. Second, exit from private self-employment will not in all cases mean entrepreneurial failure. For example, a successful entrepreneur may incorporate his or her venture. In IDA, this would be recorded as an exit from private self-employment, even though the venture is ongoing.<sup>28</sup> However, an examination of these transitions (not shown) suggests that most represent entrepreneurial failure. For example, 69% of those leaving self-employment enter employment in a different industry; transitions due to a change in incorporation status should rarely involve changes in industry. Only 3.2% transitioned to employment in upper white-collar jobs in the same industry, which is the destination we would expect if founders of a new venture had changed incorporation status.

Exit rates from self-employment are an imperfect measure of performance since people may, if sufficiently committed to the idea of self-employment, persist in self-employment even when the venture is not successful. Moreover, such persistence in the face of poor performance seems particularly likely to occur if people value the autonomy of self-employment highly. I therefore also present estimates of the effects of parental self-employment on self-employment income in the first year of self-employment. I focus on the first year of self-employment income because measures of income in



subsequent years are subject to selection processes as people leave self-employment. Moreover, the first year of self-employment is when initial differences in self-employment skills should reveal themselves most clearly. I measure self-employment income as all non-salary income while self-employed. This is not a precise measure of income derived from self-employment, since it may include other sources of income (such as interests and dividends). I therefore control for family assets in the model. Furthermore, since IDA data are only collected once a year, we do not have precise data on the amount of time an individual has been self-employed and potentially generating self-employment income. Many of these individuals therefore have salary incomes in their first year of self-employment; I control for this as well. An alternative approach involves restricting the analysis to those with no salary income. Doing so does not change the substance of the results.

Table 7 contains estimates from piecewise-constant hazard rate models of exit from private employer status as a function of parental self-employment. Failure rates decline with an individual's own work experience in the industry prior to entering, and with the presence of parents working in the same industry. This latter effect suggests that people may benefit from their parent's knowledge of the industry. However, the estimates indicate that parental self-employment has no effect on the exit rate. For example, the second model contains a dummy variable for whether an individual's parents were self-employed during the child's adolescence. If one subscribes to the argument that children acquire entrepreneurial skills by observing and working for the parents in the family home, this variable should lower the exit rate. While the coefficient estimate is in the expected direction, it is not significant. Hypothesis 5 is not supported in the analysis of failure rates.

In light of the evidence in Table 6 that the children of the self-employed are much more likely to enter their parent's industry, one might expect that they benefit from having self-employed parents in the same industry. However, there is no evidence for this. The third model suggests that current parental self-employment in general has no significant impact on the failure rate. Furthermore, there is little support for the notion that it is particularly beneficial to have a self-employed parent in the same industry; the improvement in fit between the first and fourth models in Table 7 is not significant.

Table 8 presents OLS regression estimates of the determinants of self-employment income in the first year of self-employment. Again, there is no evidence that having self-employed parents helps performance. Individuals whose parents were self-employed during their adolescence do not

**Table 7.** Piecewise-Constant Hazard Rate Models of Exit from Status as Private Employer.

Variable	(1)	(2)	(3)	(4)
0–5 Years in self-employment	0.641 (0.502)	0.683 (0.504)	0.660 (0.504)	0.649 (0.504)
5+ Years in self-employment	–0.114 (0.514)	–0.072 (0.516)	–0.095 (0.517)	–0.102 (0.516)
Prior log income	–0.011 (0.016)	–0.011 (0.016)	–0.010 (0.016)	–0.011 (0.016)
Prior non-salary income	0.000 (0.002)	0.000 (0.002)	0.000 (0.002)	0.000 (0.002)
Wage experience	–0.041 (0.022)	–0.041 (0.022)	–0.041 (0.022)	–0.042 (0.022)
Age	–0.069** (0.025)	–0.070** (0.025)	–0.070** (0.025)	–0.068** (0.025)
Self-employment income	–0.331** (0.044)	–0.331** (0.044)	–0.331** (0.044)	–0.330** (0.044)
Log salary during self-employment	0.060** (0.007)	0.060** (0.007)	0.060** (0.007)	0.060** (0.007)
Log family assets	–0.036** (0.012)	–0.035** (0.012)	–0.036** (0.012)	–0.036** (0.012)
Log family debts	0.038** (0.011)	0.038** (0.011)	0.038** (0.011)	0.037** (0.011)
Experience in industry entered	–0.078** (0.016)	–0.079** (0.016)	–0.078** (0.016)	–0.081** (0.016)
Parent(s) are in same industry	–0.496** (0.096)	–0.488** (0.097)	–0.491** (0.097)	–0.613** (0.132)
Parent(s) self-employed during adolescence		–0.065 (0.069)		
Parent(s) currently self-employed			–0.027 (0.074)	–0.077 (0.082)
Parent(s) currently self-employed × Parent(s) are in same industry				0.283 (0.195)
Log-likelihood	–2.064	–2.064	–2.064	–2.063

Note:  $N = 7,168$  spells and 937 failures.

Two-sided  $t$ -tests: \*  $p < 0.05$ ; \*\*  $p < 0.01$ .

All models include controls for sex, marital status (interacted with sex),  $N$  of children, highest education achieved, and prior occupation. Results available upon request.

have significantly higher self-employment income. Current parental self-employment does not have an impact either.

In summary, there is no evidence in either Table 7 or 8 that is consistent with Hypothesis 5. However, these tests are limited by the available data,

**Table 8.** OLS Regression Estimates of Determinants of Self-Employment Income in First Year of Self-Employment.

Variable	(1)	(2)	(3)	(4)
Prior log income	1.707** (0.540)	1.713** (0.539)	1.692** (0.540)	1.671** (0.540)
Prior non-salary income	0.275** (0.045)	0.274** (0.045)	0.274** (0.045)	0.274** (0.045)
Wage experience	1.745* (0.729)	1.756* (0.728)	1.734* (0.729)	1.720* (0.728)
Age	-1.120 (0.826)	-1.073 (0.826)	-1.039 (0.829)	-0.986 (0.830)
Log salary during self-employment	-5.778** (0.264)	-5.757** (0.265)	-5.773** (0.264)	-5.766** (0.264)
Log family assets	1.459** (0.471)	1.392** (0.473)	1.410** (0.473)	1.398** (0.473)
Log family debts	-0.420 (0.364)	-0.397 (0.365)	-0.408 (0.365)	-0.414 (0.365)
Experience in industry entered	1.825** (0.497)	1.848** (0.497)	1.834** (0.497)	1.821** (0.497)
Parent(s) are in same industry	4.445 (3.059)	3.843 (3.081)	3.857 (3.102)	1.115 (3.688)
Parent(s) self-employed during adolescence		3.664 (2.324)		
Parent(s) currently self-employed			2.858 (2.526)	0.659 (2.990)
Parent(s) currently self-employed × Parent(s) are in same industry				7.453 (5.424)
R <sup>2</sup>	0.207	0.208	0.208	0.208
N	2,634	2,634	2,634	2,634

Note: Two-sided *t*-tests: \*  $p < 0.05$ ; \*\*  $p < 0.01$ .

All models include controls for sex, marital status (interacted with sex), *N* of children, highest education achieved, and prior occupation. Results available upon request.

since IDA does not contain information on the extent to which children of the self-employed actually worked in their parent’s business. For those who did not, the opportunity to acquire entrepreneurial skills was limited to observation of the self-employed parent. It is difficult to assess the extent to which children’s exposure was limited to observing their self-employed parents, but Aldrich et al. (1998) suggest that parental self-employment during adolescence is associated with work experience in the family firm, but that this experience has no measurable impact on subsequent entrepreneurial performance.

## DISCUSSION

The transmission of self-employment is a remarkable instance of the influence of family background on career choice. Consistent with evidence from the United States, the Danish data show that having a self-employed parent doubles the raw rate of entry into self-employment. Yet despite the remarkable magnitude of this effect, sociologists have a limited understanding of the social processes behind it. While scholars have advanced a variety of different theories to explain how self-employed parents shape their children's career choices so strongly, the reliance on static comparisons of the statuses of the two generations has made it impossible to adjudicate between the different accounts. By adopting a life-course perspective and using unique, longitudinal data, I have addressed this identification problem and shed greater light on how self-employed parents shape their children's career choices. While this study does not provide a definitive explanation for the transmission of self-employment, the analyses suggest that some suspects can be eliminated from consideration in the Danish case.

First, parental wealth does not explain the transmission of self-employment in Denmark. While continuously self-employed parents had higher average wealth levels than other parents, this difference could not account for the higher rates of self-employment among children of the self-employed. Parental wealth had no direct effect on the rate of self-employment, suggesting that, in the Danish context, differential access to "family credit markets" (Dunn & Holtz-Eakin, 2000, p. 283) does not constitute a barrier to entry to self-employment. This may reflect the relative absence among Danes of the expectation that parental wealth could or should be used in this way. For example, Hancock and Bager (2001) found that despite a generally high reliance on informal sources of funding among entrepreneurs (relative to formal funding channels), only 12% of surveyed nascent entrepreneurs in Denmark expected to turn to family and relatives to fund their new ventures, and only 11% of recently launched ventures had done so. Nonetheless, this evidence echoes studies in the United States (Dunn & Holtz-Eakin, 2000; Kim et al., 2006) and Canada (Aldrich et al., 1998) that found weak effects of parental wealth at best, and the results presented here reinforce the conclusion that parental wealth is not a necessary condition for the transmission of self-employment in highly developed, industrialized societies.

Second, while parental self-employment has dramatic effects on rates of entry into self-employment, it has no measurable impact on performance in self-employment. There is no support for the hypothesis that children of the self-employed have superior self-employment skills. Children of the self-employed

do not leave self-employment at a lower rate, and do not generate more self-employment income in their first year of self-employment. A fervent supporter of the skill acquisition argument might argue that this reflects measurement error, as IDA does not contain any direct measure of work experience in the family firm. However, while exposure to parental self-employment during adolescence is not sufficient to increase the returns to self-employment, it is sufficient to generate dramatically higher rates of entry into self-employment. It is difficult to conclude from this that the transmission of self-employment can be attributed to the acquisition of entrepreneurial skills.

The evidence for the social capital hypothesis is more mixed. On the one hand, the transmission of self-employment is not limited to children entering self-employment in the same industry as their parents; parental self-employment has a more general effect on the entry rate. Thus, the effect of parental self-employment is not fully mediated by the parent's social capital, as measured here. On the other, children of the self-employed are substantially more likely to choose their parent's industry for their entrepreneurial activity than children of the employed. This is consistent with the argument that self-employed parents are in a position to communicate valuable information to their children about entrepreneurial opportunities, and broker contacts with resource holders. However, it is worth noting that the power of the social capital of self-employed parents is limited, as there are no post-entry performance benefits associated with having self-employed parents in the same industry. Roberts (1991) similarly found that while family background exerted a strong influence on rates of entry into high-technology entrepreneurship, it had no effect on entrepreneurial performance. This makes it difficult to argue that the transmission of self-employment arises from parents passing on privilege to their children, or that the children of the non-self-employed are systematically disadvantaged in the entrepreneurial process. In light of this, one might suspect that the effects on industry choice do not represent transfers of social capital, but rather a more focused form of role modeling, where the children of the self-employed not only value or aspire to self-employment, but also aspire to be in the same industry as their parents. This alternative interpretation cannot be ruled out with the current data, and so the reasons for the effects of parental self-employment on industry choice should be explored in future research.

Finally, parental self-employment during adolescence alone is sufficient to generate substantially higher self-employment rates. Parents thus appear able to shape their children's propensity to enter into self-employment relatively early in life. The magnitude and robustness of this effect is striking, particularly in light of the restrictive and somewhat arbitrary cutoff for

defining exposure during adolescence. Furthermore, there is little evidence to suggest that this effect is driven by any positional advantages possessed by these parents, since they are less wealthy than those who remain in self-employment and predominantly remain in the labor force after leaving self-employment. One prominent interpretation of this effect points to parental role modeling and job values, which scholars long have argued can exert a strong influence on the decision to enter self-employment (Hout, 1984; Aldrich et al., 1998) and on children's work values and career choices more generally (Kohn, 1969; Miller & Swanson, 1958; Johnson, 2002). Yet with only a single measure of parental status early in life, the conclusions from earlier research have rested on the assumption that the influence of parents on their children's aspirations and work values during adolescence have a lasting impact on subsequent career choices, and that adult experiences have a minor impact by comparison (e.g., Halaby, 2003, p. 257). By identifying different parental trajectories of self-employment, I avoid making this assumption and strengthened the empirical foundations of the role modeling argument. Support for the role modeling argument could be further strengthened by addressing the alternative explanation rooted in the genetic transmission of job values and personality characteristics (Aldrich & Kim, 2007). Of course, genetic transmission and parental socialization may exert concomitant effects (Maccoby, 2000).

## CONCLUSION

High levels of occupational inheritance are commonly seen as evidence of barriers to mobility into privileged class positions, as the children of class incumbents benefit from exclusive access to valuable resources. There is little evidence in this study, however, to suggest that the high rates of intergenerational transmission of self-employment result from children of the self-employed having superior access to entrepreneurial resources.<sup>29</sup> For students of entrepreneurial processes, therefore, the most striking result in these analyses is the strong effect of exposure to parental self-employment during adolescence alone. Coupled with the fact that there is little evidence to support the notion that children of the self-employed have superior entrepreneurial skills, this result is consistent with the claim that role modeling in the family environment has a dramatic and lasting impact on the likelihood of subsequent entrepreneurial activity.

While policies directed at encouraging entrepreneurial activity may focus on factors that inhibit access to entrepreneurial opportunities, this suggests

that there is an important supply-side component to the dynamics of entrepreneurship. What remains unclear is the nature of the parental influence and its impact on the transmission of self-employment. At least three different mechanisms can be identified. First, as has been argued in the literature, higher self-employment rates among children of self-employed may reflect a greater preference for autonomy in their work lives relative to other children (Hout, 1984). Second, exposure to parental self-employment may raise children's subsequent risk tolerance, or their willingness to accept greater uncertainty about their income in exchange for higher expected returns (Halaby, 2003; Miller & Swanson, 1958; although see Xu & Ruef, 2004). Both of these mechanisms suggest that parental influence operates through its effects on how children value different career options. Finally, parents' location in social structure may shape children's cognitive maps of the landscape of occupational opportunities (Carroll & Mosakowski, 1987). From this perspective, children never exposed to self-employment may not actually attach less value to self-employment, but simply be less likely to consider it as an option. Adjudicating between these accounts is an important issue for future research.

## NOTES

1. I view self-employment as a form of entrepreneurial activity, although some might argue that entrepreneurship should be more narrowly defined, in particular to exclude individuals who are independent contractors, etc. Operationally, my focus is on predicting who is likely to found a new venture (with employees).

2. It is difficult to envision many common scenarios in which self-employed parents pass material and social advantages along to their children without having some social contact with them. The personality-related genetic mechanisms discussed by Aldrich and Kim (2007) do not require social contact between parents and child. They are not studied directly here. As Maccoby (2000) argues, the fact that there is a genetic component to psychological factors does not diminish the importance of the home environment in shaping children's life-long behavior.

3. See Aldrich and Kim (2007) on the advantages of adopting a life-course perspective in the study of entrepreneurship.

4. The effect of parental self-employment on the acquisition of entrepreneurial skills can also be seen as a form of exposure process instead of closure; the difference depends on whether self-employed parents privilege their children in making opportunities available. My main focus here is on the extent of empirical support for the proposed mechanism, and not whether it is best considered a closure or exposure mechanism.

5. One could, in a similar vein, also argue that the effect of role modeling by the parent consists in limiting the range of career choices that seem viable to the child; children of professors may have little sense of the alternative. The implications for the transmission of self-employment are the same.

6. For further evidence on work experience in the family venture, see Aldrich and Kim (2007).

7. It is not entirely clear from Hout's discussion why job autonomy per se should lead to superior access to capital, except that autonomy is a consequence of ownership.

8. The classification in Fig. 1 could be elaborated with more fine-grained information on the timing of parental self-employment, but for analytic purposes here the key issue is the timing of self-employment relative to the completion of compulsory schooling.

9. Parents who are not self-employed when the children are at risk of entering self-employment may have other resources that they can use to facilitate their children's entry, but these resources are by definition not rooted in positional advantages associated with being self-employed.

10. It is unlikely that the parents' resources dissipate immediately after leaving self-employment, particularly if they have left self-employment because they have been very successful. I examine the destinations of those parents who left self-employment before their child's adulthood below.

11. Dunn and Holtz-Eakin (2000) do estimate an interaction effect between parental self-employment and parents' business assets, finding that it has a significant, positive effect on the transition probability. However, they interpret this as a measure of the parents' success in self-employment. It is also somewhat difficult to interpret, since parents who are not self-employed by definition have no business assets. They provide no test of whether the effect of non-business assets is stronger for self-employed parents.

12. This corresponds to approximately \$5,000 in current (2004) values.

13. As noted above, not all members of the original birth cohorts who were alive in 1980 are observed, since some of them left the population prior to 1994.

14. The mean number of annual observations is 16.8, so very few life histories are truncated or incomplete.

15. A further complication is that the difference between being an employer and being a VAT payer is simply a matter of whether one has any employees at the time when the data are "collected." Individuals can transition back and forth between these two states. In particular, someone who starts as a single-person business (VAT payer) may add employees after having some success. To be conservative, I did not treat transitions from VAT-payer status to employer status as transitions to self-employment. In short, for private firms, only transitions to employer status from a non-self-employed state were treated as a change in state.

16. This indirect approach will unavoidably contain some measurement error. Included among the problematic cases are: some top managers in the first year may have been hired by the true founders; the owners of a venture may simply provide capital and hire employees; and the founders' occupational codes may not indicate that they are in top management.

17. There are 1,847 transitions into non-agricultural self-employment by individuals whose parents were self-employed during the child's adulthood; of these, 146 transitions occurred simultaneously with the parent's exit from self-employment in the same industry.

18. As a share of all person-year spells in which an individual was in some form of self-employment, self-employment with unemployment insurance accounts for 3.6% of spells, "time-limited" self-employment 5.1% and assisting spouses 2.2%,



with miscellaneous self-employment 0.2%. Employer status accounts for 24.5% of the self-employment spells, while VAT-payer status accounts for 60.5% of spells.

19. The two primary paths of further education are vocational and academic upper secondary schools, although a substantial proportion of children complete an additional year (10th grade) in lower secondary school before pursuing further education or ending their formal education.

20. Almost all forms of schooling through graduate education are paid for by the state, including generous stipends for students not living with their parents.

21. I also estimated piecewise-constant (continuous-time) hazard rate models and Weibull models. Not surprisingly, the estimates for the piecewise-constant models were practically indistinguishable from the logistic regression results. However, estimation times were considerably longer. The piecewise specification (using logistic regression) is preferable to the Weibull model, since the pattern of duration dependence appears to be non-monotonic.

22. “Never” and “ever,” when referring to parental self-employment, refer to their employment histories after 1980, when the IDA data begin. It is therefore possible that children whose parents are categorized here as never self-employed were in fact self-employed prior to the child’s adolescence.

23. Separate analyses suggest that children of the self-employed do not differ dramatically from other children on two other easily observable characteristics: age of first wage experience and highest level of completed schooling. Children of the self-employed are more likely to pursue vocational educations, however. I control for educational track below.

24. For individuals with zero assets or debts, I add 1 kroner before taking the log.

25. In cases where the pattern of duration dependence is the same for both competing risks, a competing risks model can be estimated as two models: a hazard rate model of the likelihood of changing state, and a logit (or probit) model of which state was chosen, conditional on an event occurring (Petersen, 1995).

26. The results also indicate that an individual’s own work experience is a partial substitute for the social capital of a self-employed parent. This can be seen in the fact that an individual’s own work experience in the parent’s industry has a smaller effect if the parent is self-employed than if the parent is employed.

27. Parental wealth is measured in 100,000 Danish kroner units:  $\exp(2.956 * -0.012) / \exp(0.906 * -0.012) = 0.976$ .

28. The problem here is that Statistics Denmark does not release, for confidentiality reasons, workplace identification numbers for private employers.

29. The focused nature of this study makes it impossible to rule out the possibility that closure processes account for some forms of class reproduction. The extent to which this is the case can only be determined from future research. But we *can* conclude from this study that closure processes cannot provide a *general* explanation for class reproduction. Indeed, it is possible that the mechanisms underlying occupational inheritance may vary across occupations, depending on the strength and nature of occupational structuration (Grusky & Sørensen, 1998). It is also possible that in advanced capitalist societies, many of the theoretical mechanisms identified by sociologists as structural sources of inequality – such as credentials – may be more germane to career dynamics and labor market outcomes (Sørensen, 2000; Weeden, 2002).

## ACKNOWLEDGMENTS

Astrid Würtz provided invaluable research assistance. I am grateful to Søren Leth Sørensen and Jørn Hansen Schmidt of Statistics Denmark for their invaluable assistance in assembling the data extract from IDA. I am also indebted to Niels Westergaard-Nielsen, Tor Eriksson, Paul Bingley, and Henrik Mathiesen of the Center for Corporate Performance at the Aarhus School of Business for their guidance and assistance. Howard Aldrich, Diane Burton, Glenn Carroll, Roberto Fernandez, David Grusky, David Harris, Yasheng Huang, Peter Marsden, Joel Podolny, Ezra Zuckerman, the editors of this volume, and seminar participants at Cornell University, MIT and Oxford University provided helpful comments. Generous support from the Sloan School helped fund the many trips across the Atlantic. All errors remain my own.

## REFERENCES

- Aldrich, H. E., & Kim, P. H. (2007). A life course perspective on occupational inheritance: Self-employed parents and their children. In: M. Ruef & M. Lounsbury (Eds), *The sociology of entrepreneurship* (Vol. 25, pp. 31–82). Amsterdam: Elsevier.
- Aldrich, H. E., Renzulli, L. A., & Langton, N. (1998). Passing on privilege: Resources provided by self-employed parents to their self-employed children. *Research in Social Stratification and Mobility*, 16, 291–317.
- Aldrich, H. E., & Zimmer, C. (1986). Entrepreneurship through social networks. In: S. Donald & S. Raymond (Eds), *The art and science of entrepreneurship* (pp. 3–23). New York: Ballinger.
- Benz, M., & Frey, B. S. (2003). *The value of autonomy: Evidence from the self-employed in 23 countries*. Institute for Empirical Research in Economics Working Paper no. 173.
- Benz, M. (2005). *Entrepreneurship as a non-profit seeking activity*. Institute for Empirical Research in Economics Working Paper No. 243, University of Zurich.
- Blau, P. M., & Duncan, O. D. (1967). *The American occupational structure*. New York: Free Press.
- Carroll, G. R., & Mosakowski, E. (1987). The career dynamics of self-employment. *Administrative Science Quarterly*, 32, 570–589.
- Chinoy, E. (1955). *Automobile workers and the American dream*. Garden City, NY: Doubleday.
- Dunn, T., & Holtz-Eakin, D. (2000). Financial capital, human capital, and the transition to self-employment: Evidence from intergenerational links. *Journal of Labor Economics*, 18, 282–305.
- Erikson, R., & Goldthorpe, J. H. (1992). *The constant flux: A study of class mobility in industrial societies*. New York: Oxford University Press.
- Evans, D. S., & Leighton, L. (1989). Some empirical aspects of entrepreneurship. *American Economic Review*, 79, 519–535.

- Featherman, D. L., & Hauser, R. M. (1978). *Opportunity and change*. New York: Academic Press.
- Grusky, D. B., & Sørensen, J. B. (1998). Can class analysis be salvaged? *American Journal of Sociology*, *103*, 1187–1234.
- Halaby, C. N. (2003). Where job values come from: Family and schooling background, cognitive ability, and gender. *American Sociological Review*, *68*, 251–278.
- Hamilton, B. H. (2000). Does entrepreneurship pay? An empirical analysis of returns to self-employment. *Journal of Political Economy*, *108*, 604–632.
- Hancock, M., & Bager, T. (2001). *Global entrepreneurship monitor: Danish national report – 2001*. Odense, Denmark: Odense University Press.
- Hannan, M. T., & Freeman, J. (1989). *Organizational Ecology*. Cambridge, MA: Harvard University Press.
- Hauser, R. M., Tsai, S. L., & Sewell, W. H. (1983). A model of stratification with response error in social and psychological variables. *Sociology of Education*, *56*, 20–46.
- Holtz-Eakin, D., Joulfaian, D., & Rosen, H. S. (1994). Entrepreneurial decisions and liquidity constraints. *Rand Journal of Economics*, *23*, 340–347.
- Hout, M. (1984). Status, autonomy and training in occupational mobility. *American Journal of Sociology*, *89*, 1379–1409.
- Hout, M., & Rosen, H. (2000). Self-employment, family background, and race. *Journal of Human Resources*, *35*, 670–692.
- Hurst, E., & Lusardi, A. (2004). Liquidity constraints, household wealth, and entrepreneurship. *Journal of Political Economy*, *112*, 319–347.
- Johnson, M. K. (2002). Social origins, adolescent experiences, and work value trajectories during the transition to adulthood. *Social Forces*, *80*, 1307–1340.
- Kim, P. H., Aldrich, H. E., & Keister, L. A. (2006). Access (not) denied: The impact of financial, human and cultural capital on entrepreneurial entry in the United States. *Small Business Economics*, *27*, 5–22.
- Kohn, M. (1969). *Class and conformity: A study in values*. Homewood, IL: Dorsey Press.
- Kohn, M., Slomczynski, K. M., & Schoenbach, C. (1986). Social stratification and the transmission of values in the family: A cross-national assessment. *Sociological Forum*, *1*, 73–102.
- Lentz, B. S., & Laband, D. N. (1990). Entrepreneurial success and occupational inheritance among proprietors. *Canadian Journal of Economics*, *23*, 563–579.
- Maccoby, E. E. (2000). Parenting and its effects on children: On reading and misreading behavior genetics. *Annual Review of Psychology*, *51*, 1–27.
- Miller, D. R., & Swanson, G. E. (1958). *The changing American parent: A study in the Detroit area*. New York: Wiley.
- Moskovitz, T. J., & Vissing-Jørgensen, A. (2002). The returns to entrepreneurial investment: A private equity premium puzzle? *American Economic Review*, *92*, 745–778.
- Parking, F. (1979). *Marxism and class theory: A Bourgeois critique*. New York: Columbia University Press.
- Petersen, T. (1995). Analysis of event histories. In: G. Arminger, C. C. Clogg & M. E. Sobel (Eds), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 453–517). New York: Plenum Press.
- Roberts, E. B. (1991). *Entrepreneurs in high technology*. New York: Oxford University Press.
- Robinson, R. V. (1984). Reproducing class relations in industrial capitalism. *American Sociological Review*, *49*, 182–196.

- Romanelli, E., & Schoonhoven, C. B. (Eds) (2001). The local origins of new firms. In: *The entrepreneurship dynamic: Origins of entrepreneurship and the evolution of new industries* (pp. 40–67). Stanford, CA: Stanford University Press.
- Sewell, W. H., & Hauser, R. M. (1975). *Education, occupation and earnings: Achievement in the early career*. New York: Academic Press.
- Sørensen, A. B. (2000). Toward a sounder basis for class analysis. *American Journal of Sociology*, 105, 1523–1558.
- Spenner, K. I. (1988). Social stratification, work and personality. *Annual Review of Sociology*, 14, 69–97.
- Stinchcombe, A. L. (1965). Social structure and organizations. In: J. G. March (Ed.), *Handbook of organizations* (pp. 142–193). Chicago: Rand McNally.
- Thornton, P. (1999). The sociology of entrepreneurship. *Annual Review of Sociology*, 25, 19–46.
- Vollebergh, W. A. M., Idema, J., & Raaijmakers, Q. A. W. (2001). Intergenerational transmission and the formation of cultural orientations in adolescence and young adulthood. *Journal of Marriage and the Family*, 63, 1185–1198.
- Weeden, K. A. (2002). Why do some occupations pay more than others? Social closure and earnings inequality in the United States. *American Journal of Sociology*, 108, 55–101.
- Weesie, J. (1999). Seemingly unrelated estimation and the cluster-adjusted estimator. *Stata Technical Bulletin*, November, 34–47.
- Western, M., & Wright, E. O. (1994). The permeability of class boundaries to intergenerational mobility among men in the United States, Canada, Norway and Sweden. *American Sociological Review*, 59, 606–629.
- Wong, R. S. (1992). Vertical and nonvertical effects in class mobility: Cross-national variations. *American Sociological Review*, 57, 396–410.
- Xu, H., & Ruef, M. (2004). The myth of the risk tolerant entrepreneur. *Strategic Organization*, 2, 331–355.

# BOUNDARY FORMATION IN EMERGENT ORGANIZATIONS

Hongwei Xu and Martin Ruef

## ABSTRACT

*An extensive literature in organizational theory discusses how established organizations shape and maintain their boundaries but offers little guidance as to how organizational boundaries emerge in the first place. This paper examines boundary formation in business startups using a nationally representative dataset of U.S. nascent entrepreneurs. We propose several distinct roles for individuals entering into entrepreneurial activity, distinguishing between “insiders” (owner-managers) who commit both time and financial resources to these startups and “outsiders” (including passive investors and advisors) who offer more limited resource commitments. Two important criteria demarcating organizational insiders and outsiders in emergent organizations are functionality and perceived trustworthiness. Our results suggest that boundary formation is more often based on a potential member’s trustworthiness, as perceived by peers, than functionality, emphasizing considerations such as transaction cost minimization and uniqueness of resource contributions. We propose several mechanisms that may account for this result among nascent entrepreneurs, including a lack of economic sophistication, calculative trust, and the importance of social solidarity for founder recruitment.*

---

The Sociology of Entrepreneurship

Research in the Sociology of Organizations, Volume 25, 125–153

Copyright © 2007 by Elsevier Ltd.

All rights of reproduction in any form reserved

ISSN: 0733-558X/doi:10.1016/S0733-558X(06)25004-3

## INTRODUCTION

Boundaries are an essential element of most definitions of organizations (Thompson, 1967; Williamson, 1975; Pfeffer, 1997; Aldrich & Ruef, 2006). As a condition of existence, organizations maintain distinguishable boundaries that separate them from their environments, though these boundaries may be incomplete and permeable (Scott, 2003; Meyer & Lu, 2005). The processes contributing to boundary formation have been examined from several major perspectives (Santos & Eisenhardt, 2005).<sup>1</sup> *Transaction cost economics (TCE)*, emphasizing the efficiency of contractual governance, suggests that boundaries should be set at the point that minimizes the cost of governing activities for a formal organization (Coase, 1937; Williamson, 1975). That is, to set the boundary, the marginal costs of internalizing a transaction should be compared with the marginal costs of transacting with an exchange partner through the market. Scholars employing a *resource dependence perspective* (Pfeffer & Salancik, 1978) focus on power. They propose that organizations should set their boundaries to reduce their dependency on other exchange partners in the environment. A third perspective, the *resource-based view (RBV) of the firm*, emphasizes organizational competency, suggesting that boundaries should be set at the point that extracts maximum value from the organization's resource configuration (Wernerfelt, 1984; Barney, 1991). Scholars employing an *identity* approach offer yet a fourth perspective on organizational boundaries, asserting that the rationale for boundary decisions is coherence with organizational identity and image (Walsh, 1995; Kogut & Zander, 1996; Rindova & Fombrun, 2001). They consider the mental maps maintained by organizational members, in particular the shared beliefs about "who we are."

While the organizational theories described above have been fundamental in shaping our thinking about organizational boundaries, empirical work from those perspectives focuses mostly on large, established firms rather than analyzing how organizational boundaries emerge in the first place. For instance, a substantial body of empirical TCE research addresses vertical integration decisions affecting extant hierarchies (Shelanski & Klein, 1995). As David and Han (2004, p. 54) note, analyses concerning the key postulates of TCE tend to be "tests of the largest, surviving firms" and therefore exhibit a bias toward mature organizations. Recent longitudinal examinations of boundary decisions (e.g. Bigelow, 2004) avoid severe survivor bias, but still focus on populations of operational organizations rather than preoperational startups. Other perspectives on organizational boundaries simply avoid entrepreneurial organizations entirely, as intimated by the

silence on this topic in a recent review of progress in resource dependence theory (Pfeffer, 2003).

Nevertheless, it is noteworthy that the gap in the literature may not merely be empirical. Existing perspectives on boundaries – developed in the context of stable, well-established organizations – may be poorly equipped *theoretically* to explain similar phenomena when applied to entrepreneurial ventures (Santos, 2003). Considering the resource dependence perspective, for instance, it is hard for entrepreneurs to make boundary decisions based on power dependence if important players have yet to be identified in emerging markets (Aldrich & Baker, 2001). Proponents of an identity-based approach may wonder how entrepreneurial organizations can leverage identity – shared beliefs about “who we are” – to inform boundary decision, if they have yet to establish a collective image (Aldrich & Fiol, 1994; Sarasvathy, 2001). Similar issues bedevil the transaction cost and resource-based approaches to boundary definition, insofar as entrepreneurs lack a clear sense of the “assets” that will be required for a new venture. At an early stage, even those entrepreneurs that merely reproduce existing organizational templates may not be clear about the key players in a particular market, the resources that need to be deployed, or the organizational identity they seek to project.

This article examines how organizational boundaries first emerge in entrepreneurial contexts. From an evolutionary perspective, the analysis of boundary formation in emergent organizations is critical because an organization achieves standing as a population member only after it becomes a bounded entity (Aldrich & Ruef, 2006). As a bounded entity, it starts to compete and cooperate with other organizations and contributes fully to population dynamics. The boundary formation processes of nascent organizations may have implications for their long-term performance, positioning, and survival.

Emerging organizations have distinctive features that influence our conceptualization of their boundary formation processes. Unlike well-established formal hierarchies, startups are more like peer groups or other types of decentralized collaborative groups. Hence, we argue, the key to drawing their boundaries is to decide what kinds of people should be brought inside the organization, as founders or employees, and what individuals will remain outside the organization, as vendors, consultants, and the like. Organizational founding teams provide one valuable, but understudied, context for exploring boundary formation (Ruef, Aldrich, & Carter, 2003).

Drawing from the economic literature on TCE and RBV of the firm, as well as the sociological literature of network dependence and identity, we propose that two important criteria demarcating organizational insiders and outsiders are *functionality* and *trustworthiness*. We evaluate *functionality* by

examining the level of asset-specificity and uniqueness of the resources that potential organizational members possess and *trustworthiness* by studying the strength of network ties and the level of similarity in sociodemographic identity among potential members. Special attention is thus given to reformulating the implications of TCE, the RBV of the firm, resource dependence theory, and organizational identity approaches in a context where formal organizational structure has yet to emerge.

## BOUNDARY EMERGENCE AND THE STARTUP PROCESS

Initially, numerous individuals may be involved in the startup process of an organization. But at some point, especially when it is time to legalize ownership, they split themselves into organizational insiders (e.g. owners) and outsiders (e.g. helpers on contract). Specifically, we argue, startup participants tend to be sorted into four organizational roles, distinguished by two dimensions: (a) the extent to which the participant will be a regular contributor to the organization; and (b) whether the participant has a substantial ownership stake (equity) (see Fig. 1). These organizational roles need not be static. For instance, an individual initially hired as a consultant may later become an employee or investor. Nevertheless, for the sake of simplicity, we focus on the boundary formation process at a given point in time, which is the initial allocation of legal ownership and equity stakes.

Among the resulting organizational roles, those participants who become external consultants and vendors represent the clearest case of organizational

		<b>Regular Contributor to Organization</b>	
		<i>No</i>	<i>Yes</i>
<b>Significant Equity Ownership</b>	<i>No</i>	Consultant / Vendor	Employee
	<i>Yes</i>	Investor	Owner- Manager

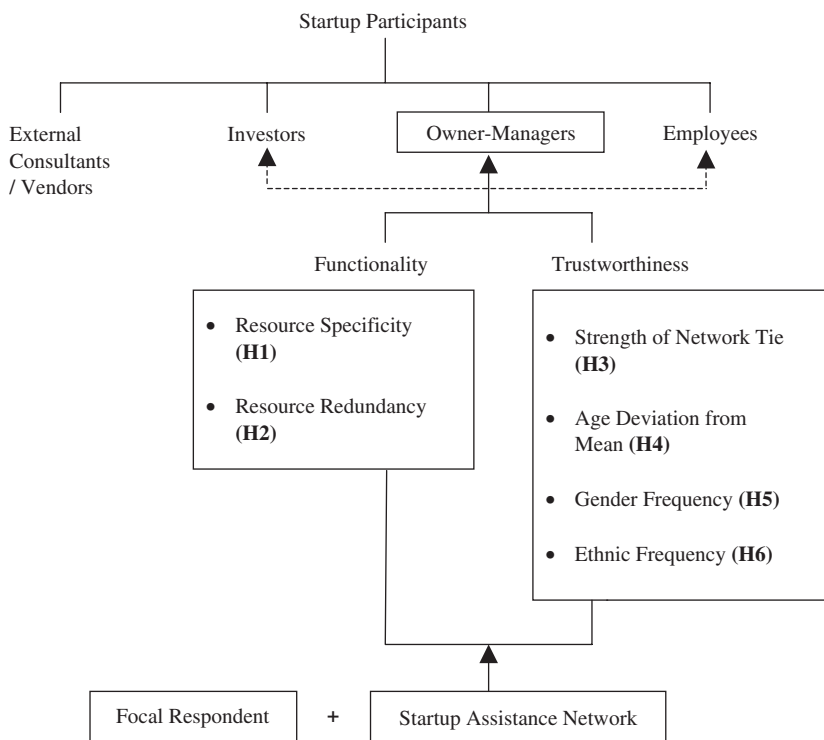
*Fig. 1.* Typology of Startup Participants.



outsiders. They engage in occasional transactions with other startup participants, typically under market-based governance, and lack any ownership stake. The roles of employees and passive investors represent intermediate cases on the continuum of insiders and outsiders. Employees are defined as regular contributors to the organization who do not share significantly in equity ownership, while investors, conversely, share significant equity but are not regular contributors. The joint occurrence of significant equity ownership and regular organizational participation defines owner-managers, the quintessential insiders in the startup context. Owner-managers are the only participants who share the ultimate risk for the emergent organization *and* interact with other startup participants in daily decision-making.

The roles identified in the cells of Fig. 1 are clearly ideal-types. The majority of business startups in the United States do not hire employees (Aldrich & Ruef, 2006), although they may still rely on regular contributions from family and friends who are not compensated through formal wage or equity arrangements. For purposes of the typology, these participants could be placed in the upper right-hand cell of the figure. In other cases, some of the roles identified in the typology may simply be absent. For instance, startups may be unable to secure investments from external parties; or they may avoid such involvement entirely, given the fickle commitment that investors often display toward the non-pecuniary goals of new ventures (Ruef, 2002). Given this variation in the role structure of organizational startups, our analytical interest centers on the boundary between owner-managers and all other “helpers” in the startup process, considered as a whole.

We conceptualize the sorting process that leads to boundary formation in emergent organizations as shown in Fig. 2. Our data are taken from the Entrepreneurial Research Consortium’s Panel Study of Entrepreneurial Dynamics (PSED) (Reynolds, 2000), a unique, nationally representative sample of nascent entrepreneurs. Survey screening identifies a focal respondent as well as their startup assistance network. Two clusters of variables help define the ongoing role relationship of individuals in the startup assistance network to the emergent organization. One cluster emphasizes the functional performance of the startup and draws primarily from the TCE and RBVs of the firm, as we discuss below. A second cluster of variables emphasizes the implicit trustworthiness attributed to different participants, drawing to a larger extent from the literature on social networks and homophily. We investigate the extent to which considerations of functionality and trustworthiness impact the selection of organizational insiders, applying a narrow definition that limits “insiders” to owner-managers in the startups,



*Fig. 2.* The Selection of Organizational Insiders.

while also conducting analyses that address the sensitivity of our findings to a broader definition that considers investors and employees as well.

## TRANSACTION COSTS AND FUNCTIONALITY

The TCE approach (Coase, 1937; Williamson, 1975) is one of the most widely used explanations for boundary decisions. Owing to the bounded rationality and possible opportunism of economic actors, transactions are costly to define, monitor, and enforce, leading to incomplete contracts. “Bounded rationality” refers to the inability of economic actors to properly anticipate the complex contingencies that might be relevant to long-term contracts. “Opportunism” is the rational pursuit by economic actors of their own advantage, with all means at their command, including guile and deceit.

In these circumstances, economic functions are performed more efficiently within the boundaries of hierarchical firms rather than by market processes that cross boundaries. When transactions are internalized into a hierarchy, the actors need not anticipate all possible contingencies; they can be handled within the firm's governance structure instead of leading to complex negotiations. Opportunism is also constrained by the authority relations and bureaucratic policies of the hierarchy; actors identified as opportunistic can be punished monetarily, withheld from promotion or fired.

Asset-specificity is one of the key transaction attributes that TCE scholars examine when predicting what transactions should be coordinated within an organization as opposed to in the marketplace. Recently, [David and Han's \(2004\)](#) systematic assessment of empirical TCE research showed that asset-specificity has been the most frequently considered independent variable, and it has fared the best in predicting the make-versus-buy decision.<sup>2</sup> Asset-specificity is related to the location of activities, dedicated physical assets, or dedicated knowledge and human skills ([Williamson, 1981](#)). The presence of asset-specificity in repeated transactions creates small number bargaining that increases the potential for hold-up by opportunistic actors ([Klein, 1988](#); [Williamson, 1991](#)). That is, asset-specificity increases bilateral dependence. Generally, the TCE literature suggests that it is more efficient to coordinate transactions of high asset-specificity in the context of a hierarchy than in the market.

One of the most common measures of asset-specificity is specialized human skills.<sup>3</sup> Therefore, applying the logic of TCE to boundary formation in emergent organizations, potential organizational members whose skills or ideas are more specific to a particular business should be more likely to become insiders.

**Hypothesis 1.** Individuals with skills or ideas specific to a startup effort will be more likely to become organizational insiders than those whose skills or ideas are not organization-specific.

## RESOURCE CONFIGURATION AND FUNCTIONALITY

An alternative perspective on functional boundary decisions in startups is offered by the RBV of the firm. RBV conceptualizes organizations as bundles of physical, human, and organizational resources that are deployed in economic activities ([Wernerfelt, 1984](#); [Barney, 1991](#); [Teece, Pisano, & Shuen, 1997](#)). Resources are often deployed in specific configurations to

increase organizational capabilities or competencies (Pralahad & Hamel, 1990; Admit & Schoemaker, 1993). Empirical research focuses on how established firms strategically shape organizational boundaries by leveraging existing resource configurations or exploring new resource configurations to adapt to changing market opportunities. RBV suggests that combinations of valuable, rare, inimitable, and non-substitutable resources can lead to sustainable competitive advantage (Barney, 1991).

Individuals bring different types of resources into an emerging organization. For instance, some offer physical capital, financial capital, or labor services; others provide creative business ideas or business training. While an emergent organization may need all those resources, participants who contribute unique and non-substitutable resources should be seen as more valuable to the startup than those whose resources overlap with the contributions of others in the startup assistance network. Based on the logic of RBV, the former individuals are more likely to be brought inside the organization as owner-managers than the latter.

**Hypothesis 2.** The higher the level of uniqueness in an individual's resource contribution to a startup effort, the more likely he or she is to become an organizational insider.

## SOCIAL NETWORKS AND TRUST

A general thesis in social network analysis is that high network density and intensity facilitate the flow of goods and communication among actors, exerting informal pressures toward normative consensus. A dense social network is one in which all counterparts have strong ties to one another. Dense social networks generate trust through norms of reciprocity that are reinforced with social sanctions against untrustworthy behavior (Granovetter, 1985). For example, opportunistic behavior by a member of a dense social network may be sanctioned by other network members through the temporary or permanent exclusion of the violating member from the group (Coleman, 1990).

The emergence of an organization is a process in which group members pool resources and then receive the benefits that are generated by pooling. Ekeh (1974) identified this process as group generalized exchange. The exchange structure involves a typical social dilemma situation. If we assume that each participant receives a specified share of the total benefits generated by resource pooling (a priori), it is rational for each person not to offer his/her full contribution in labor time or other resource commitments. Owing to

the social dilemma, participants should be selected as organizational insiders when others trust that they will not free ride. For an ego A and alter B, the stronger the tie between A and B, the more third parties A and B would be likely to share. The more shared third parties between A and B, the less likely B would free ride due to the possible punishment from those shared third parties. Based on this reasoning, the stronger the ties between an ego and an alter, the more likely the ego would like to select the alter as an insider.<sup>4</sup>

Dense networks also transmit normative expectations. In fact, Podolny and Baron (1997, p. 676) argue that within an organization, “a dense, redundant network of ties is often a prediction for: (1) internalizing clear and consistent set of expectations and values in order to be effective in one’s role; and (2) developing the trust and support from others that is necessary to access certain crucial resources (political aid, sensitive information, etc.) and to implement strategic initiatives.” The startup process involves a considerable amount of collective decision-making. To avoid friction, entrepreneurs will tend to affiliate most closely with others whom they expect to have similar views. The stronger the tie between entrepreneur A and alter B, the more likely A and B would trust each other to have consistent views in collective decision-making. Hence, we hypothesize that:

**Hypothesis 3.** The stronger the interpersonal tie between a focal entrepreneur and an alter involved in a startup effort, the more likely that alter will become an organizational insider.

## HOMOPHILY AND TRUST

The principle of homophily explains group composition in terms of the similarity of members’ characteristics. In their pathbreaking work, Lazarsfeld and Merton (1954) stated that homophily can be based on either externally salient social identities (e.g. ascribed characteristics such as gender, race, or age) or internal psychological states (values, beliefs, or norms). It is argued that, in either case, individuals of similar characteristics tend to have a greater level of interpersonal attraction, trust, and understanding than dissimilar individuals.

In an emergent organization, the founding team often needs to have a high level of interpersonal trust to survive the “liability of newness” – the high risk of dissolution or bankruptcy for young ventures (Stinchcombe, 1965; Carroll, 1983; Ruef, 2002). Given the liability of newness and the sizable investments of time and resources, the tendency toward homophily should be

especially noticeable in the boundary formation process. In support of this logic, Ruef et al. (2003) found that entrepreneurs tend to avoid dissimilar individuals in creating organizational founding teams.

Three of the most widely studied ascriptive characteristics driving homophily are age, gender, and ethnicity. Distributions of age, gender, and ethnicity are consequential for understanding conflict and turnover in organizations (Pfeffer, 1983). Gender homophily has been documented in different types of organizations, for instance, work establishments (Kalleberg, Knoke, Marsden, & Spaeth, 1996), voluntary organizations (McPherson & Smith-Lovin, 1982, 1987), and managerial networks (Ibarra, 1997). Carter (1994) examined gender homogeneity in business discussion networks. Ethnicity-based homophily has been found in workplaces (Reskin, 1999; Kalleberg et al., 1996) and classrooms (Schofeld, 1995). Ethnic homogeneity has also been identified in entrepreneurial activity, in particular among minority and immigrant groups reacting to prejudice in traditional employment contexts (Aldrich & Waldinger, 1990; Portes & Sensenbrenner, 1993).

Based on the logic of the homophily literature, individuals who match the dominant age, gender, or ethnicity in a startup group will enjoy a higher level of interpersonal trust than those who do not. With respect to the boundary formation process, we hypothesize that, for individuals in the startup assistance network:

**Hypothesis 4.** The closer an individual's age to the center of the group age distribution, the more likely that the individual will become an organizational insider.

**Hypothesis 5.** An individual who matches the dominant gender of the group involved in a startup effort is more likely to become an organizational insider than an individual who does not.

**Hypothesis 6.** An individual who matches the dominant ethnicity of the group involved in a startup effort is more likely to become an organizational insider than an individual who does not.

## DATA, MEASURES, AND METHOD

### *Data*

Our analysis of the boundary formation process in emergent organization draws on the PSED, a representative sample of individuals in the U.S. who

were in the process of starting a business between July 1998 and January 2000. Two items were used to determine whether an adult respondent qualified as a nascent entrepreneur: (1) “Are you, alone or with others, now trying to start a business?” and (2) “Are you, alone or with others, now starting a new business or venture for your employer?” If the respondent answered yes to either of the questions, two additional questions were used to qualify whether they were actively involved with the startup process, and whether he/she would share ownership in the business. Gartner, Shaver, Carter, and Reynolds (2004) offer complete details about the sampling procedure and descriptive statistics for various modules in the PSED.

The final sample of respondents in the PSED totals 830 nascent entrepreneurs, representing 830 emergent organizations. We removed 19 organizations for which respondents did not provide information about business ownership. Another seven respondents indicated that “non-persons” expected to own more than 50% of the venture. We also disqualified these startups from analysis because they could be influenced more by institutional investors than individual entrepreneurs. One respondent indicated the number of owners but failed to provide any demographic or functional information about the founding group members. The case was removed. The elimination of these 27 organizations reduces the sample size to 803 emergent organizations, of which 388 are solo-owner and 415 are multi-owner.

The inclusion of startups involving a single owner-manager warrants special attention in the following analyses. In effect, we argue that such startup endeavors *could* have involved teams of entrepreneurs, if only the level of interpersonal trust or functional complementarity among participants in the startup assistance network were higher. To evaluate this counterfactual, analysis of the solo-owner organizations had to be restricted to the 278 startups that have at least one external helper. This necessarily led to the omission of the most isolated entrepreneurs – those who could not find another individual to participate in their startup assistance network, much less someone to join them as an owner-manager of the venture. Given the inherent possibility of selection bias in this sampling approach, we urge caution in interpreting the results that include solo-owner organizations, as presented below.<sup>5</sup>

### *Measures*

The data of interest in this paper come from items on the phone interview that were designed to collect information about (1) the equity stake of each

participant involved in the emergent organizations – legal owner (owner-manager or investor) or helper (e.g. consultant, supporter, employee), (2) resources that each participant was bringing into the venture, (3) relationships between participants, and (4) sociodemographic characteristics of participants. During the PSED phone interview, the legal owners were identified by asking the interview respondents, “How many people will (or now) legally own this new business – only you, only you and your spouse, or you and other people or businesses?” If the respondent indicated others would share ownership in the venture, they were asked to identify up to five who would have highest level of ownership, and the percentage to be held by each team member. The outside helpers were identified by asking the interview respondent, “Are there other people, those that would NOT be on the startup team, who have been particularly helpful to you in getting the business started?” If the interview respondent answered yes, they were asked about the number of those outsider helpers, again up to a limit of five on the roster. Once all the legal owners and outsider helpers were identified, interview respondents provided information about each owner and outside helper, including the resources each owner and outside helper contributed to the emergent organization, the relationships between the each owner/outside helper and the interview respondent, and the sociodemographic characteristics of each owner and outside helper.

The types of *resources* contributed to the business startup include: introduction to other people, general information or advice, business-related training, financial assistance, physical resources, business services, personal services, moral or emotional support, labor, creativity, or ideas. We address the coding of this variable, with respect to asset-specificity, in greater detail below. The nature of the *relationships* between each other owner/outside helper and the focal respondent were classified into five categories: spouses/partners, relatives/family members, business associates/work colleagues, friends/acquaintances, and strangers before working together for the startup.<sup>6</sup> The *sociodemographic* characteristics include age, gender, and ethnicity.

The dependent variable is dichotomous, coded ‘1’ if an individual is an organizational insider (legal owner) and ‘0’ if they are an organizational outsider (any other startup network participant). To ensure that the owners are active owner-managers, involved in the daily operation of the startups, instead of passive owners (i.e. investors), all the owners in analysis had been working at least 15 hours per month for the emergent organization from the date of organizing the startup team to the date of final interview. It should be noted that our results are not sensitive to this criterion. The results are very similar when the owners’ time investment is not considered



or when a lower threshold (at least 5 hours per month) is applied for active participation.

To remove the concern that employees may be listed as external consultants/helpers, we also restrict the analysis to startups that have not hired any employees. Again, our results are not sensitive to this criterion and yield similar results when employers are included. Like most startups (see Aldrich & Ruef, 2006, Chapter 4), we find that only a small percentage (13%) of businesses in the PSED have hired employees in their earliest stage of development.

We analyze the multi-owner startups separately from a pooled sample of solo- and multi-owner organizations because none of the solo-owners provides information about his/her own resource contributions to the emergent organization.<sup>7</sup> In all, 584 individuals are either legal owners or helpers from the 274 multi-owner organizations. Of these 584 individuals, 286 are legal owners (48.97%) and 298 (51.03%) are helpers. Focal respondents are excluded from the sample of individuals in the analysis of startups (see Fig. 2), thus permitting the estimation of network effects. By definition, then, all 445 individuals from the solo-owner organizations are helpers.

## INDEPENDENT AND CONTROL VARIABLES

Independent variables include the asset-specificity of an individual's resource contribution to the emergent organization, the redundancy of the individual's resource contribution, the strength of the tie of the individual to the focal respondent, the deviation of the individual's age from the mean in the startup's network of participants, the frequency that other participants in the network match the gender of the individual, and the frequency that other participants match the ethnicity of the individual.<sup>8</sup>

We control for the prior startup experience of each participant in the startup assistance network, since individuals with such human capital could have more influence than those without experience. It is a dichotomous variable, coded as '1' if an individual has prior startup experience and '0' otherwise. We also include a series of dummy variables for the industries within which the startup organizations were located. Prior research suggests that technological complexity differs widely across industries (Klevorick, Levin, Nelson, & Winter, 1995). Because the most commonly used measure of uncertainty was the volatility of technology (David & Han, 2004), industry dummies should capture some cross-industry variation in uncertainty. The

industry dummies also take into account other unobserved cross-industry heterogeneity, such as variation in capital intensity and competition.

Using a combination of closed- and open-response items, the phone questionnaire allowed us to identify 10 different types of resource contributions often made by helpers or founders. All the resource variables are dichotomous, coded '1' if they are provided by a participant and '0' otherwise. Table 1 shows the descriptive statistics and correlation matrix. Since contributions involving creativity/ideas are highly correlated with labor contributions (correlation = 0.72), they are examined in separate models below to avoid multicollinearity.

Table 2 provides definitions and detailed measures for other key independent variables. Following listwise deletion for missing independent variables, we analyze 274 multi-owner and 170 solo-owner startups. Descriptive statistics and correlations are shown in Table 3. An individual's resource redundancy is operationalized as the extent to which the most important resource contribution of the individual overlaps with that of other people in the founding group.<sup>9</sup>

## OPERATIONALIZING ASSET-SPECIFICITY

The criterion demarcating asset-specific and non-asset-specific resources is whether one resource is a specialized asset to a particular organization and, therefore, more valuable in the context of that organization than elsewhere. While this definition is relatively clear in the abstract, empirical applications of TCE admit to considerable variation in the treatment of asset-specificity (David & Han, 2004). Accordingly, we classified resources into three categories: (a) those that are generally treated as asset-specific in the TCE literature (i.e. specialized training in business-related tasks/skills or ideas for a particular business); (b) those that may or may not be treated as asset-specific (introductions to other people, physical resources [such as land], personal services, and other labor); and (c) those that are generally not treated as asset-specific (general information or advice, general business services, financial capital, emotional support).

To consider the potential sensitivity of our findings to this classification, an exploratory analysis was conducted, examining the impact of the 10 different resource contributions on the probability that a given participant would become an organization insider. Contrary to the expectations of TCE, Table 4 suggests that several resource contributions with low or uncertain asset-specificity – including introductions to other people, physical

**Table 1.** Descriptive Statistics and Correlations for Resource Contribution Variables (Dummy).

Variable	Mean	S.D.	1	2	3	4	5	6	7	8	9	10
1. Intro to other people	0.67	0.47	...									
2. General information or advice	0.91	0.29	0.17	...								
3. Training in business-related tasks/skills	0.53	50.00	0.06	0.17	...							
4. Financial resources	0.29	0.45	0.04	-0.02	0.03	...						
5. Physical resources	0.41	0.49	0.14	0.04	0.16	0.22	...					
6. General business services	0.37	0.48	0.05	0.12	0.16	0.19	0.18	...				
7. Personal services	0.31	0.46	-0.04	-0.10	-0.02	0.11	0.21	0.13	...			
8. Moral or emotional support	0.20	0.40	-0.04	-0.01	-0.06	-0.01	-0.03	-0.06	0.05	...		
9. Creativity or ideas to business	0.31	0.46	0.05	0.08	0.01	0.02	0.03	0.06	0.11	0.62	...	
10. Labor	0.23	0.42	0.06	0.06	0.09	0.02	0.16	0.11	0.18	0.46	0.72	...

Note: All pairwise correlations at magnitude 0.08 or above (absolute value) are significant at  $p < 0.05$ .

**Table 2.** Definition of Key Independent Variables.

Variable	Definition
Resource asset-specificity	Coded 1 if an individual provided training in business-related tasks/skills or ideas/creativity to the business; 0 if more general resources provided (e.g. introduction to other people, general information and advice, financial assistance, physical resources, labor, etc) <sup>a</sup>
Resource redundancy	Extent to which the most important resource contribution of an individual overlaps with that of other people in the founding group. The redundancy score is between 0 and 1. The score would be closer to 0 if the individual's most important resource is unique. The score is 1 if the individual's most important resource contribution overlaps with that of all other people
Strength of the ties to the focal respondent	5 = Spouses/partners; 4 = relatives/family members; 3 = business associates/work colleagues; 2 = friends/acquaintances; and 1 = strangers before joining the team
Age deviation from the mean	Absolute value of the difference between each individual's age and the mean of the group age distribution
Gender frequency	Frequency with which other people share the gender of an individual. The standardized score is between 0 and 1. The score would be closer to 0 if the individual's gender is unique in the founding group; the score is 1 if all other people share the individual's gender
Ethnic frequency	Frequency with which other people share the ethnicity of an individual. The standardized score is between 0 and 1. The score would be closer to 0 if the individual's ethnicity is unique in the founding group; the score is 1 if all other people share the individual's ethnicity

<sup>a</sup>Sensitivity analyses were also conducted for alternative coding scheme, as noted in text.

resources, general business services, and labor – exercise positive and significant effects. Resources with high asset-specificity, on the other hand, such as specialized training and ideas for the business, have either marginal or insignificant effects. These exploratory results suggest that two different

**Table 3.** Descriptive Statistics and Correlations for Key Independent Variables.

Variable	Mean	S.D.	Minimum	Maximum	1	2	3	4	5	6
1. Resource asset-specificity	0.68	0.47	0	1	...					
2. Resource redundancy	0.22	0.28	0	1	0.07	...				
3. Strength of the tie to focal respondent	3.19	1.20	1	5	-0.09	0.02	...			
4. Age deviation from the mean	6.17	5.95	0	30.8	-0.02	-0.07	0.01	...		
5. Gender frequency	0.56	0.35	0	1	0.06	0.02	-0.49	0.04	...	
6. Ethnic frequency	0.78	0.32	0	1	-0.08	0.03	0.13	0.09	0.15	...

*Note:* All pairwise correlations at magnitude 0.075 or above (absolute value) are significant at  $p < 0.05$ .

**Table 4.** Effects of Assets-Specificity on Organizational Insider (Legal Owner) vs. Outsider (Helper) Status.

Variable	Asset-Specificity	1	2
<b>Resource contribution</b>			
Training in business-related tasks/skills	High	0.18 (0.20)	0.09 (0.20)
Creativity or ideas for business	High	0.74 <sup>†</sup> (0.43)	
Labor	Uncertain		1.38 <sup>***</sup> (0.38)
Introduction to other people	Uncertain	0.43 <sup>*</sup> (0.20)	0.43 <sup>*</sup> (0.21)
Physical resources (e.g. use of land or space)	Uncertain	0.88 <sup>***</sup> (0.20)	0.77 <sup>***</sup> (0.20)
Personal services (e.g. household help or childcare)	Uncertain	0.44 <sup>*</sup> (0.22)	0.38 <sup>†</sup> (0.22)
General information or advice	Low	-0.60 <sup>†</sup> (0.34)	-0.62 <sup>†</sup> (0.34)
General business services (e.g. legal or accounting services)	Low	0.74 <sup>***</sup> (0.20)	0.71 <sup>**</sup> (0.20)
Financial resources (e.g. loans or equity)	Low	0.20 (0.22)	0.23 (0.22)
Moral or emotional support	Low	-1.76 <sup>***</sup> (0.33)	-1.93 <sup>***</sup> (0.36)
<b>Control variables</b>			
Prior startup experience		-0.40 <sup>*</sup> (0.19)	-0.31 (0.20)
Agriculture		0.86 (0.93)	1.00 (0.89)
Construction		0.01 (0.77)	0.10 (0.72)
Manufacturing		-0.82 (0.58)	-0.73 (0.53)
Wholesale		-0.89 (0.74)	-0.78 (0.68)
Retail		-0.74 <sup>†</sup> (0.44)	-0.62 <sup>†</sup> (0.37)
Business service		-0.40 (0.44)	-0.30 (0.37)
Customer service		-0.56 (0.49)	-0.43 (0.42)
Healthcare/education		-0.71 (0.57)	-0.56 (0.51)
Constant		0.12 (0.44)	0.09 (0.41)
Within-group correlation ( $\alpha$ )		0.10	0.08
Degrees of freedom		18	18
Number of cases		583	583

Note: <sup>†</sup> $p < 0.1$ ; <sup>\*</sup> $p < 0.05$ ; <sup>\*\*</sup> $p < 0.01$ ; and <sup>\*\*\*</sup> $p < 0.001$  (two-tailed tests).

Standard errors are in parentheses.

codings of asset-specificity can be pursued: for a conservative test of the effects of asset-specificity, only contributions pertaining to specialized training or ideas should be included, while a more generous test would include the variety of resources that evidence uncertain asset-specificity. It should be noted that our multivariate results (as indicated below) are not sensitive to which operationalization is employed. Although we report results for the conservative test of asset-specificity, findings are very similar when we include resources with uncertain asset-specificity.

The exploratory results also suggest that some caution must be exercised in interpreting the effects of asset-specificity in the absence of variables that address social network effects. For instance, although personal service contributions increase the probability of inclusion among startup owners, the provision of household help or childcare is likely to be a spurious criterion for the selection of business owners. Rather, these types of contributions tend to tap into unmeasured kinship ties, which are included in our multivariate models below.

## METHOD

Because individual-level observations are clustered within startups, the probability that one individual becomes an organizational insider may depend on the likelihood that his/her peers are selected as insiders. To provide accurate prediction of organizational insiders versus outsiders in the presence of clustered observations, we employ generalized estimating equations (GEE) (Liang & Zeger, 1986). The GEE regression model is

$$g(E[Y_{ij}|x_{ij}]) = x'_{ij}\beta \quad (1)$$

where  $x_{ij}$  is a  $p$  times 1 vector of covariates,  $\beta$  consists of the  $p$  regression parameters of interest,  $Y_{ij}$  denotes the  $j$ th individual (for  $j = 1, \dots, J$ ) in the  $i$ th founding group (for  $I = 1, \dots, N$ ), and  $g(\cdot)$  is the link function. Since the dependent variable is a binary variable (legal owner or outside helper), we use the logit link function:<sup>10</sup>

$$g(E[Y_{ij}|x_{ij}]) = \log\left(\frac{E[Y_{ij}|x_{ij}]}{1 - E[Y_{ij}|x_{ij}]}\right) \quad (2)$$

## RESULTS

Table 5 reports the estimated coefficients from GEE regression analyses of the legal owner or outside helper status for the 584 individuals from the multi-owner startups (Models 1–6) as well as a pooled sample of 1,029 entrepreneurs from solo- and multi-owner enterprises (Models 7–8). We do not test the transaction cost or RBV arguments for the pooled sample, since none of the solo-owners provide information about their own resource contributions to the business.

**Table 5.** GEE Regression Coefficients Predicting Organizational Insider (Legal Owner) vs. Outsider (Helper).

Variable	Multi-Owner						All Startups	
	1	2	3	4	5	6	7	8
Resource asset-specificity	0.22 (0.19)				0.30 (0.20)	0.35 (0.21)		
Resource redundancy		0.14 (0.31)			-0.01 (0.33)	-0.07 (0.34)		
Tie strength to respondent			0.47*** (0.08)		0.57*** (0.09)		0.58*** (0.08)	
Spouse						3.28*** (0.41)		2.67*** (0.28)
Relative						0.34 (0.25)		0.41* (0.21)
Work colleague						-0.01 (0.26)		-0.03 (0.21)
Age deviation from mean				-0.06*** (0.01)	-0.07*** (0.01)	-0.05** (0.02)	-0.06*** (0.01)	-0.04** (0.01)
Gender frequency				-0.27 (0.25)	0.71* (0.31)	1.61*** (0.35)	0.50* (0.25)	1.20*** (0.28)
Ethnic frequency				0.28 (0.28)	0.17 (0.29)	0.36 (0.31)	0.22 (0.26)	0.34 (0.27)
Control variables								
Prior startup experience	-0.34* (0.17)	-0.33* (0.17)	-0.16 (0.18)	-0.29* (0.17)	-0.16 (0.18)	-0.07 (0.19)	-0.34* (0.15)	-0.27 (0.15)
Agriculture	1.47* (0.82)	1.42* (0.82)	0.99 (0.87)	1.50* (0.82)	1.04 (0.84)	0.59 (0.84)	1.13 (0.71)	0.87 (0.73)
Construction	-0.01 (0.64)	-0.05 (0.63)	-0.36 (0.67)	-0.01 (0.63)	-0.31 (0.66)	-0.40 (0.71)	-0.72 (0.55)	-0.72 (0.57)
Manufacturing	-0.45 (0.45)	-0.56 (0.44)	-0.76 (0.47)	-0.40 (0.44)	-0.39 (0.47)	-0.39 (0.49)	-0.66 (0.45)	-0.71 (0.46)
Wholesale	-0.41 (0.57)	-0.47 (0.57)	-0.73 (0.60)	-0.42 (0.57)	-0.58 (0.60)	-0.80 (0.65)	-0.73 (0.55)	-0.98 (0.57)
Retail	-0.41 (0.25)	-0.46* (0.25)	-0.72** (0.27)	-0.39 (0.25)	-0.66** (0.27)	-0.76** (0.28)	-0.82** (0.25)	-0.95*** (0.26)
Business service	-0.04 (0.26)	-0.11 (0.25)	-0.11 (0.27)	-0.14 (0.24)	-0.01 (0.27)	-0.18 (0.28)	-0.35 (0.25)	-0.46 (0.25)
Consumer service	-0.18 (0.32)	-0.24 (0.32)	-0.43 (0.34)	-0.20 (0.31)	-0.34 (0.34)	-0.42 (0.34)	-0.66* (0.31)	-0.83** (0.31)
Healthcare/education	-0.24 (0.41)	-0.27 (0.41)	-0.47 (0.44)	-0.25 (0.39)	-0.42 (0.43)	-0.38 (0.42)	-1.34** (0.39)	-1.34** (0.40)
Constant	0.23 (0.23)	0.37* (0.18)	-1.04*** (0.30)	0.72 (0.18)	-1.76*** (0.50)	-1.42** (0.42)	-2.09*** (0.41)	-1.39*** (0.34)
Wald $\chi^2$	13.72	12.62	44.75	29.29	63.48	92.85	108.43	144.53
Within-group correlation ( $\alpha$ )	0.06	0.05	0.10	0.01	0.04	0.04	0.25	0.20
Degrees of freedom	10	10	10	10	15	17	13	15
Number of cases				584			1,029	

Note: \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; and \*\*\*  $p < 0.001$  (one-tailed tests for hypothesized effects, two-tailed otherwise).

Standard errors are in parentheses.



In the initial series of multi-owner models, independent variables are entered separately to explore empirical support for Hypotheses 1–6. The equation in Model 1 includes one explanatory variable of interest, resource asset-specificity, whose coefficient is statistically insignificant. This is consistent with the exploratory analyses of individual resource variables reported in Table 4, which show that only one of the two organization-specific resources – creativity/ideas – has a marginally significant impact on insider status. Hence, while the coefficient is in the expected direction, the TCE hypothesis is not supported. In Model 2, the coefficient for resource redundancy is insignificant and positive. This result is contrary to the functionalist expectation that individuals who provide more redundant (less unique) resources are less likely to become organizational insiders (Hypothesis 2).

Hypothesis 3 asserts that the stronger the tie between an ego and an alter, the more likely that alter will become an organizational insider. The strength of the tie to the focal respondent is positive and statistically significant at the  $p < 0.001$  level (Model 3). The expected odds that the assistance network participant will become an owner-manager increases by  $(1.60 - 1)(100\%) = 60\%$  per degree of the tie strength with the focal respondent.

Our next specification (Model 4) incorporates variables examining the deviation of each network participant from the demographics of the assistance network on the whole. Hypothesis 4 – asserting that the closer an individual's age is to the mean of the founding group's age distribution, the more likely the individual is to become an organizational insider – is strongly supported. The coefficient of the age deviation from the group mean is negative and significant at  $p < 0.001$  (Model 4). The expected odds of becoming a legal owner decrease by 6% for every one year of age deviation from the group mean. Hypothesis 4 is also strongly supported in the analysis of the pooled sample (Model 7).

Surprisingly, Hypothesis 5 is not supported in the model excluding other independent variables of theoretical interest. The coefficient of gender frequency is insignificant, contrary to our expectation that the higher the frequency that other people match the gender of an individual in the startup assistance network, the more likely the individual would become an organizational insider. However, marital ties between individual founding members could account for this unexpected result.

The coefficient for ethnic frequency is statistically insignificant. Hence, Hypothesis 6 – claiming that organizational insiders tend to be chosen based on ethnic similarity – is not supported. The coefficient of ethnic frequency in the pooled sample is also insignificant and even lower in magnitude

than that estimated for multi-owner organizations. This might reflect a mechanism where entrepreneurs tend to self-select into solo-owner organizations when they are unable to cooperate with other founders from the same ethnic background. The small numbers of ethnic minorities place them at disproportionate risk of isolation in entrepreneurial activities (Ruef et al., 2003).

Model 5 combines all predictors in a multivariate specification. Considering the two variables pertaining to functionality, both resource asset-specificity (0.30) and redundancy ( $-0.01$ ) are in the expected direction, but neither is statistically significant. However, three of the four sociodemographic variables pertaining to trust are statistically significant and in the expected direction. The coefficient for the strength of network ties is positive (0.57,  $p < 0.001$ ); the coefficient of the age deviation is negative ( $-0.07$ ,  $p < 0.001$ ); and, after controlling for strength of interpersonal ties, the coefficient of the gender frequency is positive (0.71,  $p < 0.05$ ). The latter finding suggests that the negative coefficient of gender frequency in Model 4 is spurious, resulting from marital ties among business owners. Findings from the pooled sample (Model 7) suggest that these mechanisms are not substantively different when we include startups with solo owner-managers.

Two final specifications (Models 6 and 8) address heterogeneity in network tie strength – in particular, the concern that spousal relationships may be the principal driver leading to insider startup roles rather than other types of network ties. Our estimates for both the multi-owner and pooled sample suggest that the presence of spouses in the startup assistance network is very strongly associated with insider status. For instance, in the sample of all startups, members of the assistance network are 14 times as likely to be owner-managers when they are married to the focal respondent. In the pooled sample, however, there is also evidence that kinship ties function to pull or push startup participants into owner-manager roles. Thus, participants are 50% more likely to become insiders when they are relatives of the focal respondent.

The effects of a number of control variables are also worth noting. Prior startup experience has a significantly negative effect on the selection of owner-managers in four out of the eight models. Having played the role of founding owner-managers in the past, individuals with prior startup experience are often more inclined to become consultants or “angel” investors in the startup assistance network (Aldrich & Ruef, 2006, p. 88). Among the industry dummy controls, agricultural and retail startups have significant positive and negative effects, respectively, in a number of models. The former finding reflects the fact that agricultural businesses are often

family-owned and operated, thus lending themselves to a larger set of owner-managers. By contrast, given their relatively low capital requirements, retail enterprises tend to have a small set of owner-managers.

## DISCUSSION

The processes contributing to boundary formation have been explored from a variety of perspectives in organizational theory, but those theories primarily focus on how well-established organizations shape their boundaries and offer little guidance as to how organizational boundaries emerge in the first place. In this article, we sought to explain the boundary formation in emergent organizations by contrasting the functional considerations that tend to be emphasized in economic approaches with the interpersonal considerations that come into play in sociological perspectives. Analyzing a nationally representative dataset of U.S. nascent entrepreneurs, our results suggest that, in startups, the boundary decision is based more on a potential member's trustworthiness, as perceived by peers, than on the functional contribution that the individual could bring into the organization.

Why do transaction cost and resource-based views of the firm perform poorly for startups? One argument hinges on the lack of economic sophistication among many nascent entrepreneurs. Analyses of the PSED dataset suggest that 26.5% of early-stage entrepreneurs fail to ignore sunk costs, a common benchmark of managerial rationality (Morgan, 2004). By the same token, it could be argued, nascent entrepreneurs often ignore transaction costs or resource complementarities in boundary decisions. Naturally, the TCE and RBV perspectives may nevertheless hold true in an evolutionary sense if, as time goes by, the firms initiated by unsophisticated entrepreneurs are subject to higher rates of failure. This argument would be consistent with the empirical support that these perspectives have garnered in predicting boundary-making decisions and performance outcomes for large, established firms (Shelanski & Klein, 1995; Barney, 1991).

An alternative account, based on social exchange theory (Coleman, 1990; Ekeh, 1974), allows for greater sophistication on the part of nascent entrepreneurs. In this explanation, the finding that boundary formation occurs on the basis of perceived trustworthiness reflects the nature of group-generalized exchange (Ekeh, 1974) in emergent organizations. This exchange structure involves a typical social dilemma situation. As argued before, if we assume that each participant receives a specified share of the total benefits generated by resource pooling (through an equity stake) but that their

contributions (in terms of ideas, specialized skills, labor time, etc.) cannot be contracted for *ex ante*, it is rational for founders to engage in some opportunistic shirking. Trust is critical for generalized exchange because nascent entrepreneurial team members typically need to have full cooperation to survive the “liability of newness” in startups (Stinchcombe, 1965). Our findings support the intuition that individuals are more likely to trust those with whom they have strong ties and those who share similar social characteristics. Moreover, interpersonal ties and similar social characteristics not only allow an entrepreneur to trust that a startup participant is less likely to free ride, but also that the participant is more likely to have consistent views in business decision-making.

A third interpretation of our results also emphasizes interpersonal networks, but questions the calculative conceptualization suggested by social exchange theory. In this account, the key dilemma in recruiting fellow owner-managers for the startup is not opportunism at all, but interpersonal access and influence (see Bian, 1997 for a related discussion of job-seeking behavior). Because of the high rate of failure in most business startups, entrepreneurship tends to be regarded as a “foolish” activity (Aldrich & Fiol, 1994). Consequently, many individuals in the startup assistance network prefer passive to active involvement. Arguably, it is only through the solidarity afforded by network ties and homophily that nascent entrepreneurs are able to recruit others into more active involvement in these risky endeavors.

The alternative interpretations of our findings thus suggest three mechanisms that must be evaluated in future research. One of the most promising avenues for adjudicating between these mechanisms is to examine the link between the boundary formation process and the high dissolution rates of startups. According to historical surveys by the U.S. Department of Commerce, roughly two-thirds of startups close within five years (Dickinson, 1981). Based on the TCE and RBV perspectives, most of these failures will be accounted for by either a lack of transaction cost minimization or resource complementarity. The presence of such functional features among larger and more mature organizations can therefore be seen as a result of an evolutionary selection mechanism. Social exchange theorists, on the other hand, may suggest that inequity in contributions of time and ideas tends to be the culprit in the failure of organizational startups. Presumably, the more owner-managers are selected based on strong network ties and homophily, the more likely the owner-managers are to share similar information and values, and consequently, the more likely an emergent organization is to be equitable. The third mechanism argues for the importance of solidarity as a

recruitment tool for committed organizational founders. In this account, the inability to achieve such solidarity impacts organizational failure through smaller teams of owner-managers and higher turnover among them.

Following the recent organizational literature (e.g. Scott, 2004; Aldrich & Ruef, 2006), a natural extension to the present study could also examine the evolution of organizational boundaries. Such an evolutionary perspective leads to a number of analytical questions. What are the effects when startup participants switch roles across organizational boundaries? In this respect, have startups become more open and flexible over time? What are the implications of new work arrangements and “network” forms of organizations for the types of roles that tie individuals to emergent organizations (DiMaggio, 2001)? If organizational startups represent the “canary in the coal mine” for many of these societal developments, attention to their boundary formation processes promises to yield rich substantive insights.

## NOTES

1. Explanations of the boundary formation process can be roughly categorized into two types: one that highlights more rigid and instrumentally defined boundaries (i.e. “rational” system perspectives) and another that highlights the blurred nature of organizational boundaries (“natural” system perspectives) (see Scott, 2003).

2. Uncertainty, another important transaction attribute, is predicted to affect governance form only in the presence of asset-specificity (Williamson, 1985). In other words, the effects of uncertainty are contingent on the level of asset-specificity.

3. Measures that operationalize such skills also provide some of the strongest empirical support for TCE (David & Han, 2004).

4. It should be acknowledged that this argument is *not* inherently at odds with the logic of transaction cost economics, especially given the importance that TCE places on “calculative trust” (Williamson, 1993). However, the empirical TCE literature contains virtually no discussion of interpersonal network ties.

5. One analytical solution to this sample selection issue would be to generate a hypothetical pool of startup helpers, based on the characteristics of assistance networks for *all* startups in the sample. However, this exercise would push the focus of the analysis toward the determinants of assistance network composition rather than boundary delineation *per se*.

6. It should be acknowledged that we do not have fine-grained data on the strength of these ties, addressing dimensions such as emotional involvement, frequency of contact, and reciprocity (Granovetter, 1973). Our rank ordering of the strength of ties – leading to the intuition, for instance, that a spousal relationship is stronger than one involving a relative – is therefore indirect.

7. This is a function of the PSED research design. Questions concerning an internal division of labor and resource contributions among team members are only held to be meaningful in the multi-owner startup teams.

8. Wealth and income levels are not controlled for, since past research using the PSED data has found that personal financial resources are not significantly related to entry into entrepreneurial activity (see Kim, Aldrich, & Keister, 2006). However, this finding has been contested by analysts using other datasets.

9. The resource redundancy scores are based on each individual's most important resource among his/her all resource contributions. Suppose there are  $n$  individuals in a founding group, the redundancy score of individual  $j$  is calculated as  $\sum_{i=1, i \neq j}^n \delta_{ij} / n - 1$ , where  $\delta_{ij} = 1$  if  $j$ 's most important resource contribution is the same as  $i$ 's;  $\delta_{ij} = 0$  otherwise. The formula is also used for the calculation of gender and ethnic frequency scores.

10. The within-group correlation structure is treated as exchangeable, which assumes that individual observations covary equally within each startup's network of participants.

## ACKNOWLEDGMENTS

We acknowledge the National Science Foundation and Kauffman Foundation for their assistance in making the PSED data available. Earlier versions of this paper were presented at the Academy of Management Meetings in New Orleans, the American Sociological Association Meetings in San Francisco in 2004, and the London Business School. We are grateful to Howard Aldrich, James Baron, Diane Burton, Silvia Dorado, Robert Freeland, Michael Hannan, Phil Kim, Mike Lounsbury, Huggy Rao, and Jesper Sorensen for providing insightful feedback.

## REFERENCES

- Aldrich, H. E., & Baker, T. (2001). Learning and legitimacy: Entrepreneurial responses to constraints on the emergence of new populations and organizations. In: C. B. Schoonhoven & E. Romanelli (Eds), *The entrepreneurial dynamic: Origins of entrepreneurship and the evolution of industries* (pp. 207–235). Stanford, CA: Stanford University Press.
- Aldrich, H. E., & Fiol, C. M. (1994). Fools rush in? The institutional context of industry creation. *The Academy of Management Review*, 19, 645–670.
- Aldrich, H. E., & Ruef, M. (2006). *Organizations evolving* (2nd ed.). Thousand Oaks, CA: Sage.
- Aldrich, H. E., & Waldinger, R. (1990). *Ethnicity and entrepreneurship*. *Annual Review of Sociology*, 16, 111–135. Palo Alto, CA: Annual Reviews, Inc.
- Amit, R., & Schoemaker, P. J. H. (1993). Strategic assets and organizational rents. *Strategic Management Journal*, 14, 33–46.
- Barney, J. B. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, 17, 99–120.
- Bian, Y. (1997). Bringing strong ties back in: Indirect ties, network bridges, and job searches in China. *American Sociological Review*, 62, 366–385.

- Bigelow, L. S. (2004). *Make-or-buy revisited: A longitudinal, population-wide test of transaction cost alignment*. Working Paper. Olin School of Business, Washington University.
- Carroll, G. (1983). A stochastic model of organizational mortality: Review and reanalysis. *Social Science Research, 12*, 303–329.
- Carter, N. (1994). Reducing barriers between genders: Differences in new firm startups. Paper presented at the Academy of Management Meetings, Dallas, TX.
- Coase, R. H. (1937). The nature of the firm. *Economica, 4*, 386–405.
- Coleman, J. (1990). *Foundations of social theory*. Cambridge, MA: Harvard University Press.
- David, R. J., & Han, S-K. (2004). A systematic assessment of the empirical support for transaction cost economics. *Strategic Management Journal, 25*, 39–58.
- Dickinson, R. (1981). Business failure rate. *American Journal of Small Business, 6*, 17–25.
- DiMaggio, P. (2001). *The twenty-first century firm*. Princeton, NJ: Princeton University Press.
- Ekeh, P. P. (1974). *Social exchange theory: Two traditions*. Cambridge, MA: Harvard University Press.
- Gartner, W. B., Shaver, K. G., Carter, N. M., & Reynolds, P. D. (Eds) (2004). *Handbook of entrepreneurial dynamics: The process of business creation*. Thousand Oaks, CA: Sage.
- Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology, 78*, 1360–1380.
- Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *American Journal of Sociology, 91*, 481–510.
- Ibarra, H. (1997). Paving an alternative route: Gender differences in managerial networks. *Social Psychology Quarterly, 60*, 91–102.
- Kalleberg, A., Knoke, D., Marsden, P., & Spaeth, J. (1996). *Organizations in America: Analyzing their structures and human resource practices*. Thousand Oaks, CA: Sage.
- Kim, P. H., Aldrich, H. E., & Keister, L. A. (2006). Access (not) denied: The impact of financial, human capital, and cultural capital on entrepreneurial entry in the United States. *Small Business Economics, 27*, 5–22.
- Klein, B. (1988). Vertical integration as organizational ownership: The Fisher Body–General Motors relationship revisited. *Journal of Law, Economics and Organization, 4*, 199–213.
- Klevorick, A. K., Levin, R. C., Nelson, R. R., & Winter, S. G. (1995). On the sources and significance of interindustry differences in technological opportunities. *Research Policy, 24*, 185–205.
- Kogut, B., & Zander, U. (1996). What firms do? Coordination, identity, and Learning. *Organization Science, 7*(5), 502–523.
- Lazarsfeld, P., & Merton, R. K. (1954). Friendship as social process: A substantive and methodological analysis. In: M. Berger, T. Abel & C. Page (Eds), *Freedom and control in modern society* (pp. 18–66). New York: Octagon Books.
- Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73*, 13–22.
- McPherson, J. M., & Smith-Lovin, L. (1982). Women and weak ties: Differences by sex in the size of voluntary organizations. *American Journal of Sociology, 87*, 883–904.
- McPherson, J. M., & Smith-Lovin, L. (1987). Homophily in voluntary organizations: Status distance and the composition of face to face groups. *American Sociological Review, 52*, 370–379.
- Meyer, M., & Lu, X. (2005). Managing indefinite boundaries: The strategy and structure of a Chinese business firm. *Management and Organization Review, 1*, 57–86.
- Morgan, J. N. (2004). On economic sophistication. In: W. Gartner, K. Shaver, N. Carter & P. Reynolds (Eds), *Handbook of entrepreneurial dynamics: The process of business creation* (pp. 214–219). Thousand Oaks, CA: Sage.

- Pfeffer, J. (1983). Organizational demography. In: L. L. Cummings & B. M. Staw (Eds), *Research in organizational behavior* (Vol. 5). Greenwich, CT: JAI Press.
- Pfeffer, J. (1997). *New directions for organization theory: Problems and prospects*. New York: Oxford University Press.
- Pfeffer, J. (2003). Introduction to the classic edition. In: J. Pfeffer & G. Salancik (Eds), *The external control of organizations* (pp. xi–xxix). Stanford, CA: Stanford University Press.
- Pfeffer, J., & Salancik, G. (1978). *The external control of organizations: A resource dependence perspective*. New York: Harper & Row.
- Podolny, J. M., & Baron, J. (1997). Resources and relationships: Social networks and mobility in the workplace. *American Sociological Review*, 62, 673–693.
- Portes, A., & Sensenbrenner, J. (1993). Embeddedness and immigration: Notes on the social determinants of economic action. *American Journal of Sociology*, 98, 1320–1350.
- Pralahad, C. K., & Hamel, G. (1990). The core competence of the organization. *Harvard Business Review*, 68, 79–91.
- Reskin, B. (1999). The determinants and consequences of workplace sex and race composition. *Annual Review of Sociology*, 25, 335–361.
- Reynolds, P. D. (2000). National panel study of U.S. business start-ups: Background and methodology. In: J. Katz (Ed.), *Advances in entrepreneurship, firm emergence, and growth* (Vol. 4). Stamford, CT: JAI Press.
- Rindova, V. P., & Fombrun, C. J. (2001). Entrepreneurial action in the creation of the specialty coffee niche. In: C. B. Schoonhoven & E. Romanelli (Eds), *The entrepreneurial dynamic: Origins of entrepreneurship and the evolution of industries* (pp. 236–261). Stanford, CA: Stanford University Press.
- Ruef, M. (2002). Unpacking the liability of aging: Toward a socially-embedded account of organizational disbanding. *Research in the Sociology of Organizations*, 19, 195–228.
- Ruef, M., Aldrich, H., & Carter, N. (2003). The structure of founding teams: Homophily, strong ties, and isolation among U.S. entrepreneurs. *American Sociological Review*, 68, 195–222.
- Santos, F. M. (2003). *Constructing niches and shaping boundaries: Entrepreneurial action in nascent fields*. Unpublished dissertation. Department of Management Science and Engineering, Stanford University, Stanford, CA.
- Santos, F. M., & Eisenhardt, K. M. (2005). Organizational boundaries and theories of organization. *Organization Science*, 16(5), 491–508.
- Sarasvathy, S. (2001). Causation and effectuation: Toward a theoretical shift from economic inevitability to entrepreneurial contingency. *Academy of Management Review*, 26(2), 243–288.
- Schofield, J. W. (1995). Review of research on school desegregation's impact on elementary and secondary school students. In: J. Banks & C. McGee (Eds), *Handbook of research on multicultural education*. New York: Macmillan.
- Scott, W. R. (2003). *Organizations: Rational, natural and open systems* (5th ed.). Upper Saddle River, NJ: Prentice-Hall.
- Scott, W. R. (2004). Reflections on a half-century of organizational sociology. *Annual Review of Sociology*, 30, 1–21.
- Shelanski, H. A., & Klein, P. (1995). Empirical research in transaction costs economics: A review and assessment. *Journal of Law, Economics, and Organization*, 11, 335–361.
- Stinchcombe, A. (1965). Social structure and organizations. In: J. March (Ed.), *Handbook of organizations*. Chicago: Rand McNally.



- Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, 18(7), 509–533.
- Thompson, J. D. (1967). *Organizations in action*. New York: McGraw-Hill.
- Walsh, J. P. (1995). Managerial and organizational cognition: Notes from a trip down memory lane. *Organization Science*, 6(3), 280–321.
- Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic Management Journal*, 5, 171–180.
- Williamson, O. E. (1975). *Markets and hierarchies: Analysis and antitrust implications*. New York: Free Press.
- Williamson, O. E. (1981). The economics of organizations: The transaction cost approach. *American Journal of Sociology*, 87(3), 548–577.
- Williamson, O. E. (1985). *The economic institutions of capitalism*. New York: Free Express.
- Williamson, O. E. (1991). Comparative economic organization: The analysis of discrete structural alternatives. *Administrative Science Quarterly*, 36(2), 269–296.
- Williamson, O. E. (1993). Calculativeness, trust, and economic organization. *Journal of Law and Economics*, 36, 453–486.

# REVISITING THE ENCLAVE HYPOTHESIS: MIAMI TWENTY-FIVE YEARS LATER<sup>☆</sup>

Alejandro Portes and Steven Shafer

## ABSTRACT

*We review the empirical literature on ethnic economic enclaves after the concept was formulated 25 years ago. The balance of this literature is mixed, but many studies reporting negative conclusions were marred by faulty measurement of the concept. We discuss the original theoretical definition of enclaves, the hypotheses derived from it, and the difficulties in operationalizing them. For evidence, we turn to census data on the location and the immigrant group that gave rise to the concept in the first place – Cubans in Miami. We examine the economic performance of this group, relative to others in this metropolitan area, and in the context of historical changes in its own mode of incorporation. Taking these changes into account, we find that the ethnic enclave had a significant economic payoff for its founders – the earlier waves of Cuban exiles – and for their children, but not for refugees who arrived in the 1980 Mariel exodus and*

<sup>☆</sup>Paper originally presented at the Conference on New Directions in Equality and Stratification. Princeton University, April 6–8, 2006. We acknowledge the very helpful commentaries on the original version by Philip Kasinitz, Patricia Fernandez-Kelly, and Min Zhou. Responsibility for the contents is exclusively ours.

---

**The Sociology of Entrepreneurship**  
**Research in the Sociology of Organizations, Volume 25, 157–190**  
Copyright © 2007 by Elsevier Ltd.  
All rights of reproduction in any form reserved  
ISSN: 0733-558X/doi:10.1016/S0733-558X(06)25005-5

*after. Reasons for this disjuncture are examined. Implications of these results for enclave theory and for immigrant entrepreneurship in general are discussed.*

## INTRODUCTION

In 1980, Wilson and Portes published an article in the *American Journal of Sociology* that identified ethnic enclaves as a distinct form of economic adaptation. These formations were characterized by the spatial concentration of immigrants who organize a variety of enterprises to serve their own market as well as the general population. A substantial proportion of workers of the same nationality were employed in these firms and the latter were found in a variety of manufacturing and commercial sectors rather than being limited to a single economic niche. Wilson and Portes' (1980) analysis made three principal claims: (1) that the enclave was a distinct economic sector, separate from the "primary" and "secondary" sectors of the mainstream labor market; (2) that the economic payoff for human capital brought from the home country was higher in the enclave than in the sectors of the mainstream economy that employed immigrant labor; (3) that enclave entrepreneurs received higher economic returns than co-ethnics with comparable human capital, even when the latter were employed in the better sectors of the mainstream economy.

Wilson and Portes provided empirical evidence for their argument on the basis of data from a sample of Cuban exiles who arrived in the United States in the early 1970s and were followed throughout the decade. Discriminant analysis was used to demonstrate the existence of an enclave sector, different from the primary and secondary sectors of the mainstream labor market. Earning regressions, within and across sectors, were then used to support the remaining hypotheses.

The enclave hypothesis garnered attention at the time because it ran contrary to conventional economic and sociological wisdom. The latter argued that ethnic economies were at best an employer of last resort with distinct disadvantages relative to entrance into mainstream sectors of the economy. At worst, they were characterized as "mobility traps" confining immigrant and ethnic minorities to a condition of permanent disadvantage (Borjas, 1986, 1990; Bates, 1987, 1989). The opposite tack was retaken by Portes and Bach (1985) in their comparative analysis of employment and income among Cuban and Mexican immigrant males in the late 1970s. Their

results went beyond those reported earlier by focusing on determinants of enclave entrepreneurship. Using the same data set as Wilson and Portes, they found that the principal determinants of self-employment among Cubans in Miami were parental human capital, education, and most importantly having been employed in a co-ethnic firm at an earlier time (Portes & Bach, 1985, pp. 226–232).

The physical “look” of immigrant enclaves was described as follows:

Near downtown Los Angeles there is an area approximately a mile long where all commercial signs suddenly change from English to strange pictorial characters. Koreatown, as the area is known, contains the predictable number of ethnic restaurants and grocery shops; it also contains a number of banks, import–export houses, industries, and real estate offices. Signs of “English Spoken Here” assure visitors that their links with the outside world have not been totally severed...

A similar urban landscape is found near downtown Miami. Little Havana extends in a narrow strip for about five miles, eventually merging with the southwest suburbs of the city. Cuban firms are found in light and heavy manufacturing, construction, commerce, finance, and insurance. An estimated 60 percent of all residential construction in the metropolitan area is now done by these firms (Portes & Rumbaut, 1996, pp. 20–21).

### *Testing the Enclave Hypothesis*

The interest produced by the notion that ethnic enclaves could be “mobility machines” rather than “traps” led to a number of studies that sought to test the idea or put it to use. Looking back in history, Portes and Bach had identified several similar communities constructed by immigrant groups at the beginning of the 20th century. Two, in particular, deserved attention: the Jewish enclave in the Lower East Side of Manhattan and the Japanese enclave in Los Angeles. Both served as effective platforms for the economic incorporation of these immigrant groups leading, by the third generation, to levels of income and occupational status that exceeded those of the native white population (Portes & Bach, 1985, pp. 38–48).

Early attempts at applying the concept of ethnic entrepreneurial enclaves tended to yield positive results. Zhou (1992) for example applied the concept to her ethnographic study of New York’s Chinatown finding that employment in this enclave had positive effects both for entrepreneurs, in the form of higher incomes and, for employees, in the form of opportunities for self-employment. While workers could initially receive low wages and worked longer hours, these disadvantages were compensated for by on-the-job training and social ties, facilitating their acquisition of their own business after some years. These advantages applied only to male workers, however,

as women remained confined to the role of supporting the entrepreneurial initiatives of their husbands or other male kin (Zhou & Logan, 1989; Zhou, 1992).

These gender differences in access and economic returns to self-employment were replicated in a subsequent study of the original Cuban enclave in Miami (Portes & Jensen, 1989). On the basis of a different data set and drawing largely on the economic experiences of immigrant minorities in New York, Bailey and Waldinger (1991) defined the enclave labor market as a “training system” where the example and skills of successful entrepreneurs and ethnic bonds of solidarity facilitated the emergence of new firms which produced, in turn, new employment opportunities. In his studies of the needle trade and of various New York public bureaucracies, Waldinger (1986, 1996) showed how access to information made available through social networks led to the construction of “niches” dominated by members of a particular immigrant or ethnic group.

Taking a different tack, Logan, Alba, and McNulty (1994) sought to measure the range and depth of ethnic economies created by different immigrant groups on the basis of census data. For this purpose, they classified all private sector workers in each target metropolitan area into self-employed vs. employed by others, and by industry sector, recombining two-digit industry codes into 47 categories. “Enclave economies” included those sectors where both self-employed and employees from a particular ethnic or immigrant group were overrepresented. “Entrepreneurial niches” comprised those sectors where the self-employed, but not workers from a given group were overrepresented.

On the basis of these criteria, Logan et al. were able to “map” the ethnic economies of different groups in metropolitan areas nationwide. In agreement with earlier descriptions by Portes and Bach, they found that well-developed enclave economies were exceptional, being associated with only a few immigrant groups, such as the Korean and Chinese. None, however, came close to the original enclave in Miami in terms of number of sectors and strength of the immigrant presence in those sectors. Unlike most ethnic groups in other cities, which were overrepresented in only a few sectors of the economy, the entrepreneurial and wage labor presence of Cubans in the Miami economy was widespread. These findings replicated those of Wilson and Martin (1982) a decade earlier who found not only great diversity in the internal composition of the Cuban enclave, but also dense networks among these firms so that the “outputs” of some were utilized as “inputs” for others. Such tight social and economic networks gave enclave firms a

distinct advantage, allowing them to compete, despite their smaller size, with those in the mainstream economy.

Arrayed against these generally supportive findings, a number of studies emerged that questioned the benefits of ethnic entrepreneurship and even the very existence of entrepreneurial enclaves. Several economists have taken a dim view of ethnic entrepreneurship and self-employment in general. Borjas (1990), for example, flatly asserts that, "There is no evidence that immigrant entrepreneurs are particularly successful. The presumption that many immigrant entrepreneurs begin with small shops and through their ability and hard work accumulate substantial wealth is a myth" (pp. 163–164). Similarly, Bates and Dunham (1991) dismiss the prevalence of self-employment among immigrant groups with the comment that "it may reflect in part the language difficulties that restrict employment alternatives" (p. 12). Bates (1993, p. A14) concludes that because of the "meager returns" to entrepreneurship only immigrants who are not fluent in English follow the entrepreneurial path.

These assertions were made despite consistent evidence that immigrant and native entrepreneurs receive higher earnings than their co-ethnic wage earners. Borjas, for example, sees as a myth that entrepreneurship helps immigrants despite his own evidence that self-employed immigrants earn about 48 percent more than their waged counterparts. He discounted this difference because "the income advantage of self-employed over salaried workers disappears after controlling for the large differences in demographic characteristics between the two groups" (Borjas, 1990, p. 164–65). Similarly, Bates (1989) and Bates and Dunham (1991) argue that the earnings advantage of immigrant entrepreneurs is a result of their greater human capital. Arguing that self-employment does not "pay" after rendering entrepreneurs and wageworkers artificially equivalent is questionable. More telling is that immigrants with higher education and other skills commonly choose the entrepreneurial route as a way of making those skills pay off. That the better qualified often choose entrepreneurship runs contrary to assertions that this path is a mobility trap.

Portes and Zhou (1996, 1999) have further demonstrated that self-employment has significant net positive effects on absolute earnings among both native and immigrant groups, even after controlling for human capital and demographic characteristics. Borjas et al. reached their conclusions by relying on the log linear functional form of the earnings equation. This function achieves a better fit to the assumption of normally distributed residuals, but only at the cost of obscuring substantively important

information. Positive outliers (i.e. those with very high earnings) who are “pulled” toward the mean of the distribution by the log linear form are disproportionately entrepreneurs. Put differently, among individuals who have achieved extraordinary economic success, the self-employed are heavily represented. This is pertinent information to the question of whether this route to economic adaptation pays or not, but it disappears from view by the statistical choice of regressing only logged dollars.

Sociological critics of the enclave hypothesis have questioned less the positive effects of self-employment on earnings than the effect that enclave employment has for workers. Sanders and Nee (1987), who took issue with Wilson and Portes’ (1980) positive view of enclave employment, recognized nevertheless that “self-employment in an enclave economy appears to be well rewarded for a set of human capital characteristics” (p. 763). In assessing these and other critics, it should be noted that the original formulation of the hypothesis never asserted that enclave employees would have higher earnings, on average, than their counterparts in the mainstream economy. Instead, the point was that enclaves were entrepreneurial incubators, showing wage workers the ropes to become self-employed themselves and that this shift, in turn, led to improved economic circumstances (Portes & Bach, 1985; Bailey & Waldinger, 1991).

A common problem in attempts to measure effects of enclave employment on economic outcomes is the difficulty of operationalizing the concept. Wilson and Portes (1980) and Portes and Bach (1985) defined the enclave as firms located in geographically circumscribed areas with high concentrations of other co-ethnic enterprises. Owners of these firms were defined as enclave entrepreneurs and workers in them as enclave workers. The original formulations of the hypothesis were grounded on survey data that unambiguously classified employed adults into these categories. In the absence of the necessary information in census data, investigators made use of various approximations of different degrees of plausibility.

As mentioned previously, Logan et al. used overrepresentation in collapsed two-digit census industrial categories to identify enclave sectors in specific metropolitan areas. While reasonable, this approach neglected the possibility that self-employed and waged workers in overrepresented sectors were not really part of the ethnic enclave and that those in underrepresented sectors were. There is considerable slippage in this approximation, making any conclusion about individual economic outcomes based on it rather tenuous (Logan et al., 1994).

More problematic still was the approach taken by Sanders and Nee (1987), who equated enclave participation with *living* in the area of

entrepreneurial co-ethnic concentration. Residents in these areas may or may not work in enclave firms and the latter may employ or be owned by individuals who do not live in close proximity to them. [Portes and Jensen \(1989\)](#) demonstrated that enclave entrepreneurs generally do not live in areas where their firms are located, but in better-off suburbs. Conversely, areas of high ethnic business concentration tend to be inhabited by low-income workers, whether enclave employees or not. It is this residential difference that [Nee and Sanders](#) captured in their analysis, erroneously attributing them to the deleterious economic effects of enclave work.

Another attempt to test the hypothesis defined “enclave” as all Hispanic-owned businesses in New York City and went on to report that immigrant women working in these firms experienced lower wages and poorer economic conditions than those employed elsewhere ([Gilbertson, 1995](#)). Obviously, there is no such thing as a “Hispanic” enclave in New York City since there is no immigrant nationality that goes by that name. There are concentrations of Dominican-owned firms in the Washington Heights area of upper Manhattan and of Colombian-owned firms in Queens ([Guarnizo, 1994](#); [Guarnizo, Sanchez, & Roach, 1999](#); [Itzigsohn, Dore, Fernandez, & Vazquez, 1999](#)) and these concentrations approach, in some dimensions, the characteristics of an enclave. However, this study did not focus on those areas, nor explained the dynamics of inter-ethnic relationships in them in a manner remotely resembling the fine-grained study of New York’s Chinatown by [Zhou \(1992\)](#).

All that could be said on the basis of [Gilbertson’s](#) findings is that firms that are the property of Latin immigrants or their descendants tend to pay lower wages to their female employees, a result that in no way bears on the predictions of the theory. A similar conclusion applies to [Hum’s \(2000\)](#) attempt to test the enclave hypothesis on the basis of data from the Los Angeles Study of Urban Inequality (LASUI) ([Bobo, Oliver, Johnson, & Valenzuela, 2000](#)). [Hum](#) correctly criticized prior attempts to operationalize the “ethnic economy” on the basis of place of residence or industrial sector. Instead, this author assigned workers to the “primary”, “secondary”, and “ethnic” labor market on the basis of job characteristics. “Ethnic economy” workers were those who had a co-ethnic supervisor and whose co-workers were mostly of the same ethnicity in firms no larger than 100 workers. Four nationalities were included: Chinese, Korean, Mexican, and Central American. On the basis of this operationalization, [Hum](#) conducted a series of multivariate analyses leading to the conclusion that “...mobility opportunities are limited in the ethnic economy. Contrary to its projection as facilitating the assimilation and mobility of new immigrants, the immigrant



ethnic economy does not typically provide a meaningful alternative to exploitative dead-end work” (Hum, 2000, p. 305).

Despite its use of a superior data set, this study exemplifies well the reigning confusion in this area. The study draws extensively from immigrant enclave theory and derives all its predictions from it. Then it proceeds to operationalize them, on the basis of a different concept, the “ethnic economy”, which bears little relationship to the original definition of the enclave. Mexican workers laboring next to other Mexicans and with a Mexican supervisor may be part of an “ethnic economy”, but they are certainly not part of an enclave. To be so, the firm had to be owned by Mexicans and located in an area of geographic concentration of such firms. The study provides no clue as to firm ownership or location. In the Los Angeles Metropolitan Area, there is no evidence that either Mexican or Central American immigrants have created enclaves making results about their respective economies irrelevant to the theory.

On the other hand, Koreans do possess a well-defined enclave in the area, as described by numerous past studies (Light & Bonacich, 1988; Light & Gold, 2000; Zhou, 2004). Despite Hum’s idiosyncratic measures, it is among this group where the “ethnic economy” does pay off. Thus, “the experience of Korean immigrants...emphasizes the centrality of self-employment and co-ethnic work relationship...Clearly, the quality of Korean ethnic economy employment stands out relative to the Chinese, Mexican, and Central American economies” (p. 305).

This conclusion is questionable, as well, with respect to the Chinese who also possess a well-defined enclave in suburban Monterrey Park (Fong, 1994). Its location probably escaped the LASUI sample, which also included a small number of Chinese workers ( $N = 154$ ). We have gone into such detail because this study represents the most recent and clearest example of the mischaracterization of the concept of ethnic enclave. The study draws its ideas and predictions from it, operationalizes them in ways bearing little relationship to the original definitions, and on the basis of predictable negative results concludes that the original hypotheses were wrong. The concept of enclave is quite distinct from that of “ethnic economy”. According to authors working on the latter tradition (Light & Rosenstein, 1995; Light & Gold, 2000), all groups possess an ethnic economy of some sort. On the contrary, as the original theory and subsequent analyses have noted, enclaves are exceptional and have emerged only among a few immigrant minorities (Portes & Bach, 1985; Portes & Jensen, 1989; Zhou, 2004).

*Revisiting the Cuban Enclave in Miami*

The enclave hypothesis was grounded on the remarkable economic experience of Cuban exiles in Miami in the 1960s and 1970s. Many escaped the communist revolution on the island with literally their “shirts on their backs”; few had property or capital in the United States; and practically none planned on permanent resettlement here, as they hoped for a prompt return to their country. The Bay of Pigs defeat of 1961 and the Missile Crisis episode the following year, where the Kennedy Administration traded the removal of Russian missiles from Cuba for a promise to rein exile efforts to overthrow the Castro regime, put an end to these hopes (Portes & Stepick, 1993; Garcia, 1996).

Highly concentrated in Miami for historical and geographical reasons, Cubans proceeded to rebuild lives and families there. Since many of the early exiles were entrepreneurs and professionals, they mobilized their skills, along with their networks to launch fledgling enterprises. Through devices such as the “character loans”, which Cuban bank officers advanced to clients without collateral on the basis of their business reputation in the island, many exile firms got a start (Portes & Stepick, 1993, pp. 132–135). Initially, these firms located in areas of co-ethnic concentration in the “Little Havana” section of Miami and in the city of Hialeah, but eventually they expanded throughout the entire metropolitan area (Perez, 1992; Stepick, Grenier, Castro, & Dunn, 2003).

Cuban-owned firms grew nationwide from an estimated 919 in 1967 to about 36,000 in 1982 and 61,500 in 1987. Most of these were found in the Miami/Ft. Lauderdale metropolitan area. Aggregate receipts of Hispanic firms in Miami were \$3.8 billion in 1987, a figure that exceeded by \$400 million that of second-ranking Los Angeles and was three times that of New York despite these cities having much larger Hispanic populations (Portes & Stepick, 1993, pp. 146). By 2000, the rate of self-employment per thousand employed persons nationwide was 93.5, while for Cubans it reached 127.3 (U.S. Bureau of the Census, 2000). By 2001, there were 125,273 Cuban-owned firms in the United States, with receipts of over 26.4 billion dollars. The majority of these concentrated in Miami/Ft. Lauderdale. There were 144 firms per 100,000 Cuban-origin population, a rate that quadrupled the figure among Mexicans, the largest Hispanic group, sextupled that among African-Americans, the largest domestic minority, and exceeded by a significant margin the rate among the native-born population as a whole (U.S. Bureau of the Census, 2002).

The remarkable story of the Cuban economic and, subsequently, political rise in South Florida has been told many times and with very different hues before. The concept of the enclave, born out of this experience, was then applied to other immigrant entrepreneurial concentrations, such as Chinatown in New York, Koreatown in Los Angeles, and Little Saigon in Orange County (Rumbaut, 1990, 1994; Zhou, 1992; Zhou, 2004); as well as large working-class immigrant communities such as Mexicans in the Pilsen area of Chicago (Rajiman & Tienda, 1999). As seen in the prior section, numerous attempts to test this hypothesis stretched the original definition of enclave beyond recognition. It seems, therefore, appropriate to return to Miami to see what happened to the group that created the phenomenon in the first place and to its descendants. If the hypothesis that concentrated entrepreneurship provides a viable form of economic adaptation for immigrant and ethnic minorities holds, it is among Cubans in Miami where these outcomes should be most prominent and visible.

Alternatively, if ethnic entrepreneurship and ethnic business concentrations are “mobility traps” (Bates, 1989), then the economic profile of a group that decisively opted for this route should show these results most clearly. The 5 percent micro-data sample (IPUMS) from the 2000 Census provides appropriate data to put these competing ideas to a test. Not only does the sample contain a sufficient number of cases and information on all relevant variables, but it comes from a census taken two decades after the original hypothesis was formulated. This allows us to examine the long-term evolution of this immigrant community and the results of its chosen path of economic adaptation.

Credible attempts to apply the concept of enclave to other immigrant/ethnic groups have always been grounded on historical knowledge and direct observations of the dynamic of these communities (Zhou, 1992; Guarnizo, 1994; Rajiman & Tienda, 1999). On the contrary, misapplications of the concept and faulty conclusions have generally been based on secondary analyses of census or survey data without any close familiarity with the groups studied, their histories, and modes of incorporation. For this reason, an analysis of 2000 Census data for Cubans in Miami must take into account the evolution of this community in the *interim* period, that is from 1980 to 2000.

The defining event for Cubans during this period was the Mariel exodus of 1980. The decision of the Cuban government to open the port of Mariel to all exiles wishing to take their relatives out of the island triggered a massive and chaotic new wave that brought to South Florida 125,000 refugees in less than six months. The Cuban government took advantage of the

episode to empty its jails and mental hospitals, putting the inmates aboard the boats. The spectacle of a chaotic flotilla and of boats of ragged people arriving in the Florida keys everyday triggered a strong negative reaction in the United States. The Carter Administration refused to grant the new arrivals refugee status, categorizing them as “entrants, status pending”. The public reaction to the Mariel episode shifted the perception of Cubans from a “model minority” and “the builders of the new Miami” to one of the most unpopular foreign minorities in the nation. Aware of this shift, the older exile community distanced itself from the new arrivals, coining the derogatory term *marielitos* to refer to them (Camayd-Freixas, 1988; Perez, 1992; Portes & Stepick, 1993).

Mariel marked a pivotal shift in modes of incorporation for Cubans arriving in U.S. shores. Prior to 1980, exiles which included almost the entire pre-revolutionary elite, had been warmly received by the federal government, were eligible for extensive resettlement assistance, and were perceived by the general public as a deserving group, loyal U.S. allies in the global struggle against communism. During Mariel and after, all of this changed. Cubans became just another Third World impoverished minority seeking to crash the doors of the nation; the federal government focused its efforts on stopping the inflow and treated new arrivals with much less benevolence than their predecessors. Following an accord with the Cuban government in 1994, all new refugees caught by the U.S. Coast Guard at sea were promptly returned to Cuba (Perez, 2001; Stepick et al., 2003).

More important still, the old middle-class Cuban population of Miami largely severed their ties with the new arrivals. They were not part of old Cuba, having been raised during the revolutionary period, and lacked strong kinship and friendship ties with the established Miami Cuban community. Perceiving the *marielitos* and post-Mariel entrants as responsible for the rapid decline of Cubans’ public image in the United States and having few social links with them, pre-1980 exiles came to regard the newcomers as a group different from themselves. The separation was physical, as well as social – the old middle-class Cuban population settled in the comfortable suburbs of Coral Gables and Kendall; Mariel and post-Mariel refugees crowded in the poor city of Hialeah and the deteriorating “Little Havana” quarter of Miami (Portes & Stepick, 1993; Garcia, 1996).

This rupture in modes of incorporation means that Mariel and post-Mariel refugees benefited little from the pre-1980 enclave and its internal ties of solidarity and mutual business support. No “character loans” were available to prospective new entrepreneurs for they were not known to Cuban bank officials. No tips about new business opportunities were

available to them. While many eventually went to work for Cuban-owned firms in Miami, and some eventually learned the ropes of self-employment, their links to business owners and to the established Cuban community at large were much weaker (Portes & Jensen, 1989; Stepick et al., 2003). These differences in the internal composition and evolution of the Cuban community must be taken into account when analyzing its economic performance by the end of the 20th century.

## RESULTS

### *Preliminary Findings*

The IPUMS files provide data for 66,955 adults, ages 18–65 who earned more than \$500 in the Miami/Ft. Lauderdale MSA. This large sample can be divided into ethnic categories of Non-Hispanic White, Non-Hispanic Black, Cuban, Other Hispanic, and Other. The last category is composed mainly of first- and second-generation Asians. The Cuban-origin sample is composed of 12,004 that can be further subdivided into Pre-1980 (pre-Mariel) arrivals, Mariel (1980) and later arrivals, and the U.S.-born second generation. Table 1 presents personal and family incomes for these ethnic categories.

Results show that non-Hispanic Whites ( $N = 25,387$ ) occupy the top of the economic hierarchy with incomes higher than any other group. All other groups have average personal incomes that are significantly lower, but, of these, only pre-1980 Cuban arrivals and the U.S.-born Cuban second generation have incomes that exceed the mean for the metropolitan area. In particular, only pre-1980 Cubans significantly exceed \$40,000 in average personal income, approaching the non-Hispanic White figure. The trend is the same when we consider family incomes. In this case, only non-Hispanic Whites and pre-Mariel Cubans exceed \$80,000 in annual income, with the Cuban second generation shy of that amount by a little more than \$1,000. The second generation is a group composed, almost exclusively, of persons born before 1980 (age 20 or higher in 2000) and, hence, the offspring of pre-Mariel Cubans. By contrast refugees arriving during the Mariel exodus and after have not succeeded economically. Their personal and family incomes are even lower than other Hispanic immigrants in the area and are not statistically different from Non-Hispanic Blacks, the bottom ethnic category.

**Table 1.** Personal and Family Incomes of Racial and Ethnic Groups in Miami/Ft. Lauderdale Metropolitan Area (Adults 18–65).

Racial/Ethnic Group	Total MSA	Non-Hispanic White	Non-Hispanic Black <sup>a</sup>	Cuban			Other Hispanic	Other
				Pre-1980 Migrant	1980 or After Migrant	U.S. Born		
Personal Income	37,407 (47,396)	49,812 (60,297)	26,642*** (24,931)	45,218*** (51,489)	23,961*** (25,754)	34,971** (40,169)	27,671*** (34,289)	32,724*** (41,831)
N	66,955	25,383	12,003	4,914	5,021	2,069	14,546	3,019

Racial/Ethnic Group	Total MSA	Non-Hispanic White	Non-Hispanic Black	Cuban			Other Hispanic	Other
				Pre-1980 Migrant	1980 or After Migrant	U.S. Born		
Family Income	68,720 (67,753)	84,842 (82,165)	51,361*** (41,434)	82,589 (74,208)	51,071*** (43,838)	78,739*** (62,310)	56,371*** (54,670)	59,674*** (59,775)
N	66,955	25,383	12,003	4,914	5,021	2,069	14,546	3,019

Source: IPUMS 2000.

Note: Universe includes adults aged 18–64, who are not unemployed, and whose annual income is greater than or equal to \$500. Raw *N* values included; person weights used. Standard deviations in parenthesis.

<sup>a</sup>Non-Hispanic White is the reference category. Significant differences from this category are noted by asterisks. \*\*\* $p < 0.001$ ; \*\* $p < 0.01$ ; \* $p < 0.05$ , two-tailed.

Table 2 disaggregates these figures further into the categories of self-employed and wage earners among adult males (there are not sufficient numbers of self-employed females to permit inter-ethnic comparisons). Self-employed Cubans in Miami are, almost by definition, ethnic enclave entrepreneurs since, as explained previously the original business networks of the enclave grew out of the city of Miami to encompass the entire metropolitan area. This is especially the case among pre-1980 exiles – founders of the original enclave and their offspring. By contrast, Cuban waged and salaried workers may or may not be employees of the enclave economy. While the data provide no means to establish this comparison, they allow us at least to examine the economic situation of Cuban workers who arrived at the time the business enclave was being built and those who came during the Mariel exodus and later.

The same ethnic hierarchy observed earlier holds, with Non-Hispanic Whites at the top, pre-Mariel Cubans close behind, followed by their children and everyone else below. As repeatedly noted in prior studies, entrepreneurs (the self-employed) enjoy a substantial economic advantage in terms of personal and, especially, family incomes. Non-Hispanic Whites and pre-1980 Cuban entrepreneurs are the only groups to exceed an annual family income of \$100,000, with the difference between the two groups just shy of \$1,000. They are followed, at some distance, by second generation Cubans, and then all others. Differences among waged and salaried workers follow exactly the same pattern: There is no statistical difference between the average incomes of Non-Hispanic White and Pre-Mariel Cuban workers, while all other groups fall significantly behind. This result indicates that those Cuban workers who arrived at the time that the ethnic enclave developed have done rather well economically, whether they are employed by these firms or not.

The difference is stark with the economic performance of Mariel and post-Mariel refugees whose income levels are at the bottom of the ladder for both the self-employed and the employees and in personal as well as family incomes. Noteworthy in particular is the dismal performance of entrepreneurs from this group whose economic rewards are the lowest of all ethnic categories, being surpassed by other Hispanics and African-Americans. It is evident from these results that the original advantages conferred on would-be entrepreneurs by networks within the Cuban enclave did not extend to the later arrivals. They are, as it were, a group apart. Part of this disadvantage may be due to their recency of arrival, a possibility that will be examined in the next section. However, second generation Cubans entered the local labor force at about the same time, i.e., they were also “recent”

**Table 2.** Personal and Family Incomes of Working, Self-Employed, and Wage/Salaried Males by Racial and Ethnic Group in Miami/Ft. Lauderdale Metropolitan Area.

Racial/Ethnic Group	Total MSA	Non-Hispanic White	Non-Hispanic Black <sup>a</sup>	Cuban			Other Hispanic	Other
				Pre-1980 Migrant	1980 or After Migrant	U.S. Born <sup>a</sup>		
(a) Personal Income								
Working Adults	45,010 (57,377)	61,712 (72,504)	28,929*** (27,642)	56,541*** (63,132)	26,918*** (27,670)	39,265*** (46,748)	32,414*** (40,756)	37,383*** (48,084)
Self-Employed	64,968 (86,719)	80,948 (97,159)	39,255*** (57,350)	71,302 (87,575)	28,250*** (34,865)	65,594 (84,173)	49,033*** (71,487)	57,369 (91,733)
Wage/Salaried	41,639 (49,995)	57,474 (65,088)	28,244*** (24,281)	52,523** (53,955)	26,647*** (25,963)	35,983*** (38,525)	30,254*** (34,242)	34,429*** (36,706)
(b) Family income								
Working Adults	70,500 (70,057)	88,226 (85,302)	53,078*** (42,213)	87,404*** (78,036)	50,109*** (42,850)	77,688*** (63,267)	55,409*** (52,973)	58,974*** (60,620)
Self-Employed	90,618 (100,842)	106,667 (110,504)	60,443*** (64,862)	105,921 (108,997)	53,545*** (52,097)	94,683 (92,853)	71,671*** (85,701)	79,990** (111,365)
Wage/Salaried	67,102 (62,767)	84,162 (78,094)	52,589*** (40,224)	82,363 (66,334)	49,410*** (40,689)	75,569** (58,267)	53,294*** (46,670)	55,868*** (48,138)

Source: IPUMS 2000.

Note: Universe includes adult males aged 18–64, who are not unemployed, and whose annual income is greater than or equal to \$500. Person-weights used; standard deviations in parenthesis.

<sup>a</sup>Non-Hispanic White is the reference category. Significant differences from this category are noted by asterisks. \*\*\*  $p < 0.001$ ; \*\*  $p < 0.01$ ; \*  $p < 0.05$ , two-tailed.



workers with little work experience. Yet, their economic achievements as both new entrepreneurs and workers far exceed those of their Mariel compatriots.

This is not because of lack of trying among recent arrivals. Table 3 presents self-employment rates for all adult males in the Miami metropolitan area by ethnic categories in 2000. Two trends are apparent in these results: First, pre-Mariel Cubans are the most entrepreneurial group, with a rate significantly higher than non-Hispanic Whites. This result accords with the historical role of this group as builders of the enclave economy. Second, Mariel and post-Mariel Cubans have also been strongly inclined to follow that route with self-employment rates not significantly lower than non-Hispanic Whites. However, as just seen, their efforts have not paid off so far. This outcome is largely attributable to their exclusion from the resources in credit, business information, and opportunities that flowed from the original networks of the enclave.

Despite these disadvantages, the existence of the Cuban enclave seems to have benefited *workers* of the same nationality, regardless of the time of arrival. In one of the original formulations of the theory, Portes and Bach (1985) argued that the existence of a dense network of co-ethnic enterprises allow recent arrivals to put to use the human capital brought from the home country, even without knowledge of English. While in the mainstream economy, lack of English condemns immigrant workers to the most menial occupations; this is not true in enclave enterprises where the language of work is frequently that of the home country.

Table 4 presents the incomes of immigrant waged and salaried workers in the Miami/Ft. Lauderdale MSA who did not speak English well at the time of the 2000 Census. The four sub-tables present data on personal and family incomes, for the male and total populations. In all cases, results indicate that Cuban workers receive incomes significantly higher than other immigrants in the same situation. Since the data also indicate that immigrants without English do not differ significantly across nationalities in other dimensions of human capital, the results suggest that Cubans were better able to put to use whatever human capital they brought from their country. This result is in agreement with the original hypothesis. The existence of the Cuban enclave has given new arrivals and older workers without much education (those who most commonly do not know English) economic opportunities absent in the mainstream economy. Immigrants from other nationalities have not been so fortunate.

Finally, the general pattern of results observed in the Census data repeats itself in other recent surveys conducted in the area. Table 5 presents findings

**Table 3.** Self-employment Rates among Males by Racial and Ethnic Group in Miami/Ft. Lauderdale Metropolitan Area (Adults 18–65).

Racial/Ethnic Group	Total MSA	Non-Hispanic White	Non-Hispanic Black <sup>a</sup>	Cuban			Other Hispanic	Other
				Pre-1980 Migrant	1980 or After Migrant	U.S. Born		
Self-Employment	0.14	0.18	0.06***	0.21***	0.17	0.11***	0.12***	0.13***
N	35,285	13,780	5,430	2,398	3,077	1,051	7,872	1,677

Source: IPUMS 2000.

Note: Universe includes adult males aged 18–64, who are not unemployed, and whose annual income is greater than or equal to \$500. Raw N values included; person-weights used.

<sup>a</sup>Non-Hispanic White is the reference category. Significant differences from this category are noted by asterisks. \*\*\* $p < 0.001$ ; \*\* $p < 0.01$ ; \*  $p < 0.05$ , two-tailed.

**Table 4.** Personal and Family Incomes of Foreign-Born Males and Individuals (18–65) Who did not Speak English Well or at all in Miami/Ft. Lauderdale Metropolitan Area, 2000.

Racial/Ethnic Group	Total	Other Hispanic	Cuban <sup>a</sup>	Non-Hispanic Other
(a) Personal incomes of foreign-born males				
Working Adults	21,679 (24,212)	20,753 (25,349)	22,939** (23,166)	20,331 (18,595)
Wage/Salaried	20,639 (19,780)	19,455 (20,514)	22,365*** (18,200)	18,228 (12,439)
N (Working; Wage/Salaried)	3,907; 3,320	1,791; 1,584	1,673; 1,339	144 124
(b) Family incomes of foreign-born males				
Working Adults	41,943 (40,877)	39,458 (39,736)	45,772*** (44,042)	35,818 (25,329)
Wage/Salaried	40,635 (37,366)	38,160 (35,922)	44,596*** (40,617)	34,378 (22,480)
N (Working; Wage/Salaried)	3,907; 3,320	1,791; 1,584	1,673; 1,339	144; 124
(c) Personal incomes of foreign-born individuals				
Working Adults	18,687 (22,322)	17,568 (21,542)	20,080*** (23,443)	17,362 (15,781)
Wage/Salaried	17,895 (18,728)	16,937 (18,494)	19,111*** (18,494)	16,156 (11,530)
N (Working; Wage/Salaried)	6,922; 5,982	3,253; 2,821	2,806; 2,376	250; 222
(d) Family incomes of foreign-born individuals				
Working Adults	42,981 (43,602)	41,390 (45,018)	46,083*** (43,754)	37,300* (25,704)
Wage/Salaried	41,806 (40,280)	40,198 (42,177)	44,906*** (39,370)	35,975* (23,319)
N (Working; Wage/Salaried)	6,922; 5,982	3,253; 2,821	2,806; 2,376	250; 222

Source: IPUMS 2000.

Note: Universe includes adults aged 18–64, who are not unemployed and whose annual income is greater than or equal to \$500. Raw *N* values included; person-weights used. Standard deviations in parenthesis.

<sup>a</sup>Other Hispanic is the reference category. Significant differences from this category are noted by asterisks. \*\*\* $p < 0.001$ ; \*\* $p < 0.01$ ; \* $p < 0.05$ , two-tailed.

from a large survey of immigrant parents, conducted in 1996 in conjunction with the first follow-up wave of the Children of Immigrants Longitudinal Study (CILS) in South Florida. The original sample of second-generation 8th and 9th graders was representative of the relevant population in the

**Table 5.** Family Incomes of Immigrant Parents in Miami/Ft. Lauderdale Metropolitan Area.

	Total Sample	Nationality		Cuban	
		Cuban <sup>a</sup>	Non-Cuban	Self-employed <sup>b</sup>	Other
Family Income	40,218 (38,334)	47,517*** (42,245)	36,350 (35,518)	61,986** (58,774)	43,960 (36,334)
<i>N</i>	1,068	698	370	73	297

Source: CILS Parental Survey; Standard deviation in parenthesis.

<sup>a</sup>Non-Cuban is the reference category. Significant differences from this category are noted by asterisks. \*\*\* $p < 0.001$ ; \*\* $p < 0.01$ ; \* $p < 0.05$ , two-tailed.

<sup>b</sup>Non-self-employed is the reference category.

schools of the Miami/Ft. Lauderdale metropolitan area. The parental survey included a random sample of 50 percent of the original respondents, making it, in turn, representative of that universe. As seen in Table 5, Cuban parents had significantly higher family incomes on average than those of all other nationalities combined. In turn, enclave entrepreneurs in this sample (self-employed Cuban parents) had the highest annual earnings of any category, exceeding co-ethnic employees by almost \$20,000 and other immigrant parents by over \$25,000.

*Multivariate Analysis: Income Determinants*

The average inter-ethnic differences observed previously are suggestive, but are subject to the objection that they reflect differences in levels of human capital, including education and work experience. As seen previously, some economists make much of the fact that entrepreneurs tend to have higher levels of human capital, using this fact to conclude that the economic gain from self-employment is a “myth”. More reasonably, it could be argued that pre-Mariel Cubans represent an older and, hence, more experienced population and that this is the root cause of their superior economic performance.

In the following multivariate analysis, we use actual dollars throughout for reasons explained in detail in Portes and Zhou (1996). The log linear form of income regressions expresses the proportional net gain or loss associated with a unit change of each predictor, relative to the average of the distribution. As Hodson (1985, p. 387) has noted, the proportional net

effects produced by the log linear form can obscure real differences between categories of individuals in a comparative analysis. This occurs because the log linear form's relative effects are a partial function of each category's mean income.<sup>1</sup> Further, the log linear form achieves a better approximation to normality by neutralizing the influence of income outliers. For the purpose at hand, this means suppressing valuable information since entrepreneurs are disproportionately represented among positive income outliers, this result being indicative of their economic success.

For reasons of space, we restrict the presentation of results to family income, which provides a more accurate measure of overall family economic well-being. We conducted a parallel analysis of personal incomes and its results tend to reproduce, in all essential aspects, those to be discussed next. Table 6 presents regressions of family incomes in actual dollars for the adult working population of Miami/Ft. Lauderdale on indicators of human capital, self-employment, and ethnic origins. The first model presents the effects of all main ethnic categories, while the second disaggregates the Cuban-origin population into the sub-categories discussed previously.

Three main findings emerge from this analysis. First, education and work experience, indexed by age, have the expected strong net effects on incomes, as does gender. College and post-college graduates derive an advantage to their education measured in the tens of thousands of dollars relative to high-school dropouts (the reference category). Women suffer the well-demonstrated income handicap relative to statistically equivalent males. Second, after controlling for these predictors, self-employment continues to have a positive and significant effect. Compared to workers of the same education, work experience, gender, and ethnicity, the self-employed have a net annual income advantage of \$14,000 in this sample.

Third, all ethnic categories earn significantly less than Non-Hispanic Whites, even after controlling for human capital, gender, and self-employment. These differences are statistically significant in the first model, but disappear when the Cuban-origin sample is disaggregated. Pre-Mariel Cubans and the U.S.-born Cuban second generation now exhibit incomes that are higher than comparable Non-Hispanic Whites, although the differences are not statistically significant. In contrast, Mariel and post-Mariel Cubans continue to experience an income disadvantage that is higher than that suffered by any other ethnic category, including African-Americans. The annual income gap relative to statistically comparable native whites is a startling \$23,000 for this segment of the Cuban population.

With human capital indicators and gender controlled, the differences show clearly the advantages of entrepreneurship in general and of

**Table 6.** Regressions of Family Income on Ethnicity and Selected Variables (Adults 18–64).

	I	II
<i>Ethnicity:</i>		
Cuban <sup>a</sup>	-9511.28*** (796.40)	
Pre-1980 Cuban		212.53 (1206.35)
1980 or After Cuban		-23552.77*** (876.18)
U.S.-born Cuban		114.91 (1580.00)
Black	-21608.99*** (642.99)	-21912.73*** (641.92)
Hispanic	-19352.02*** (715.17)	-19680.03*** (714.77)
Other	-19793.26*** (1271.88)	-20046.82*** (1271.86)
Female	-1925.97*** (529.81)	-2342.27*** (529.07)
Age	392.40** (147.95)	442.76** (148.60)
Age <sup>2</sup>	-0.93 (1.86)	-1.79 (1.86)
<i>Education:</i>		
High School	7031.37*** (698.97)	6269.24*** (697.53)
Some College	17832.03*** (721.40)	16139.94*** (723.35)
College	37975.68*** (981.78)	36265.05*** (980.72)
Post-Graduate	60470.47*** (1479.38)	58872.11*** (1471.76)
Self-Employed	14167.01*** (1198.81)	14133.69*** (1194.73)
Intercept	44109.86*** (2882.71)	45180.10*** (2898.06)
<i>N</i>	66,955	66,955
<i>R</i> <sup>2</sup>	0.12	0.12

Source: IPUMS 2000 (5% microsample). Standard deviations in parentheses.

Note: Universe includes adults aged 18–64, who are not unemployed and whose annual income is greater than or equal to \$500. Raw *N* values included; person-weights used. \*\*\**p*<0.001; \*\**p*<0.01; \**p*<0.05, two-tailed.

<sup>a</sup>Non-Hispanic White is the reference category.

involvement in the enclave economy, in particular. It is the cohort associated with its creation in the 1960s and 1970s and its descendants that have been most able to reap its benefits, matching or exceeding the economic performance of native whites. Mariel and post-Mariel Cubans have been largely excluded from the resources of the enclave, and, in consequence, have been relegated to the economic level of other disadvantaged racial and ethnic minorities.

A possible objection to these findings is that age is not a good indicator of U.S. work experience. Mariel and post-Mariel entrants of the same age as pre-1980 Cubans may have much fewer years of experience in the American labor market because of their recency of arrival. In order to take into account this possibility, we substituted U.S. work experience for age across all ethnic groups. For the native-born, work experience is computed in standard fashion as age minus education minus six. For adult immigrants, it is years since U.S. arrival minus  $y$ , where  $y$  is 0 for those who arrived after age 18 and  $(18 - \text{age at arrival})$  for those who arrived as minors. Table 7 presents the result for the adult universe and for males only.

With work experience brought into the equation, results are still more powerful than before. The Cuban second generation remains statistically indistinguishable from Non-Hispanic Whites, although they receive slightly lower average incomes. On the other hand, pre-Mariel Cubans males now exceed the incomes of the reference category by a significant margin. Net of work experience and other control variables, this group of earlier exiles displays a net income advantage over native white males of almost \$4,000.

A second pertinent question is whether income determinants differ among ethnic groups and, in particular, whether the overall positive effect of self-employment on annual incomes varies significantly among them. The enclave hypothesis predicts that it does and that the effect is more marked among the more entrepreneurially inclined groups, especially among those in dense co-ethnic concentrations (Wilson & Martin, 1982; Zhou, 2004). Entrepreneurship should “pay” more in this situation because of access to otherwise unavailable or more expensive resources, such as in-site business apprenticeship, subsidized credit, captive markets, and disciplined labor (Bailey & Waldinger, 1991; Raijman & Tienda, 1999). Table 8 presents relevant results for the total adult working population of Miami/Ft. Lauderdale with Cubans as a single ethnic category. Significant differences from Non-Hispanic Whites are indicated by the dagger sign (†).

There are indeed numerous differences among all ethnic groups relative to native whites, indicating that the income attainment process varies

**Table 7.** Regression of Family Income on Ethnicity and Selected Variables Substituting Work Experience for Age.

	Adults (18–64)	Males (18–64)
<i>Ethnicity:</i> <sup>a</sup>		
Pre-1980 Cuban	2023.88 (1220.47)	3849.55* (1777.46)
1980 or After Cuban	–14956.88*** (1027.94)	–16442.62*** (1381.96)
U.S.-born Cuban	–222.71 (1574.38)	–372.66 (2272.18)
Black	–19062.42*** (676.70)	–17100.27*** (974.96)
Hispanic	–14163.78*** (813.21)	–15565.28*** (1115.40)
Other	–14091.57*** (1345.03)	–15856.60*** (1834.92)
Female	–2408.99*** (528.30)	
Work Experience (U.S.) <sup>b</sup>	940.46*** (58.77)	934.68*** (100.78)
Work Experience (U.S.)	–15.81*** (1.89)	–13.00*** (2.65)
<i>Education:</i> <sup>c</sup>		
High School	5787.45*** (697.25)	5109.78*** (893.68)
Some College	15452.91*** (717.47)	14750.29*** (938.03)
College	36585.79*** (977.00)	37062.47*** (1295.85)
Post-Graduate	60334.30*** (1469.41)	66736.50*** (2010.92)
Self-Employed <sup>d</sup>	14869.66*** (1189.95)	15281.24*** (1451.51)
Intercept	50467.92*** (1017.30)	49398.98*** (1371.06)
<i>N</i>	66,955	35,285
<i>R</i> <sup>2</sup>	0.13	0.15

Source: IPUMS 2000 (5% microsample). Standard deviations in parentheses.

Note: Universe includes adults aged 18–64, who are not unemployed, and whose annual income is greater than or equal to \$500. Raw *N* values included; person weight used. \*\*\* $p < 0.001$ ; \*\* $p < 0.01$ ; \* $p < 0.05$ , two-tailed.

<sup>a</sup>Non-Hispanic White is the reference category.

<sup>b</sup>See text for definition of this variable.

<sup>c</sup>Less than high school is the reference category.

<sup>d</sup>Wage/salaried worker is the reference category.



**Table 8.** Within-Group Regression of Family Income on Ethnicity and Selected Variables (Adults 18–64).

	Non-Hispanic White	Black	Cuban	Hispanic	Other
Female	–4117.59*** (1049.19)	–5008.40*** (796.64)	–1054.42 <sup>†</sup> (1200.25)	1855.88 <sup>†</sup> (967.26)	2623.37 <sup>†</sup> (2256.60)
Work Experience (U.S.)	1319.20*** (138.60)	214.09* <sup>†</sup> (88.74)	2815.86*** <sup>†</sup> (195.65)	805.76*** <sup>†</sup> (156.96)	1199.97** (392.69)
Work Experience (U.S.)	–22.9*** (3.22)	1.61 <sup>†</sup> (2.38)	–72.42*** <sup>†</sup> (8.30)	–10.91*** (6.00)	–28.14* (10.93)
<i>Education:</i> <sup>a</sup>					
High School	4052.18 (2252.73)	5926.08*** (1212.17)	10176.28*** <sup>†</sup> (1387.11)	6808.05*** (1204.49)	9475.05*** (2256.53)
Some College	13502.13*** (2220.29)	15881.94*** (1284.21)	25641.71*** <sup>†</sup> (1591.06)	14422.51*** (1148.32)	16554.00*** (2262.79)
College	40040.90*** (2419.75)	30141.33*** <sup>†</sup> (1791.54)	42045.92*** (2173.00)	32784.91*** <sup>†</sup> (1833.56)	34114.95*** (3185.11)
Post-Graduate	68345.77*** (2908.90)	45497.28*** <sup>†</sup> (2801.10)	61501.53*** (3308.8)	49010.75*** <sup>†</sup> (3101.34)	67574.67*** (7093.77)
Self-Employed <sup>b</sup>	16962.83*** (1971.13)	4584.36 <sup>†</sup> (2736.14)	14778.79*** (2529.12)	12206.26*** (2190.85)	18345.39** (6228.57)
Intercept	46554.31*** (2427.05)	38002.81*** <sup>†</sup> (1206.58)	31336.33*** <sup>†</sup> (1416.20)	36313.28*** <sup>†</sup> (1014.81)	32046.61*** <sup>†</sup> (2221.40)
<i>N</i>	25,383	12,003	12,004	14,546	3,019
<i>R</i> <sup>2</sup>	0.09	0.08	0.12	0.07	0.12

*Source:* IPUMS 2000 (5% microsample). Standard deviations are in parentheses.

*Note:* Universe includes adults aged 18–64, who are not unemployed, and whose annual income is greater than or equal to \$500. Raw *N* values included; person weight used.

<sup>a</sup>Less than high school is the reference category.

<sup>b</sup>Wage/salaried individuals is the reference category. \*\*\* $p < 0.001$ ; \*\* $p < 0.01$ ; \* $p < 0.05$ ; two-tailed.

<sup>†</sup> = significantly different from Non-Hispanic Whites.

significantly between them. There are several suggestive results in these models, such as the linear effect of work experience – stronger for Cubans than for any other group – and the net effect of gender (female), which turns positive among Other Hispanics and Others (mostly Asians), after controlling for other predictors. The key results of interest are the coefficients associated with self-employment, which again vary widely across groups. The net effect is always positive, but it is too small to be significant among African-Americans. In agreement with the enclave hypothesis, the effect is strongest among the three most entrepreneurial groups – Non-Hispanic Whites, Others (Asians), and Cubans – and there are no statistically significant differences among them. Cubans, however, derive a smaller absolute benefit from entrepreneurship, a result which runs contrary to expectations since they are the only group associated with a full-fledged enclave in the area. Results limited to adult males (not shown) reproduce this pattern.

Results in Table 9 clarify this anomaly: Pre-Mariel Cubans receive the highest payoff for entrepreneurship of any group, exceeding by over \$6,000 the comparable figure among Non-Hispanic Whites. In agreement with prior results, the gain associated with self-employment for Mariel and post-Mariel Cubans, while still significant, is less than one-fourth of the figure among their pre-1980 compatriots. A final result of note is that the net effect of self-employment is statistically insignificant among the Cuban second generation. This indicates that economic advantages observed previously for this group *do not depend* on its being highly entrepreneurial. Offspring of successful immigrant businessmen do not necessarily follow in the footsteps of their parents. They may use instead resources accumulated by the first generation for an advanced education and for entry into well-paying professional careers.

This pattern accords with the experience of earlier immigrant enclaves chronicled in the literature, such as those created by Russian Jews on the Lower East Side of Manhattan and by the Japanese in Los Angeles and San Francisco (Rischin, 1962; Petersen, 1971; Howe, 1976). As in these earlier formations, enclaves do not seem to be self-perpetuating phenomena, but “platforms” for successful entry into the social and economic mainstream by the children and grandchildren of the original entrepreneurs (Portes & Bach, 1985).

#### *Multivariate Results: Determinants of Self-Employment*

Given the consistent positive economic effect of entrepreneurship for the entire sample and for most ethnic groups, it is worthwhile to examine what

**Table 9.** Within-Group Regression of Family Income on Ethnicity and Selected Variables: Non-Hispanic Whites and Disaggregated Cuban Population.

	Non-Hispanic White	Pre-1980 Cuban	1980+ Cuban	U.S.-born Cuban
Female	-4117.59*** (1049.19)	-5024.23* (2160.30)	2101.20† (1424.20)	-1563.30 (3036.46)
Work Experience (U.S.)	1319.20*** (138.60)	2624.27***† (619.31)	1608.37* (687.42)	749.24 (518.04)
Work Experience (U.S.) <sup>2</sup>	-22.9*** (3.22)	-61.96***† (18.95)	-172.32***† (62.47)	-16.31 (16.29)
<i>Education:</i> <sup>a</sup>				
High School	4052.18 (2252.73)	16151.95***† (2904.91)	3537.81* (1672.30)	8916.98 (5391.65)
Some College	13502.13*** (2220.29)	31908.14***† (3167.08)	19179.55*** (2234.00)	22576.09*** (5268.87)
College	40040.90*** (2419.75)	58307.49***† (4108.60)	16689.70***† (2957.62)	42263.08*** (5888.30)
Post-Graduate	68345.77*** (2908.90)	86260.38***† (5801.42)	24439.22***† (3415.14)	62305.33*** (9110.61)
Self-Employed <sup>b</sup>	16962.83*** (1971.13)	23032.79*** (4509.87)	5900.75*† (2537.09)	11171.43 (7328.04)
Intercept	46554.31*** (2427.05)	-24690.68***† (5703.16)	39608.79***† (1734.45)	46889.03*** (5644.14)
<i>N</i>	25,383	4,914	5,021	2,069
<i>R</i> <sup>2</sup>	0.09	0.11	0.04	0.08

Source: IPUMS 2000 (5% microsample). Standard deviations in parentheses.

Note: Universe includes adults aged 18–64, who are not unemployed, and whose annual income is greater than or equal to \$500. Raw *N* values included; person weight used.

<sup>a</sup>Less than high school is the reference category.

<sup>b</sup>Wage/salaried individuals is the reference category.\*\*\* $p < 0.001$ ; \*\* $p < 0.01$ ; \* $p < 0.05$ , two-tailed.

† = significantly different than Non-Hispanic Whites.

leads to entry into this economic route. For this purpose, we modeled the probability of entrepreneurship (self-employment = 1) on the same set of regressors used previously in a series of logistic regressions. We examine first the main effects of ethnicity on self-employment and then the interaction effects across the various groups. Table 10 presents results in the form of net odd-ratios taken to the fourth significant digit. The table shows that the principal determinant of entrepreneurship is an advanced education. The odds of people with a post-graduate degree becoming self-employed are 1.36 to 1.00, relative to those with less than a high-school education. The effect of education on entrepreneurship is nonmonotonic since those with a

**Table 10.** Determinants of Self-Employment in Miami-Ft. Lauderdale, 2000 (Adults 18–64)<sup>a</sup>.

	I	II
<i>Ethnicity:</i> <sup>b</sup>		
Cuban	1.0975* (0.0425)	
Pre-1980 Cuban		1.2844*** (0.0621)
1980 or After Cuban		1.1072 (0.0646)
U.S.-born Cuban		0.6387*** (0.0572)
Black	0.3724*** (0.0190)	0.3716*** (0.0191)
Hispanic	0.9029** (0.0366)	0.8960** (0.0371)
Other	0.8422** (0.0589)	0.8352** (0.0588)
Female	0.4931*** (0.0139)	0.4925*** (0.0138)
Work Experience (U.S.)	1.0076* (0.0034)	1.0057 (0.0037)
Work Experience (U.S.)	1.0002** (0.0001)	1.0002** (0.0001)
<i>Education:</i> <sup>c</sup>		
High School	0.9136 (0.0432)	0.9199 (0.0436)
Some College	0.8010*** (0.0380)	0.8138*** (0.0387)
College	0.9001* (0.0467)	0.9131 (0.0475)
Post-Graduate	1.3611*** (0.0732)	1.3752*** (0.0741)
<i>N</i>	66,955	66,955
Pseudo <i>R</i> <sup>2</sup>	0.04	0.04
Wald $\chi^2$	1400.57	1465.33

*Note:* Universe includes adults aged 18–64, who are not unemployed, and whose annual income is greater than or equal to \$500. Raw *N* values included; person weight used.

*Source:* IPUMS 2000 (5% microsample).

<sup>a</sup>Logistic regression, net odds-ratios to the fourth digit. Standard deviations in parentheses.

<sup>b</sup>Non-Hispanic White is the reference category.

<sup>c</sup>Less than high school is the reference category. \*\*\**p* < 0.001; \*\**p* < 0.01; \**p* < 0.05, two-tailed.

high school degree or college are *less* likely to be self-employed than high-school dropouts. This suggests a bimodal pattern in which entrepreneurship is embraced by the most educated as a vehicle for economic advancement and by those with the least education as a possible vehicle for survival. We will examine whether this pattern is reproduced among all ethnic groups or whether it is exclusive to some of them. In agreement with past results in the literature, females are significantly less likely to be self-employed.

The ethnic coefficients reveal an interesting pattern of results: relative to Non-Hispanic Whites, all groups, except Cubans, are significantly less likely to be self-employed. The odds among African-Americans are particularly miniscule, 0.37 to 1.00. This pattern indicates that the group normally ranked at the top of the ethnic hierarchy, native whites, not only derives significant economic benefits from independent enterprise (see Table 8), but also chooses this route far more frequently than others, especially the most downtrodden ethnic minorities. The only exception in Miami/Ft. Lauderdale are Cubans who, controlling for human capital and gender differences, are significantly more likely to become entrepreneurs than Non-Hispanic Whites. This result again supports prior results highlighting the impact of the emergence of the enclave economy in this area.

The second panel of Table 10 disaggregates the Cuban population showing results that would, by now, be unsurprising. The real high level of entrepreneurship is found among pre-Mariel Cubans whose odds of doing so, controlling for human capital and gender, are 1.29 to 1.00, relative to native whites. Later arrivals are not significantly different from the reference category in pursuing the entrepreneurial route – a result indicating that a significant number do, albeit with poorer economic results, shown previously. The U.S.-born Cuban second generation is significantly *less* likely to be self-employed. This finding confirms prior ones to the effect that this group neither derives major income benefits from entrepreneurship, nor is it likely to follow that route. Its economic prowess comes from elsewhere, namely using the resources of the enclave to pursue highly paid professional careers. The well-documented rapid ascent of the Jewish second generation out of the Lower East Side and into the professional ranks in New York City during the mid-twentieth century (Rischin, 1962; Goldschneider, 1986) is closely mirrored in this Miami pattern by the end of the century.

Inter-ethnic differences on determinants of self-employment are presented in Table 11. The principal story here is the contrasting effects of educational achievement on employment for native whites and Asians (grouped in the “Other” category), on the one hand, and Cubans, on the other. For the first two groups, *any* level of education above high school has a positive effect on

**Table 11.** Within-Group Determinants of Self-Employment in Miami-Ft. Lauderdale, 2000 (Adults 18–64)<sup>a</sup>.

	White	Black	Pre-1980 Cuban	1980+ Cuban	U.S.-born Cuban	Other Hispanic	Other
Female	0.4375*** (0.0183)	0.5714***† (0.0548)	0.3628*** (0.0328)	0.3399***† (0.0382)	0.3439*** (0.0684)	0.7943***† (0.0471)	0.4932*** (0.0679)
Work Experience (U.S.)	1.0103 (0.0056)	0.9764*† (0.0101)	0.9899 (0.0234)	1.211***† (0.0616)	1.2154***† (0.0384)	0.9972 (0.0098)	0.9904 (0.0168)
Work Experience (U.S.)	1.0001 (0.0001)	1.0009***† (0.0003)	1.0010 (0.0008)	0.9855**† (0.0052)	0.9960*** (0.0010)	1.0001 (0.0003)	1.0001 (0.0004)
<i>Education:</i> <sup>b</sup>							
High School	1.0828 (0.1128)	1.1229 (0.1520)	0.7515*† (0.1114)	0.8660 (0.1122)	0.3999**† (0.1535)	0.9567 (0.0841)	1.3862 (0.3190)
Some College	1.0257 (0.1051)	0.7683 (0.1158)	0.7254* (0.1064)	0.5783***† (0.0931)	0.5452† (0.1899)	0.8164* (0.0732)	1.7405*† (0.3960)
College	1.1942 (0.1251)	0.8778 (0.1766)	0.6598*† (0.1144)	0.5184**† (0.1138)	0.4668*† (0.1789)	0.9948 (0.1031)	1.2036 (0.3010)
Post-Graduate	1.7088*** (0.1823)	1.3166 (0.2853)	1.4737* (0.2590)	0.9572† (0.1936)	1.6421 (0.6362)	1.3070* (0.1473)	2.0515**† (0.5325)
<i>N</i>	25,383	12,003	4,914	5,021	2,069	14,546	3,019
Pseudo <i>R</i> <sup>2</sup>	0.03	0.02	0.05	0.05	0.10	0.001	0.03
Wald $\chi^2$	552.16	84.32	163.67	150.57	93.93	39.21	50.38

Source: IPUMS 2000 (5% microsample).

Note: Universe includes adult males aged 18–64, who are not unemployed, and whose annual income is greater than or equal to \$500. Raw *N* values included; person weight used.

<sup>a</sup>Logistic regression, net odds-ratios to the fourth digit. Standard deviations in parentheses.

<sup>b</sup>Less than high school is the reference category. \*\*\**p* < 0.01; \*\**p* < 0.01; \**p* < 0.05, two-tailed.

† = significantly different than Non-Hispanic Whites.

self-employment, although only a post-college education yields a significant effect for both groups. For Cubans, on the other hand, all educational levels below an advanced post-graduate education have *negative* effects. This is as true of pre-Mariel Cubans as of later arrivals and the second generation.

Since educational coefficients are computed relative to high-school dropouts, this pattern of results suggests dual entry routes into entrepreneurship. It also indicates that the previously observed finding of higher rates of self-employment among high-school dropouts is entirely due to the economic behavior of Cubans in this area. Both the highly educated (best represented among the earlier exiles) and the least educated (better represented among the post-1979 cohorts) took the entrepreneurial route in significantly higher numbers than the rest of their co-ethnics and other ethnic groups in the city. For those at the top, it has represented a vehicle for rapid economic advancement. For those at the bottom, it has been primarily a vehicle for survival since, as seen previously, the group most clearly associated with this form of self-employment did not derive a significant economic payoff from it. The story suggested by these results is that educationally disadvantaged refugees, common among the later Cuban waves, did try to emulate the business prowess of established enclave entrepreneurs, but without the same results. This reflects again the social and economic bifurcation of the Miami Cuban community, as described previously.

## CONCLUSION

The development of theoretical concepts – what Weber ([1904]1949) called “ideal types” – represents the culmination of successful long-term inquiry into any area of social or historical reality. Concepts are valuable heuristic tools that highlight certain features of the social world and summarize, in a word or phrase, what is known about them. By the same token, concepts stimulate further investigation through the explanations and predictions (hypotheses) that they generate. However, once formulated, concepts have the character of a “public good”, freely available to everyone. This availability is, in principle, a good thing, but it runs the risk of misrepresentation or overuse of the original idea. In their quest to make their mark on the scientific world, investigators may stretch the intended meaning of the concept, all the while claiming that they are putting it to rigorous empirical test. The more a concept gains visibility and popularity among the lay or scientific publics, the more it is prone to this pattern of distortion (Merton, 1968, Chapter 1).

Something of the sort took place after the formulation of the concept of ethnic enclave a quarter of a century ago. As seen in the introductory sections, attempts at testing its implications have been frequently marred by faulty operationalizations or by outright misinterpretations of the theory. While it is true that finding appropriate measures of what an enclave is and of who is or is not an enclave entrepreneur is difficult, there is also a willful resolve – most common among some economists – to demonstrate that the independent business route does not pay for minorities and, hence, that the possibility of successful ethnic entrepreneurship is a “myth”.

In this context, a retrospective glance at the concept of enclave and at the immigrant group most closely associated with its formulation seems worthwhile. The two resources needed to accomplish this are fortunately at hand: (a) recent census data on the economic performance of different ethnic groups in the relevant metropolitan area; and (b) knowledge of the history of the target immigrant nationality in the two intervening decades. Without this kind of knowledge, the analysis and interpretation of census figures would become meaningless, a lesson that extends to the investigation of similar ethnic formations elsewhere.

Results of this analytic exercise correspond fairly well to the original theoretical expectations in several ways: the total and net economic payoffs to entrepreneurship; the prevalence of this form of economic adaptation among certain groups, but not others; and the determinants of entry into it within and between these groups. A 20-year retrospective offers an authoritative standpoint to assess the economic situation and performance of any immigrant minority. From this perspective, we have seen that the economic trajectory of Cuban exiles in Miami has been enviable, placing them at par, if not higher in some respects, than the white native segment of the population. This overall characterization offers, in our view, the best validation of the original concept as it is clear that this group could not have moved ahead without a rapid process of firm creation, backed by dense business networks. The pursuit of this route placed the earlier cohorts of Cuban exiles and their offspring at the top of an economically dynamic city. No other immigrant or ethnic group in it, except native whites themselves, have approached this situation.

A hypothetical parallel analysis of the same Census data that ignored the historical development of the Miami Cuban community would have completely distorted the facts by aggregating all members of this population into a single ethnic category. The emerging results would have been quite different and would have largely negated expectations stemming from the enclave hypothesis. As we have seen, the actual story has been different: benefits of



enclave entrepreneurship accrued primarily to its original builders and, secondarily, to their offspring. The latter did not follow in their parents' footsteps but used their accumulated resources to obtain advanced degrees and enter well-paid careers. Later Cuban arrivals also followed the entrepreneurial route, but they were hampered in their efforts by the rupture of networks (and hence the opportunities and resources that they provided) with the older exile community.

Throughout this period, the only other group that paralleled pre-Mariel Cubans in rates of entrepreneurship and in the benefits derived from it were native whites themselves. This should put to rest some economists' stories that the only "real" route to economic mobility is through salaried employment. When the group at the top of the ethnic hierarchy eagerly pursues entrepreneurship as an alternative to salaried work, we can be certain that its returns are anything but a "myth". For the Cubans, as for other immigrant groups such as the Chinese and the Koreans, as well as for the Russian Jews and the Japanese in the early 20th century, the building of a business enclave was the key tool that allowed them to carve a socially respectable and economically viable place in the midst of American society. Absent this tool, their fate in a highly competitive and racially stratified economy would have been quite different.

## NOTES

1. For example, a low absolute rate of return per year of education for women may yield a high proportional rate because the log linear coefficient reflects an effect relative to the mean earnings of women, which tend to be much lower than men's.

## REFERENCES

- Bailey, T., & Waldinger, R. (1991). Primary, secondary, and enclave labor markets: A training system approach. *American Sociological Review*, 56, 432-445.
- Bates, T. (1987). Self-employed minorities: Traits and trends. *Social Science Quarterly*, 68, 539-551.
- Bates, T. (1989). The changing nature of minority business: A comparative analysis of Asian, non-minority, and black-owned business. *The Review of Black Political Economy*, 18, 25-42.
- Bates, T. (1993). Helping black entrepreneurs. *Wall Street Journal*, December 13, p. A14.
- Bates, T., & Dunham, C. (1991). The changing nature of business ownership at a route to upward mobility of minorities. Presented at the conference on urban labor markets and labor mobility, Sponsored by the Urban Institute, March 7-8, Arlie House, VA.

- Bobo, L. D., Oliver, M. H., Johnson, J. H., & Valenzuela, A. (2000). *Prismatic metropolis: Inequality in Los Angeles*. New York: Russell Sage Foundation.
- Borjas, G. J. (1986). The self-employment experience of immigrants. *The Journal of Human Resources*, 21, 485–506.
- Borjas, G. J. (1990). *Friends or strangers, the impact of immigrants on the U.S. economy*. New York: Basic Books.
- Camayd-Freixas, Y. (1988). *Crisis in Miami: Community context and institutional response in the adaptation of Mariel Cubans and undocumented Haitian entrants in South Florida*. Commissioned Report. Boston Urban Research and Development.
- Fong, T. P. (1994). *The first suburban Chinatown: The remaking of Monterey Park, California*. Philadelphia: Temple University Press.
- García, M. C. (1996). *Havana USA: Cuban exiles and Cuban Americans in South Florida, 1959–1994*. Berkeley, CA: University of California Press.
- Gilbertson, G. A. (1995). Women's labor and enclave employment: The case of Colombian and Dominican women in New York City. *International Migration Review*, 29, 657–670.
- Goldscheider, C. (1986). *Jewish continuity and change: Emerging patterns in America*. Bloomington, IN: Indiana University Press.
- Guarnizo, L. E. (1994). Los 'Dominican Yorkers': The making of a binational society. *Annals of the American Academy of Political and Social Science*, 533, 70–86.
- Guarnizo, L. E., Sanchez, A. I., & Roach, E. (1999). Mistrust, fragmented solidarity, and transnational migration: Colombians in New York and Los Angeles. *Ethnic and Racial Studies*, 22(March), 367–396.
- Hodson, R. (1985). Some considerations concerning the functional form of earnings. *Social Science Research*, 14, 374–394.
- Howe, I. (1976). *World of our fathers*. New York: Harcourt, Brace, and Jovanovich.
- Hum, T. (2000). A protected niche? Immigrant ethnic economies and labor market segmentation. In: L. D. Bobo, M. H. Oliver, J. H. Johnson & A. Valenzuela (Eds), *Prismatic metropolis: Inequality in Los Angeles* (pp. 279–314). New York: Russell Sage Foundation.
- Itzigsohn, J., Dore, C., Fernández, E., & Vazquez, O. (1999). Mapping Dominican transnationalism: Narrow and broad transnational practices. *Ethnic and Racial Studies*, 22(March), 316–339.
- Light, I., & Bonacich, E. (1988). *Immigrant entrepreneurs: Koreans in Los Angeles 1965–1982*. Berkeley: University of California Press.
- Light, I., & Gold, S. J. (2000). *Ethnic economies*. San Diego, CA: Academic Press.
- Light, I., & Rosenstein, C. (1995). Expanding the interaction theory of entrepreneurship. In: A. Portes (Ed.), *The economic sociology of immigration* (pp. 166–212). New York: Russell Sage.
- Logan, J. R., Alba, R. D., & McNulty, T. L. (1994). Ethnic economies in metropolitan regions: Miami and beyond. *Social Forces*, 72, 691–724.
- Merton, R. K. (1968). *Social theory and social structure* (Enlarged ed.). New York: Free Press.
- Perez, L. (1992). Cuban Miami. In: G. J. Grenier & A. S. Gainesville (Eds), *Miami now* (pp. 83–108). FL: University Press of Florida.
- Perez, L. (2001). Growing up Cuban in Miami: Immigration, the enclave, and new generations. In: R. G. Rumbaut & A. Portes (Eds), *Ethnicities: Children of immigrants in America* (pp. 91–125). Berkeley, CA: University of California Press and Russell Sage Foundation.
- Petersen, W. (1971). *Japanese Americans: Oppression and success*. New York: Random House.
- Portes, A., & Bach, R. L. (1985). *Latin journey: Cuban and Mexican immigrants in the United States*. Berkeley, CA: University of California Press.

- Portes, A., & Jensen, L. (1989). The enclave and the entrants: Patterns of ethnic enterprise in Miami before and after Mariel. *American Sociological Review*, 54, 929–949.
- Portes, A., & Rumbaut, R. (1996). *Immigrant America: A portrait* (2nd ed.). Berkeley: University of California Press.
- Portes, A., & Stepick, A. (1993). *City on the edge: The transformation of Miami*. Berkeley: University of California Press.
- Portes, A., & Zhou, M. (1996). Self-employment and the earnings of immigrants. *American Sociological Review*, 61, 219–230.
- Portes, A., & Zhou, M. (1999). Entrepreneurship and economic progress in the 1990s: A comparative analysis of immigrants and African Americans. In: F. Bean & S. Bell-Rose (Eds), *Immigration and ethnicity in the United States* (pp. 143–171). New York: Russell Sage.
- Raijman, R., & Tienda, M. (1999). Immigrants' socio-economic progress post-1965: Forging mobility or Survival? In: P. K. C. Hirschman & J. DeWind (Eds), *The handbook of international migration* (pp. 239–256). New York: Russell Sage Foundation.
- Rischin, M. (1962). *The promised city: New York Jews 1870–1914*. Cambridge, MA: Harvard University Press.
- Rumbaut, R. G. (1990). The structure of refuge: Southeast Asian refugees in the United States, 1975–1985. *International Review of Comparative Social Research*, 1, 95–127.
- Rumbaut, R. G. (1994). Origins and destinies: Immigration to the United States since World War II. *Sociological Forum*, 9, 583–621.
- Sanders, J., & Nee, V. (1987). Limits of ethnic solidarity in the enclave economy. *American Sociological Review*, 52, 745–773.
- Stepick, A., Grenier, G., Castro, M., & Dunn, M. (2003). *This land is our land: Immigrants and power in Miami*. Berkeley: University of California Press.
- U.S. Bureau of the Census. (2000). *Five percent public use microdata sample*. Washington, DC: U.S. Department of Commerce.
- U.S. Bureau of the Census. (2002). *Survey of business owners – Hispanic-owned firms*. Washington, DC: U.S. Department of Commerce.
- Waldinger, R. (1996). *Still the promised city? African-Americans and new immigrants in post-industrial New York*. Cambridge, MA: Harvard University Press.
- Waldinger, R. (1986). *Through the eye of the needle: Immigrants and enterprise in the New York garment trade*. New York: New York University Press.
- Weber, M. ([1904]1949). *The methodology of the social sciences*. Translated by E. A. Shils and H. A. Finch. New York: The Free Press.
- Wilson, K., & Martin, W. A. (1982). Ethnic enclaves: A comparison of the Cuban and black economies in Miami. *American Journal of Sociology*, 88, 135–160.
- Wilson, K., & Portes, A. (1980). Immigrant enclaves: An analysis of the labor market experiences of Cubans in Miami. *American Journal of Sociology*, 86(September), 295–319.
- Zhou, M. (1992). *New York's Chinatown: The socioeconomic potential of an urban enclave*. Philadelphia: Temple University Press.
- Zhou, M. (2004). Revisiting ethnic entrepreneurship: Convergencies, controversies, and conceptual advancements. *International Migration Review*, 38(Fall), 1040–1074.
- Zhou, M., & Logan, J. (1989). Returns on human capital in ethnic enclaves: New York City's Chinatown. *American Sociological Review*, 54(October), 809–820.

# THE FORCE OF REGULATION IN THE LAND OF THE FREE: THE PERSISTENCE OF CHINATOWN, WASHINGTON DC AS A SYMBOLIC ETHNIC ENCLAVE ☆

Ching Lin Pang and Jan Rath

## ABSTRACT

*Like many other cosmopolitan cities, Washington, DC has a Chinatown, a site of leisure and consumption, based on the commodification and marketing of ethno-cultural diversity. The successful transformation of an ethnic precinct into a tourist attraction depends on supportive economic and social infrastructure as well as on the flourishing of small-businesses, commodifying ethnic features. For sure, this Chinatown does not represent the nodal point of a vibrant community. On the contrary, it is artificially kept alive by city planners and a handful of self-appointed Chinese spokespersons through its inclusion in DC's regulatory structures that strongly support and promote ethnic theming.*

☆The authors carried out fieldwork during a visitorship at the Migration Policy Group in Washington, DC in 2003. They wish to express their gratitude to MPI – Demitris Papademitriou and Joanne van Selm in particular – for their support, and Kirin Kalia for her comments on an earlier version of this article.

---

**The Sociology of Entrepreneurship**

**Research in the Sociology of Organizations, Volume 25, 191–216**

**Copyright © 2007 by Elsevier Ltd.**

**All rights of reproduction in any form reserved**

**ISSN: 0733-558X/doi:10.1016/S0733-558X(06)25006-7**

## INTRODUCTION

Like many other cosmopolitan cities, Washington, DC, promotes its Chinatown. This ethnic precinct is roughly located north of a line extending the midpoint of G Place eastward from 9th to 6th Street, only a few blocks northwest of Capitol Hill. Each year, hundreds of thousands of visitors come to the precinct by public transportation – there is a Chinatown Metro Station at Gallery Place – or by other means, and find their way to the cafes, restaurants and shops and the neighboring Verizon Center, home to professional sports teams, pop concerts and family events. The ‘world’s largest single-span Chinese arch’ at 7th and F Streets, the banners and store signs in Chinese characters, the decorative fences and brick pavers, the annual Chinese New Year’s Day Parade, and various other markers help Chinatown play up its ethnic character (Picture 1).

DC’s Chinatown exemplifies a familiar phenomenon, namely the proliferation of urban space as site of leisure and consumption, based on the commodification of ethno-cultural diversity, located in a formerly derelict ethnic precinct. In itself this phenomenon is not particularly new, certainly



*Picture 1.*

not in traditional countries of settler immigration, such as the United States, Canada, and Australia. As far back as the 1880s, it became fashionable for middle-class New Yorkers to go slumming or 'rubbernecking' in Chinatown (Lin, 1998). Many others have followed these trendsetters and today, more than one century later, various ethnic 'enclaves' or 'ghettos' are indeed destinations of mass tourism (Anderson, 1990; Conforti, 1996; Shaw, Bagwell, & Karmowska, 2004; Bodaar & Rath, 2005; Rath, 2005; and various others).

While representing a wide range of political, economic, social, and cultural spaces within the urban landscape, Chinatown as an urban locale evokes various images in the 'Western' popular mind. These images are related to 'vice' and 'exoticism' are rather persistent. Chinatown is repulsive and seductive at the same time. Interestingly, Chinatown as an urban space has undergone significant transformations, largely as a consequence of globalization on diasporic communities in the North American context. As a result, the nature of migration flows have changed, cities have taken on a more global than local role, social identities have become deterritorialized, and cosmopolitan cultures have emerged (Cohen, 1997). These changes are also reflected in ethnic settlements, transforming the traditional ethnic enclave into 'ethnoburbs' (Li, 1998) or heterolocal communities (Zelinsky & Lee, 1998). Whereas the 'traditional' Chinatown represents inner-city ethnic enclaves, ethnoburbs are multiethnic communities in which one ethnic group has a significant position (Miller, 2004). Incidentally, we are aware of the debate, held among North American scholars in the 1990s, about the phenomenon of ethnic enclaves. This debate revolved around niche-like concentrations of immigrant and ethnic minorities in the urban economy and focused on the role of ethnic entrepreneurs and ethnic workers (see for instance Portes & Manning, 1986; Portes & Jensen, 1987; Sanders & Nee, 1987; Waldinger, 1993; Logan, Alba, & Jones, 2003). In this article, however, we define enclaves in a more simple way, namely as a territory that distinguished itself in a political or cultural way from its surroundings. We are, moreover, aware when discussing these enclaves that a North American typology of Chinatown dominates, a typology that may not fully appreciate European (Christiansen, 2003) and Asian experiences (Yamashita, 2003).

Anyway, visiting Chinatown, but also Little Italy, Little Saigon, Finntown, Banglatown or whatever other place, has become part of the more general phenomenon of 'cultural tourism'. Cultural tourism is based on the use of cultural symbols and reflects various societal changes, one of them being the changing perception and appreciation of the cultural Other. While ethno-cultural enclaves were still very much associated with lower social classes in the 1950s and 1960s and in some places even in the 1970s, they no longer

stood exclusively for insularism, poverty, inferiority, vice, and social backwardness by the mid-1980s (Wong, 1995; Li, 1998). Another change refers to the transformation of urban economies that are no longer based on manufacturing industries but ever more on service industries, and that revolve around the production, circulation, and consumption of information (Tsu, 1999; cf. Rath, 2006). The commodification of real or perceived (ethno-)cultural symbols is contingent on the expansion of cultural economies and simultaneously contributes to it and may, at the same time, foster the transformation of dilapidated ethnic streetscapes into places of cultural consumption (Zukin, 1995). A growing number of leisure seekers, visitors, travelers, and business persons gravitate to these places and enjoy the products and services that are offered in these new cultural economies. In doing so, they strengthen these developments and encourage local entrepreneurs, business developers, and city boosters to continue on that track.

At present – and notwithstanding economic recession, threats of terrorism and bird flu in several parts of the world – the tourism and leisure economy is a growing sector. What makes this development even more interesting is the fact that the tourism and leisure economy is one of the few growth sectors that are all-inclusive: it provides numerous jobs and business opportunities to both high- and low-skilled immigrants of both genders (World Travel and Tourism Council, 2005). Immigrants are, sure enough, involved in this economy as wage laborers or as entrepreneurs. In their capacity as entrepreneurs, immigrants are active as producers of a range of tourist services and attractions, varying from restaurants, travel agents, and gift shops to festivals and street parades. There can be no mistake that these entrepreneurs are central to the transformation of shopping strips or shopping malls into ‘exotic’ ethnic precincts.

An ethnic precinct’s tourist potential is exploited best when it is embedded in a larger tourism industry (Hope & Klemm, 2001). Ethnic precincts – except perhaps the manufactured ones such as the Chinatown in Las Vegas or the China Pavilion in Walt Disney’s Epcot theme park in Orlando – are typically the product of immigrant ethnic communities. Most grew without any organized plans. They exist because immigrants have carved out spaces that have served as nodal points of community life. For the exploitation of its tourism potential immigrants do not actually have to live in that area. Leichhardt is Sydney’s Little Italy, but has ceased to be home to the Italian population. The Italian community nonetheless meets in Leichhardt, where the sights, sounds, flavors and irresistible aromas of Italy come alive in numerous Italian bars and cafes (Collins & Castillo, 1998; cf. Halter, 2006). The Eden Center just outside Washington, DC, is described as the heart and

soul of the Vietnamese community for the entire East Coast. Eden Center continues to grow and extend its financial and community support into the surrounding Vietnamese and Asian community. The Center, however, is 'just' a shopping mall, not a residential area, that serves a population that shows a dispersed pattern of residential location (Wood, 1997; Zelinsky & Lee, 1998). What matters is that this shopping mall provides a public space where people can meet co-ethnics and behave in a manner familiar to them, and where tourists can indulge and consume cultural diversity.

In the eyes of tourists, these are shops, i.e. a particular type of shops, that give the neighborhood its ethnic flavor and foster the kind of public life that give cultural tourists an excuse to linger. Imagine an ethnic enclave without shops and businesses. Only a few cultural tourists will be interested in strolling along nondescript houses or be attracted to, say, Korean accountancies, Indian construction businesses, or Hispanic exhaust centers. On the other hand, easily accessible book and music stores, gift shops, bric-a-brac shops, travel agents, and especially restaurants, groceries, and supermarkets do have the capacity to attract non-coethnic leisure seekers. This is no coincidence, as food is one of the cultural features that people tend to retain over a longer period of time and that may help bridge cultural differences at the same time (Van den Berghe, 1980; see also Diner, 2001; or Valle & Torres, 2000, for more nuanced and critical perspectives). Ethnic precincts hold additional attraction when cultural tourists are able to visit particular ethnic institutions (such as churches, temples, and mosques, but also community center; see Lalich, 2003) or attend cultural events, such as New Year parades, food festivals, or other public manifestations. Whatever combination of products and services are supplied, the point is that immigrant entrepreneurs are key figures in these developments (Kunz, 2005).

Let us return to Chinatown, Washington, DC. There can be no mistake that a 'real' Chinatown does exist. That is, city maps and 'official' tourist guides invite visitors to this 'colorful, diverse neighborhood' and its 'numerous restaurants'. However, key actors involved do not seem to be satisfied with current developments. Community leaders, real estate developers, city planners, business support people, cultural tourism marketers, and local researchers express serious concerns about the future of DC's Chinatown. A typical account of the situation is this newspaper report:

But Washington's Chinatown has been surrounded and flooded by dramatic change. Seventh Street NW has been transformed into a strip of restaurants and trendy stores. The block north of MCI Center is home to the 275,000-square-foot Gallery Place with a 14-screen theater, fashionable shops and a spa. Upscale apartment buildings stand to the north and the east. More are on the way. The neighborhood has become a boomtown,



but there's increasingly less 'China' in Chinatown. As the area's Chinese community gathered yesterday for its New Year's celebration and parade, there was a palpable sense that something is being lost, if not by subtraction then by dilution. "In another few years, you won't see Chinatown," said Thomas Lee, past president of the Chinese American Citizens Alliance. "You'll be hard-pressed to find the arch because it will be dwarfed by everything else." (*Washington Post*, February 14, 2005)

The critical voices revolve especially around issues of authenticity and the credibility of its Chinese distinctiveness, and thus about the unique selling point of this precinct (Pictures 2 and 3).

Indeed, when visiting DC's Chinatown, one first feels more disappointment than exaltation. This Chinatown is clearly *not* the nodal point of a vibrant community. Strolling along the many stores, one hardly passes by Chinese residents or Chinese consumers. Mainstream chain stores, such as Starbucks and Fadó, but also less swanky places such as Hooters, Fuddruckers, Ruby Tuesday, Radioshack, and CVS Pharmacy, and especially the huge Verizon Center – previously named the MCI Center – dominate the streetscape. Anyone can observe how these stores symbolically and sometimes also literally overshadow the distinctive 'Chinese-ness' of Chinatown. The Chinese



Picture 2.



*Picture 3.*

presence seems to be confined to just one block. There are the Wah Luck House, which is a home for Chinese elderly, a dozen or more inexpensive eateries and a few more up-market restaurants, including the Hunan Chinatown, the Golden Palace, and Tony Cheng's Mongolian restaurant. New mainstream developments, however, dwarf these places. Some, therefore, cynically refer to Chinatown as the 'China-block' (Pictures 4 and 5).

The local actors find themselves in a puzzling and paradoxical situation. These actors seem to be searching for credible representation and imaging, but the outcome resembles a makeshift, Disneylandish Chinatown. Ethnic Chinese are conspicuous by their absence and one misses the bustle that is so characteristic for other Chinatowns like the ones in San Francisco or Manhattan, New York. Indeed, it seems that this precinct has ceased to be the spatial heart of a vigorous ethnic community. Assuming that this is the case, the real interesting question becomes: what accounts for the persistence of Chinatown as a symbolic ethnic enclave? Why do entrepreneurs continue to attract clients by collectively playing up a real or pretended Chinese character. And what does this tell us about the structural determinants of small entrepreneurship, ethnic as much as nonethnic?



*Picture 4.*

In this article, we will argue that the revitalization of DC's Chinatown coincides with the expansion of DC's central business district and concomitant penetration of corporate capitalism, but that the precinct continues to exist, albeit in a more symbolic way, first and foremost because of governmental regulation. In what follows, we will first briefly discuss ethnic/immigrant entrepreneurship theory and introduce the mixed embeddedness approach. Thereafter, we will describe and analyze the process of regularization of Chinatown as a one-of-a-kind commercial enclave.

## **ENTREPRENEURSHIP THEORY**

Let us now examine the literature of immigrant/ethnic entrepreneurship. Most studies of immigrant/ethnic entrepreneurship focus on entrepreneurs only and are mainly interested in explaining the proclivity of certain groups toward entrepreneurship and their paths to entrepreneurial success.<sup>1</sup> Scholars studying these questions have developed several theoretical approaches, ranging from those emphasizing the cultural endowments of



Picture 5.

immigrants (i.e., certain groups are culturally inclined towards risk-taking behavior; see Light, 1972; Metcalf, Modood, & Virdee, 1996), to others that highlight racist exclusion and blocked mobility in the regular labor market (i.e., marginalized individuals are driven towards entrepreneurialism as a means of escaping unwelcoming labor markets; see Ram, 1993; Collins, Gibson, Alcorso, Tait, & Castles, 1995; Barrett, Jones, & McEvoy, 1996; Saxenian, 1999).

A number of years ago, Waldinger and his associates (1990) developed a composite theory that brought together these views, based on the principle that entrepreneurship is the product of the interaction between group characteristics and the opportunity structure. As such their *interactive model* combines ethno-cultural and socio-cultural factors (agency) with politico-economic factors (structure). According to Waldinger et al., the latter entail market conditions (particularly access to ethnic/nonethnic consumer markets) and access to ownership (in the form of business vacancies, competition for vacancies, and government policies). This interactive model has been appreciated as an important step towards a more comprehensive theoretical approach, even though it is more of a classification than an

explanatory model. However, it has also been subjected to criticism. Its shortcomings included its methodology (Light & Rosenstein, 1995), the lack of attention devoted to issues of class and gender (Morokvasic, 1993; Collins et al., 1995), insufficient emphasis on processes of racialization of immigrants (Collins et al., *ibid.*), a priori categorization of immigrants as ethnic groups and the concomitant assumption that immigrants as ethnic entrepreneurs act differently than mainstream entrepreneurs (Kloosterman & Rath, 2003), and the narrow and static way economic and politico-regulatory factors are dealt with (Bonacich, 1993; Rath, 2000, 2002). As regards the latter, the authors conceive market conditions in terms of the ethnicization or de-ethnicization of consumer markets, and confine politico-regulatory factors to a shortlist of laws and regulations that specifically apply to immigrants.

Theoretical development has continued but, oddly enough, this has led to a convergence of approaches to issues of *social embeddedness*, that is, the assumption that individual entrepreneurs participate in ethnically specific economic networks that facilitate their business operations (especially in acquiring knowledge, distributing information, recruiting capital and labor, and establishing relations with clients and suppliers). This would suggest that the proliferation of Chinatown as a tourist attraction would be mainly the result of the mobilization of the Chinese entrepreneurs' ethno-social networks. Their social embeddedness enables them to reduce transaction costs by eliminating formal contracts, gaining privileged access to vital economic resources, and providing reliable expectations as to the effects of malfeasance. Particularly in cases where the entrepreneurs' primary input is cheap and flexible labor, as is true of some parts of the tourism industry, the reduction of transaction costs by mobilizing social networks for labor recruitment seems key. Many students of immigrant entrepreneurship, especially in the United States, are indeed fervent adherents to a version of economic sociological thought that focuses on the entrepreneurs' social networks and impact on entrepreneurship (see for example Zhou, 1992; Wong, 1998). However, taking advantage of social embeddedness is a complex and dynamic process, is connected to cultural, human, and financial capital (Light & Gold, 2000), is contingent on the goals pursued and the political and economic forces at work (Granovetter, 1995; Kumcu, 2001), and is the product of the interaction of structural factors such as migration history and processes of social, economic and political incorporation in the mainstream as well as their spatial variations (Rath, 2002). These intricacies, however, have not always been adequately addressed.

In recent years, continental European researchers criticized this economic sociological thought for focusing on the supply side of entrepreneurship only. They argued, moreover, that these theories of immigrant entrepreneurialism assume an unregulated and undifferentiated economy, whereas in reality economies are never unregulated and undifferentiated. Scant attention has subsequently been paid to the array of *regulatory structures* that promote certain economic activities while inhibiting others. For example, while virtually anyone can establish a private business in the United States, in Germany and even more so in Austria individuals must apply for special licenses even to sell flowers in restaurants and bars, and they need the approval of a particular organization to engage in most forms of production or service (see for instance [Haberfellner, 2003](#)). It is thus important to address these highly relevant forms of regulation. Next to that, one must also fully appreciate the *economic dynamics* of a market. It does not require much sociological imagination to see that designers of virtual tourist guides, pencils vendors, or take-out restaurateurs operate in entirely different markets. Different markets obviously offer different opportunities and obstacles, demand different skills, and lead to different outcomes in terms of business success or – at a higher level of agglomeration – a different ethnic division of labor.

Acknowledging the salience of regulation as well as market dynamics, researchers have proposed a *mixed embeddedness* approach to immigrant entrepreneurship ([Kloosterman, van der Leun, & Rath, 1999](#); [Kloosterman & Rath, 2001, 2003](#); [Rath, 2002](#)). The multi-scalar approach is considered to be more appropriate, since it relates social relations and transactions to wider political and economic structures. It acknowledges the significance of immigrants' concrete embeddedness in social networks, and conceives that their relations and transactions are embedded in a more abstract way in wider economic and politico-institutional structures. While appreciating the relevance of social and cultural structures for economic development, this article must be situated within this emerging analytical approach.

How does regulation work? To begin with, regulation should not be confused with legislation, as there are two other forms of regulation. There are 'sticks', which [Engelen \(2001\)](#) refers to as 'legislation per se', and 'carrots' (financial incentives and disincentives) or 'sermons' (persuasion), all different forms in complex packages that define what is 'possible' in a market. Nor should regulation be confused with state regulation. A multitude of agents play a role in regulation processes, such as local, national or international governmental agents, unions, quangos, not-for-profit organizations, voluntary associations, and individual and their social networks.

Regulation can be manifested in thick or thin ways or can either be imposed or enforced or be a matter of voluntary action.

These notions are important, as they make it clear that regulation is not just a matter of repression and constraining, but also of enabling. Suppressing illicit practices such as dodging taxes and labor and immigration laws by prosecuting the perpetrators are important manifestations of regulation (repression), but so are decisions to tolerate these practices and not prosecute them. The plethora of business support programs also constitutes forms of regulation (Dreef, 2004), as these are efforts to change the market landscape.

These notions make it clear that regulation occurs in advanced welfare states, but also in liberal welfare states. The United States government, for instance, is admittedly a relatively lean government that supposedly has less means to regulate economic life, but this is amply compensated for by the regulation of a more voluntary nature. In the land of the free, economic life (and not only economic life) is severely dogged by litigation. In addition, the federal government has various instruments to regulate markets. Its relatively open immigration programs for professionals and businessmen have enhanced the proliferation of money-makers in Silicon Valley, where immigrant entrepreneurs own a quarter of the high tech companies (Saxenian, 1999). In the same vein, local governments or private organizations or coalition of the two may deploy a plethora of instruments to interfere in the market economy, varying from business support schemes, economic development zone programs, zoning laws, place marketing and so forth. The tourism industry, as we will see, is a case in point.

Let us turn now to the economic processes that foster the growth of urban cultural tourism industry. The growth of this industry is intricately linked with the rapid transformation of the manufacturing economy to the information economy and beyond. Deindustrialization resulted in the need for localities to differentiate themselves in order to attract a share of this spatially mobile capital. In the case of cities in particular, authorities ranging from local governments to marketing consortia have been striving to present localities as attractive to potential investors, employers, inhabitants, and tourists (Kearns & Philo, 1993). Urban cultural diversity is then a vital resource for the prosperity of cities and a potential catalyst for socio-economic development, particularly since business investors consider this diversity as one of the factors determining the location of businesses.

Cities, faced with job losses and decay, engage in 'a desperate struggle for survival' and one after the other bet on the tourism and leisure industry, a sector with few barriers to entry and the potential for large returns

(Judd & Fainstein, 1999; Hall, 2000). The commodification and marketing of diversity, i.e. the commercial use of the presence of the ethnic Others or their symbols, fits in well with this process (Halter, 2000). The chances of this occurring are obviously contingent on the level of living, lifestyle and consumption patterns of those living in Western cities and the degree to which they develop a distinctive taste for cultural products offered by migrant and minority groups.

Zukin describes the growing enthusiasm for 'interesting' landscapes that have the potential to draw tourists (Zukin, 1995), and explores the relationship between industrial restructuring and the deterioration of factory landscapes vs. the growing significance of places of consumption. Zukin (1991, p. 16; see also Zukin et al., 1998) reminds us that landscapes are 'contentious, compromised product[s] of society' that create visual order and, in so doing, both reveal and conceal social processes. She is particularly concerned with the growing social polarization evident in many Western societies and devotes much of her effort to understanding places that appeal to affluent consumers. These include landscapes of leisure, such as Coney Island, Disney World, or Las Vegas, as well as gentrified inner-city neighborhoods that contain mixed land uses. These places contribute both materially and symbolically to the urban economy and are therefore highly prized by planners and city boosters (Zukin, 1998).

To foster this process, city governments attempt to attract investment (or invest themselves) in high-profile events, institutions, and symbolic land uses, such as Olympic games, international sports teams, and towers or special bridges. Some also emphasize areas of the city that may interest local consumers or tourists, including ethnic festivals and ethnic precincts that offer a wealth of goods and services that appear exotic, exciting, and authentic (Knecht & Soysal, 2005). As Zukin notes, this has led to a sea-change in the way these types of areas are understood and represented by the state: 'Elected officials who, in the 1960s, might have criticized immigrants and nontraditional living arrangements, now consciously market the city's diverse opportunities for cultural consumption' (1998b, p. 836). This process leads to a commodification of diversity and has led to a situation where culture – particularly the more 'sanitized' manifestations of immigrant and minority cultures – can be seen as an economic resource for cities. In practice, governmental and nongovernmental regulation may support, or at least not thwart the transformation of ethnic precincts into tourist attractions. This can be accomplished by passing favorable zoning regulations, creating a clean and safe environment and ensuring the area's accessibility.



## CHINATOWN AND THE REGULATION OF ETHNIC THEMING

How did this work out in our case? The Chinese presence in Washington, DC, dates from the mid-19th century (Chow, 1996). DC was a secondary destination, never a primary destination, as the Chinese who arrived in the 20th century came there from other US states. They were low skilled, suffered from racial exclusion, and flocked to a neighborhood somewhere between Pennsylvania Avenue, Constitution Avenue, and 15th Street, NW. In the 1920s and 1930s, they were forced to vacate their houses and businesses in order to make place for the construction of a series of public buildings in what later became the Federal Triangle government office complex. The Chinese relocated to a neighborhood that was originally inhabited by German and Jewish immigrants, i.e. the location of today's Chinatown. As elsewhere in the United States, the Chinese faced many difficulties in the labor market and, consequently, gravitated to self-employment. They managed to carve out a niche in laundry services and, when this industry became obsolete, they entered the catering business. Until the 1960s, the overwhelming majority of Chinese in the District of Columbia lived in Chinatown. Many houses and shops were marked with decorative metal latticework and railings as well as Chinese signage.

In the 1960s and 1970s, major societal changes unfolded. At this juncture, many cities, especially those that were dependent on manufacturing economies, were in decline. Washington, DC, being the center of public administration in the United States, never had an economy that was strongly reliant on manufacturing industries. Yet, many Washingtonians moved to greener pastures, and in so doing undermined the city's economy and tax base. This had an enormous impact on various neighborhoods, including the residential areas in downtown DC, as it magnified their sorry plight and enhanced the neglect of its public spaces. Ethnic Chinese residents who could afford it, like many other middle-class city dwellers, moved into the suburbs. This process was fostered by the upward social mobility that many Chinese had experienced, and this held particularly true for better-educated, second-generation immigrants who were increasingly fed up with the rising crime, rising taxes, and deteriorating business climate. The average age of the population increased, while the average level of education decreased. Many other ethnic enclaves or ghettos witnessed similar developments (cf. Wilson, 1987; Zhou, 1992). But what distinguished DC's Chinatown from other Chinatowns, such as the ones in New York, was that only a few new immigrants arrived to fill the vacancies of those who left the

neighborhood. Chinatown, consequently, gradually ceased to be the vibrant heart of the Chinese community. Admittedly, Chinese immigration did continue, but most newcomers were college-educated professionals who moved directly into the suburbs. The newcomers did not see Chinatown's symbolic and economic potential in the gentrification of metropolitan neighborhoods. Instead, heterolocalism became the new way (Wood, 1997; Zelinsky & Lee, 1998), while Chinatown became an ever more unsafe, run-down place.

Next to these spectacular socio-economic developments, there were important political developments. The political leadership of DC passed through a serious crisis following major riots in the wake of the assassination of Martin Luther King in 1968. The riots, that affected Washington, DC and 110 other American cities, revolved around issues of equal treatment of racial minorities. The civil unrest directly and indirectly devastated the economy of Washington, DC. It accelerated the closure of many businesses, the redundancy of thousands of workers, and the departure of many city dwellers of all racial and ethnic groups for the suburbs. As a result of this, property values depressed, crime increased, and new investments were discouraged. The blight of the city prompted the local government to take serious steps to address urban decline and to interface with minority groups at last. The interests of minority groups were suddenly on the political agenda and this enhanced the empowerment of African-Americans and other ethnic minority groups. It is important to note that these events took place in an era in which the assimilationist orthodoxy lost its natural dominance. With hindsight, we know that multiculturalism became *de rigueur* and this too contributed to the empowerment of minority groups that stake out claims for citizenship rights.

The city embarked on an urban renewal process, among others in the eastern part of downtown, i.e. Chinatown. Improving the streetscape and attracting businesses were given top priority, so as to stimulate the advancement of corporate businesses in Washington, DC's central business district. One of the plans designed to boost the inner city's economy entailed the development of a convention/sports center, to be located in Chinatown. This plan stirred the Chinese community for obvious reasons: the locals feared that the redevelopment would wipe out their houses and businesses and a new political crisis was lurking that could possibly thwart the process of urban renewal. In a series of meetings, self-proclaimed community leaders and city officials eventually reached a compromise. The way in which ethnic Chinese leaders and city officials defused this crisis turned out to be critical for all further developments.

First, the ethnic Chinese leaders and the city officials reached a consensus that the Chinatown area needed to be economically revitalized in a city that was becoming increasingly more gentrified and up-market. Secondly, both parties acknowledged the historic and contemporary value of Chinatown as an ethnic cultural area, and subsequently embarked on a plan to preserve and enhance the neighborhood. In so doing, the city officially acknowledged the ethnic Chinese heritage, and underscored the potential of Chinatown as a tourist attraction. Thirdly, it was decided that the convention center – not a sports center – would be built on the edge of Chinatown. Several blocks of houses had to be demolished, but in return the city supported the construction of the Wah Luck house, an apartment complex designed by a Chinese architect offering affordable residence to elderly Chinese.

This compromise marked the start of a close collaboration between city officials and Chinese community leaders who seemed to share the same objectives. This collaboration was sealed with a series of legal actions securing the protection and enhancement of Chinatown as downtown’s only ethnic cultural area, for instance in the Zoning laws and District of Columbia Municipal Regulations (see Figs. 1 and 2). This legitimized the claims of the Chinese community, legally acknowledged the existence of the precincts, and formally fixed its boundaries. Today, these rules and

7	<b>CHINATOWN</b>
927.1	The Chinatown objectives are as follows: <ul style="list-style-type: none"><li>(a) Retain and enhance Chinatown as a thriving, mixed-use Downtown community including substantial housing with community and cultural facilities, street-level retail with related wholesale operations, supporting office and professional uses, and hotels;</li><li>(b) Obtain a concentration of land uses consisting of ethnically-oriented ground floor retail uses, substantial housing and office uses, community facilities, and hotel uses as appropriate;</li></ul>
927.2	The policies established in support of the Chinatown objectives are as follows: <ul style="list-style-type: none"><li>(a) Develop a physical design criteria for new and rehabilitated buildings which will reinforce the definition and identity of Chinatown as a special cultural district;</li><li>(b) Develop a range of special design guidelines which would include building design guidelines, historic preservation relationships, and streetscape and sign criteria and that are supportive of creating a special Chinatown cultural district;</li></ul>

Fig. 1. Zoning Laws

Title 11	District of Columbia Municipal Regulations
1705.1	<p>The principal policies and objectives from the Comprehensive Plan for the Chinatown area are to:</p> <ul style="list-style-type: none"> <li>(a) Protect and enhance Chinatown as Downtown’s only ethnic cultural area;</li> <li>(b) Maintain and expand the existing concentration of retail uses emphasizing Chinese and Asian merchandise and related wholesale operations serving residents, visitors, tourists, and business travelers;</li> <li>(c) Reinforce the area’s economic viability by encouraging mixed use development, including substantial housing, cultural and community facilities, offices, retail and wholesale businesses, and hotels; and</li> <li>(d) Protect existing housing and the most important historic buildings with suitable preservation controls, residential and commercial zones, and economic incentives.</li> </ul>

Fig. 2. District of Columbia Municipal Regulations

regulations still exist. They stipulate that the city is to treat Chinatown specially to boost its ‘ethnic cultural’ character through a variety of land use and design guidelines, incentives, and special programs. These serve the preservation of the historic character and structural integrity of the precincts.

The city–community collaboration had been institutionalized by the establishment of the Chinatown Steering Committee. This committee was to monitor the developments and to act as interlocutor between the community and the government. This collaboration clearly showed the city’s willingness to make this project work. It should be noted that to date no other ethnic group in Washington, DC, has enjoyed such a favorable position.

Over the course of time, many actors actively participated in the reshaping and enhancing of the Chinatown area, including the DC Planning Office, the Downtown BID (Business Improvement District), Mainstreet Development Programs, Heritage Preservation, etc. One of the issues pertained to the distinctiveness of Chinatown, insofar as expressed in the streetscape. The Chinatown Steering Committee was encouraged by city planners to design guidelines for Chinatown. In 1976, the District government’s Chinatown Program called for design guidelines to reinforce the distinctiveness of the precinct. The Mayor’s Downtown Committee in 1982 and the comprehensive plan in 1984 once again called for that objective. This resulted in

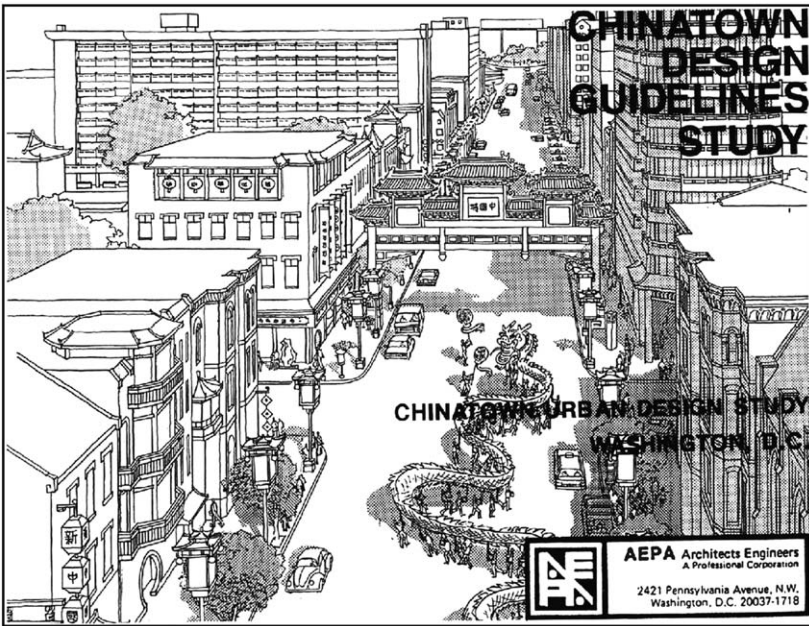


Fig. 3. The Chinatown Design Guidelines Study

*The Chinatown Design Guidelines Study*, a book written by a local consulting firm – the Architects Engineers – to enhance the Chinese character of Chinatown. The guidelines, issued in December 1988, can be seen as the crowning piece of earlier initiatives (see Fig. 3).

In the book's preface the authors write:

The Chinatown Design Guidelines Study is intended to lead to the adoption of building design guidelines and streetscape standards that will guide and assist architects, developers, and planners involved in development and renovation in Washington, DC's Chinatown. It is hoped that the criteria and guidelines suggested herein will help create an enhanced Chinatown with a strong Chinese character.

This guidebook is a clear attempt to underline and enlarge the distinctive Chinese-ness of the Chinatown by codifying Chinese culture and cultural characteristics. Drafted with care and based on scientific research, the guidelines provide very specific criteria and allow for a range of styles including traditional, modern and postmodern adaptation, while referring to existing building in China that have similar features. The near-scientific approach largely neutralizes the critique that enhancing 'Chinese-ness'

would create a caricatural version of Chinese culture. In fact, the architects sought inspiration in existing postmodern and hybrid buildings in China. The Chinatown Steering Committee and the city planners believed this initiative would make Chinatown appealing to tourist and leisure seekers, but also to overseas businesses people and investors, especially from Taiwan. The latter might be related to the fact that some committee members were connected to Taiwan. They also expected local mom-and-pop stores, the ‘local touch’, to flourish and to add to the Chinese ‘flavor’.

While the City Planning Office adopted only a short version of the design guidelines, architects were expected to take into account the symbolism of Chinatown and Chinese spirit, and thus to combine Chinese traditions and modern architecture. In reality, the prominent Chinese architecture – be it traditional, modern or postmodern adaptation – is hardly noticeable when entering DC’s Chinatown (except perhaps the Wah Luck House, the roofing of some of the restaurants, or some of the architectonic ornaments of the Verizon Center). Some striking particularities of DC’s Chinatown are Chinese signs for mainstream chain stores including Starbucks, McDonalds, and Hooters, which seem exotic in a Western city (Picture 6).



*Picture 6.*

Rather than enhancing the neighborhood's Chinese-ness, these chain stores with signs in Chinese characters look odd and therefore unconvincing. Apart from the incorrect – on purpose or not – rendition of Hooters into Chinese (the translation is 'owl'), a Starbucks with Chinese signage (but with Hispanic and other non-Chinese workers) does not launch us into the Chinese realm. One can even state that the legal provision imposing all store owners to have Chinese signage has a reverse effect. It actually tends to underline the unconvincing, unnatural nature of Chinatown instead of contributing to it. This is partly due to the fact that there is a lack of continuing input of the ethnic group itself. No progress has been made into the full-fledged development of a Chinatown in which mom-and-pop stores are thriving and larger companies capitalize on the mobilization of transnational business linkages.

Chinatown reached its peak in terms of small-scale mom-and-pop stores and restaurants in the 1980s. Up until that period, DC did not offer an extensive choice of restaurants and bars that were open during the night. For the longest time, Chinatown had been the only place offering late-night leisure activities. However, neighborhoods such as Adams Morgan and Georgetown have taken over this role. The expectation of grand-scale investment from Taiwan and the establishment of family-run business have not materialized either. On the contrary, more and more mainstream stores have found their way into the Chinatown area, watering down the 'Chinese-ness' of the area. As of 2003, 1500 new residential units have been established in Chinatown and its vicinity. Yet this development has not lead to more Chinese business in the neighborhood, and it seems more likely that high-income gentrification will continue to make its mark in the area that has de facto become part of the central business district.

As has been said, nobody seems to be sufficiently satisfied with the outcome, let alone being compelled by it. There is a smoldering conflict about the Chinese character of the area. Some rumors have it that especially the Chinese entrepreneurs in the Chinatown Steering Committee were defending their personal or business interest instead that of the larger Chinese community. The representative nature of the Chinese Steering Committee has been a bone of contention. Unlike other advisory neighborhood commissions, the members have not been elected. Furthermore, members of the committee do not live in Chinatown but in the suburbs. These self-proclaimed leaders have divergent views on the development of Chinatown: one group, comprising the more affluent and more successful business persons, wants to reach out to mainstream capital and aspires after the development of grand projects that fit into the central business district; the other group, comprising leaders

of community organizations, claims ownership of the ethnic heritage and aspires after a more festive Chinatown. According to the latter group, the members of the Chinatown Steering Committee saw to it that the area would not become too Chinese, as that would deter corporate capitalism.

## CONCLUSIONS

We argued that transforming an ethnic precinct into a tourist attraction is embedded in the symbolic and political economy. Although the main actors in this process – city planners and the Chinese community including its entrepreneurs – seem at first sight to support both the economic revival and the symbolic transformation of the Chinatown area, the DC case shows us the difficulty of successfully reshaping an ethnic precinct into a thriving business and tourist area.

This reshaping requires a social infrastructure that is able to support the development of a regular precinct into a tourist attraction as well as a proliferation of small-businesses that commodify ethnic features. Both conditions are no longer fulfilled and this is probably related to a combination of social processes, including the social and spatial mobility of second and third generation Chinese and the gradual change of the place – ethnic boundary nexus. Furthermore, DC's Chinatown never had a history of attracting newcomers 'fresh off the boat'. The formal recognition of Chinatown did not change this pattern, as new immigration of ethnic Chinese into the area hardly occurs. If it did, it could help foster the maintenance of ethnic groups and boundaries and the sustenance of ethnic community life in the way that [Glazer and Moynihan \(1964\)](#) described four decades ago. Ethnic Chinese community life is obviously changing and may even fade away in more advanced stages of assimilation.

This, however, is only one part of the story. There are also changing market conditions, notably the process of gentrification. The ethnic Chinese business community that once constituted the economic structure of the precinct has been shrinking in the past few decades. At the same time corporate business entered the precinct. The city's urban renewal programs and economic development programs enticed private capital to invest in Chinatown, and various mainstream corporations were indeed most willing to spend large numbers of dollars in such a centrally located neighborhood. The establishment of the Washington Convention Center and later also the MCI Center – recently renamed Verizon Center – have had a tremendous impact on the neighborhood. Numerous visitors and spectators flock



to Chinatown whenever there is an event and this has certainly helped improving the quality of the neighborhood and the local economy. The spending power of high numbers of consumers did attract mainstream chain stores and corporations. The flip side of this development was that small one-of-a-kind stores were pushed or bought out of the market, as they could not counter balance the power of corporate capitalism.

These mainstream corporations take the Chinese heritage for granted but hardly include it in their marketing. The Marriot Hotel, for instance, located on 900 F Street, thus on a location that once constituted the heart of Chinatown, does not even bother to refer to Chinatown on its web site:

With its spectacular \$25 million renovation, the Courtyard by Marriott Washington Convention Center has beautifully transformed the historic Riggs Bank Building into one of the most sought after Washington, DC hotels. Sharing a neighborhood with some of the city's finest restaurants, foremost businesses and government offices, you can easily walk to the Metro subway, famed museums and the Verizon Center. Enter this grand hotel, and you'll find the amenities and services that make business travel easier ...<sup>2</sup>

Under these conditions, there is no reason to assume that an ethnic precinct such as Chinatown is given a perpetual life span. There is no sustained inflow of new Chinese migrants into Chinatown. The same settlement pattern applies to newcomers, mainly highly educated migrants with a high income. Furthermore, the second generation Chinese do not show any inclination to settle in Chinatown. Instead, like their middle class counterparts, they avoid the city center. Thus with a declining presence of ethnic Chinese, a declining demand for Chinese goods and services, and a declining proliferation of Chinese merchants and other small entrepreneurs, Chinatown as an ethnic commercial enclave is slowly but surely disappearing.

Yet, there is a regulatory environment that strongly supports and promotes ethnic theming. Self-appointed Chinese spokespersons – including a number of successful entrepreneurs – have been regarded as representatives and guardians of ethnic authenticity and have, consequently, managed to gain authority. As a result of that, and perhaps at odds with the dominant representation of the United States of America as the land of free enterprise, Chinatown's symbolic economy has been included in DC's regulatory structures. Ethnic theming is obviously not required on the companies' web sites, but it is required on the streets. It is this regulation that accounts for the fact that Chinatown continues to have a 'Chinese' streetscape and that all businesses – including mainstream ones – collectively play up an 'exotic' Chinese character, in line with the popular 'Western' image.

## NOTES

1. The literature is inconclusive as to the use of concepts such as immigrant entrepreneurship and ethnic entrepreneurship (cf. Rath, 2002, 23–24).
2. <http://marriott.com/property/propertypage/WASCN>.

## REFERENCES

- Anderson, K. J. (1990). “Chinatown re-oriented”. A critical analysis of recent redevelopment schemes in a Melbourne and Sydney enclave. *Australian Geographical Studies*, 28, 131–154.
- Barrett, G., Jones, T., & McEvoy, D. (1996). Ethnic minority business. Theoretical discourse in Britain and North America. *Urban Studies*, 33(4–5), 783–809.
- Bodaar, A., & Rath, J. (2005). Cities, diversity and public space. *Metropolis World Bulletin*, 5, 3–5.
- Bonacich, E. (1993). The other side of ethnic entrepreneurship: A dialogue with Waldinger, Aldrich, Ward and associates. *International Migration Review*, 27, 685–692.
- Chow, E. N. (1996). From Pennsylvania Avenue to H. Street, NW. The transformation of Washington’s Chinatown. In: F. C. Cary (Ed.), *Urban Odyssey. A multicultural history of Washington, DC* (pp. 190–207). Washington: Smithsonian Institution Press.
- Christiansen, F. (2003). *Chinatown, Europe. Identity of the European Chinese towards the beginning of the 21st century*. London: Routledge Curzon.
- Cohen, R. (1997). *Global diasporas. An introduction*. London: UCL Press.
- Collins, J., & Castillo, A. (1998). *Cosmopolitan Sydney. Explore the World in one city*. Annandale, N.S.W.: Pluto Press with Comerford and Miller.
- Collins, J., Gibson, K., Alcorso, C., Tait, D., & Castles, S. (1995). *A shop full of dreams. Ethnic small business in Australia*. Sydney: Pluto Press.
- Conforti, J. M. (1996). Ghettos as tourism attractions. *Annals of Tourism Research*, 23(4), 830–842.
- Diner, H. R. (2001). *Hungering for America. Italian, Irish, and Jewish foodways in the age of migration*. Cambridge, MA: Harvard University Press.
- Dreef, M. E. P. (2004). *Politiek, Migranten en Informele Economie. Politieke en Bestuurlijke Ontwikkelingen ten aanzien van de Amsterdamse Confectieateliers, 1980–1997*. Ph.D. thesis, University of Amsterdam, Amsterdam.
- Engelen, E. (2001). Breaking in and breaking out. A Weberian approach to entrepreneurial opportunities. In: R. Kloosterman, & J. Rath (Eds), *Journal of Ethnic and Migration Studies*, Special issue on ‘Immigrant Entrepreneurship’, 27(2), 203–223.
- Glazer, N., & Moynihan, D. P. (1964). *Beyond the melting pot. The Negroes, Puerto Ricans, Jews, Italians and Irish of New York City*. Cambridge, MA: MIT Press.
- Granovetter, M. (1995). The economic sociology of firms and entrepreneurs. In: A. Portes (Ed.), *The economic sociology of immigration. Essays on networks, ethnicity, and entrepreneurship* (pp. 128–165). New York: Russell Sage Foundation.
- Haberfellner, R. (2003). The case of Austria: Immigrant self-employment in a highly regulated economy. In: R. Kloosterman & J. Rath (Eds), *Immigrant entrepreneurs. Venturing abroad in the age of globalization* (pp. 213–232). Oxford: Berg.

- Hall, P. (2000). Creative cities and economic development. *Urban Studies*, 37(4), 639–649.
- Halter, M. (2000). *Shopping for identity. The marketing of ethnicity*. New York: Schocken Books.
- Halter, M. (2006). Tourists ‘R’ us: Immigrants, ethnic tourism and the marketing of metropolitan Boston. In: J. Rath (Ed.), *Tourism, ethnic diversity and the city*. London: Routledge.
- Hope, C. A., & Klemm, M. S. (2001). Tourism in difficult areas revisited. The case of Bradford. *Tourism Management*, 22, 629–635.
- Judd, D. R., & Fainstein, S. S. (Eds) (1999). *The tourist city*. New Haven: Yale University Press.
- Kearns, G., & Philo, C. (Eds) (1993). *Selling places. The city as cultural capital past and present*. Oxford: Pergamon Press.
- Kloosterman, R., & Rath, J. (Eds) (2001). Immigrant entrepreneurs in advanced economies. Mixed embeddedness further explored. *Journal of Ethnic and Migration Studies*, Special issue on ‘Immigrant Entrepreneurship’, 27(2), 189–202.
- Kloosterman, R., & Rath, J. (Eds) (2003). *Immigrant Entrepreneurs. Venturing abroad in the age of globalization*. Oxford/New York: Berg/University of New York Press.
- Kloosterman, R., van der Leun, J., & Rath, J. (1999). Mixed embeddedness. (In)formal economic activities and immigrant business in the Netherlands. *International Journal of Urban and Regional Research*, 23(2), 253–267.
- Knecht, M., & Soysal, L. (Eds) (2005). *Plausible Vielfalt: wie der Karneval der Kulturen denkt, lernt und Kultur schafft I*. Berlin: Panama Verlag.
- Kumcu, A. (2001). *De fil en aiguille: Genèse et déclin des ateliers de confections turcs d’Amsterdam*. Amsterdam: Thela Thesis.
- Kunz, P. (2005). *Ethnic entrepreneurs, ethnic precincts and tourism. Missed opportunities from Sydney*. A thesis submitted for the Degree of Doctor of Philosophy. Sydney: University of Technology Sydney.
- Lalich, W. F. (2003). *Ethnic community capital. The development of ethnic social infrastructure in Sydney*. Ph.D. thesis, University of Technology Sydney, Sydney.
- Li, W. (1998). Anatomy of a new ethnic settlement. The Chinese ethnoburb in Los Angeles. *Urban Studies*, 35(3), 479–501.
- Light, I. (1972). *Ethnic enterprise in America. Business and welfare among Chinese, Japanese, and Blacks*. Berkeley, CA: University of California Press.
- Light, I., & Gold, S. J. (2000). *Ethnic economies*. San Diego, CA: Academic Press.
- Lin, J. (1998). *Reconstructing Chinatown. Ethnic enclave, global change*. Minneapolis and Globalization and Community, Vol. 2. London: University of Minnesota Press.
- Logan, J. R., Alba, R., & Jones, B. J. (2003). Enclaves and entrepreneurs. Assessing the payoff for immigrants and minorities. *International Migration Review*, 37(2), 344–388.
- Metcalf, H., Modood, T., & Virdee, S. (1996). *Asian self-employment. The interaction of culture and economics in England*. London: Policy Studies Institute.
- Miller, V. (2004). ‘Mobile Chinatowns: the future of community in a global space of flows’, *Social issues*, 2(1), <http://www.whb.co.uk/socialissues/vol2vm.htm> (accessed on November 22, 2006).
- Morokvasic, M. (1993). Immigrants in garment production in Paris and Berlin. In: I. Light & P. Bhachu (Eds), *Immigration and entrepreneurship. Culture, capital, and ethnic networks* (pp. 75–96). New Brunswick: Transaction Publishers.
- Portes, A., & Jensen, L. (1987). What’s an ethnic enclave? The case for conceptual clarity. A comment. *American Sociological Review*, 52(6), 768–773.

- Portes, A., & Manning, R. (1986). The immigrant enclave: Theory and empirical example. In: S. Olzak & J. Nagel (Eds), *Competitive ethnic relations* (pp. 47–68). New York: Academic Press.
- Ram, M. (1993). *Managing to survive. Working lives in small firms*. Oxford: Blackwell.
- Rath, J. (Ed.) (2000). *Immigrant businesses: The economic, political and social environment*. Houndmills, Basingstoke, Hampshire, and New York: Macmillan Press and St. Martin's Press.
- Rath, J. (Ed.) (2002). *Unravelling the rag trade. Immigrant entrepreneurship in seven world cities*. Oxford: Berg.
- Rath, J. (2005). Feeding the Festive City. Immigrant Entrepreneurs and Tourist Industry. In: E. Guild & J. van Selm (Eds), *International Migration and Security: Opportunities and Challenges* (pp. 238–253). London and New York: Routledge.
- Rath, J. (Ed.) (2006). *Tourism, Ethnic Diversity and the City*. (Contemporary Geographies of Leisure, Tourism and Mobility Series). London and New York: Routledge.
- Sanders, J., & Nee, V. (1987). Limits of ethnic solidarity in the ethnic enclave. *American Sociological Review*, 52(6), 745–767.
- Saxenian, A. (1999). *Silicon valley's new immigrant entrepreneurs*. San Francisco: Public Policy Institute of California.
- Shaw, S., Bagwell, S., & Karmowska, J. (2004). Ethnoscapes as spectacle. Reimaging multicultural districts as new destinations for leisure and tourism consumption. *Urban Studies*, 41(10), 1983–2000.
- Tsu, T. Y. (1999). From ethnic ghetto to “Gourmet Republic”. The changing image of Kobe's Chinatown in modern Japan. *Japanese Studies*, 19(1), 17–32.
- Valle, V. M., & Torres, R. D. (2000). Mexican cuisine. Food as culture. In: V. M. Valle & R. D. Torres (Eds), *Latino metropolis* (pp. 67–99). Minneapolis/London: University of Minnesota Press.
- Van den Berghe, P. L. (1980). Tourism as ethnic relations. A case study of Cuzco, Peru. *Ethnic and Racial Studies*, 3(4), 375–392.
- Waldinger, R. (1993). The ethnic enclave debate revisited. *International Journal of Urban and Regional Research*, 17(3), 428–436.
- Waldinger, R., R. Aldrich, R. Ward and Associates (1990). *Ethnic entrepreneurs. immigrant business in industrial societies*. Newbury Park: Sage.
- Wilson, W. J. (1987). *The truly disadvantaged. The inner city, the underclass and public policy*. Chicago: University of Chicago Press.
- Wong, B. (1998). *Ethnicity and entrepreneurship: The New Chinese immigrants in the San Francisco bay area*. Boston: Allyn and Bacon.
- Wong, K. S. (1995). Chinatown. Conflicting images, contested terrain. *MELUS*, 20(1), 3–15.
- Wood, J. (1997). Vietnamese American place making in Northern Virginia. *Geographical Review*, 87(1), 58–72.
- World Travel and Tourism Council. (2005). *Travel & tourism sowing the seeds of growth: The 2005 travel & tourism economic research: Executive summary*. London: World Travel & Tourism Council.
- Yamashita, K. (2003). Formation and development of Chinatown in Japan. Chinatowns as tourist spots in Yokohama, Kobe and Nagasaki. *Geographical Review of Japan*, 76, 910–923.
- Zelinsky, W., & Lee, B. A. (1998). Heterolocalism. An alternative model of the sociospatial behaviour of immigrant ethnic communities. *International Journal of Population Geography*, 4, 281–298.

- Zhou, M. (1992). *Chinatown. The socioeconomic potential of an urban enclave*. Philadelphia: Temple University Press.
- Zukin, S. (1991). *Landscapes of power. From Detroit to Disney World*. Berkeley, CA: University of California Press.
- Zukin, S. (1995). *The cultures of cities*. Cambridge, MA: Blackwell.
- Zukin, S. (1998). Urban lifestyles. Diversity and standardization in spaces of consumption. *Urban Studies*, 35(5-6), 825–839.
- Zukin, S., Baskerville, R., Greenberg, M., Guthreau, C., Halley, J., Halling, M., Lawler, K., Nerio, R., Stack, R., Vitale, A., & Wissinger, B. (1998). From Coney Island to Las Vegas in the urban imaginary. Discursive practices of growth and decline. *Urban Affairs Review*, 33(5), 627–654.

# FROM VULNERABLE TO VENERATED: THE INSTITUTIONALIZATION OF ACADEMIC ENTREPRENEURSHIP IN THE LIFE SCIENCES

Jeannette A. Colyvas and Walter W. Powell

## ABSTRACT

*We examine the origins, acceptance, and spread of academic entrepreneurship in the biomedical field at Stanford, a university that championed efforts at translating basic science into commercial application. With multiple data sources from 1970 to 2000, we analyze how entrepreneurship became institutionalized, stressing the distinction between factors that promoted such activity and those that sustained it. We address individual attributes, work contexts, and research networks, discerning the multiple influences that supported the commercialization of basic research and contributed to a new academic identity. We demonstrate how entrepreneurship expands from an uncommon undertaking to a venerated practice.*

---

The Sociology of Entrepreneurship

Research in the Sociology of Organizations, Volume 25, 219–259

Copyright © 2007 by Elsevier Ltd.

All rights of reproduction in any form reserved

ISSN: 0733-558X/doi:10.1016/S0733-558X(06)25007-9

## INTRODUCTION

Much contemporary discussion of entrepreneurship celebrates risk-taking individuals, who either by dint of their skill or charisma forge new paths, be these in the form of companies, social movements, or causes. Whether such activity requires the formation of a new organization or involves the reform of an existing one, these efforts are seen as valorous. This present-day veneration often ignores, however, the social context in which entrepreneurship initially occurs. Immigrant entrepreneurs are seen as creating jobs and wealth for their ethnic communities, but the strains and challenges to a stable social order are often downplayed (but see [Portes & Sensenbrenner, 1993](#)). Champions inside companies are viewed as virtuous reformers, but the settled ways of doing things provided comfort, routine, and solidarity to many. We seek to more closely analyze the settings out of which enterprising efforts emerge, and understand the risks involved for those pursuing new paths as well as the possible costs of a change to an existing order. We emphasize that entrepreneurship is rarely a single momentous act, but an incremental process and its reinforcement is very much a social accomplishment. The manner in which efforts at entrepreneurship are supported or contested is critical to its reproduction.

To explore these issues, we examine the origins, acceptance, and spread of academic entrepreneurship in the biomedical field at Stanford, a university that championed efforts at translating basic science into commercial application. We define academic entrepreneurship as the practice of disclosing inventions, filing for patents, or working with biomedical companies. With the use of multiple data sources spanning 1970–2000, we document a slow rise in these endeavors, with considerable discussion and uncertainty surrounding early enterprising efforts, then a contested period, followed by a growing acceptance, much increased activity, and eventual celebration. Our focus is on how entrepreneurial pursuits became institutionalized; hence, we examine the feedback processes that buttress such activity and contribute to a new academic identity. The data afford assessment of how commercially relevant science spreads through a key academic department. Over this period, academic entrepreneurship expands from an uncommon undertaking to become a venerated practice. We focus on individual characteristics, the work context, and the research networks within and across laboratories. Our aim is to examine which faculty are initially engaged in commercialization, which ones persist at such activity, who new entrants invent with, and how the composition of inventors changes over time.

We take a broad view of entrepreneurship, attending not only to the formation of business ventures, as is common in the literature, but also the creation of new organizational identities and practices (Hwang & Powell, 2005). We follow the classic formulation of Schumpeter (1934), who argued that entrepreneurship involved the recombination of existing resources and practices to introduce either novel products, methods of production, sources of supply, markets, or modes of organization (Swedberg, 2000; Fagerberg, 2003). We extend this approach to analyze how scientific identities that combined academic norms with industrial considerations emerged and spread. In our view, academic entrepreneurship was an integration of novel roles and resources into existing organizational contexts, triggering the creation of new models of what a researcher should be doing. This transformation was accompanied by an expansion of university administrative procedures to support these new activities (Colyvas & Powell, 2006). Eventually, academic entrepreneurship became both highly scripted and widely valued. Seen more abstractly, the argument we advance also provides insight into the emergence and development of entrepreneurial activity in other settings where such efforts were once viewed as unconventional.

Our starting point is a discussion of invention and entrepreneurial activity within the context of university-based science. Having chosen a university that has in place many of the elements that researchers identify as necessary to support academic entrepreneurship, we advance arguments that stress the feedback mechanisms that reinforce academic entrepreneurship. The research site, Stanford University, and our data are introduced next. Our analysis highlights the process by which entrepreneurial behavior spread within a basic biological science department, using data on inventors and their attributes, number of disclosures, research expenditures, and revenues. We utilize visualization tools to represent the relational networks that constituted invention disclosure teams. These methods allow us to capture the trajectory of growth of inventive activity. We examine the relationships between scientists and their discovery efforts, analyzing how the practices and meanings associated with entrepreneurship change over time.

## **ENTREPRENEURSHIP IN THE UNIVERSITY CONTEXT**

For much of the last century, the practice of academic patenting was uncommon, especially in the life sciences (Mowery, Nelson, Sampat, & Ziedonis, 2004). The norms and reward system of science did not place a



high value on faculty involvement in commercializing research findings. A 1968 letter to *Science* made the point that publishing was the primary objective of university researchers, and “many academic investigators ... overlook or ignore the invention and patentable results of their work.” (Macy, 1968). The author was speaking from the point of view of the Research Corporation, the foundation that handled most university licensing prior to 1970. As one of the earliest technology licensing professionals, he underscored the importance of understanding academic science, and the objectives of university researchers: “If the objective of the research is new scientific knowledge, or the introduction of students to meaningful investigation, once this objective is reached, the sole remaining step taken by the investigator is to publish” (Macy, 1968).

Gradually, however, universities such as Stanford became more directly involved in marketing basic science discoveries. Such steps were not without controversy. A few high profile cases emerged that challenged both the desirability and appropriateness of commercial activity. “I do not want to sign a letter saying that I was just another laboratory worker,” commented a University of Michigan professor to *Nature* in 1980 when the now famous recombinant DNA patents were under scrutiny (Dickson, 1980, p. 388). The patenting process, which required ‘disclaimers’ on the patenting application from co-authors and other scientists that they were not inventors, contravened scientific conventions of dissemination and credit. One of four authors on the original publication, this professor challenged the idea that his scientific input was marginal: “I was part and parcel of the whole thing; I don’t feel that I should sign something that I do not believe is true” (Dickson, 1980, p. 388). The logic of the patent system conflicted with ideas about the cumulative aspects of scientific discovery, putting scientist-inventors under professional scrutiny.

Much ferment was underway in national policy debates in the late 1970s about the lack of university contribution to industrial competitiveness. In 1980, Federal legislation was passed encouraging universities to facilitate technology transfer. Most notably, the Bayh-Dole Act harmonized institutional patent agreements across federal funding agencies, allowing universities to take title to inventions developed as a result of government sponsored projects. The same year, a milestone Supreme Court ruling, *Diamond versus Chakrabarty*, allowed the patenting of human life forms, while one of the first biotech companies, Genentech, had a hugely successful initial public offering. These developments helped fuel an upsurge in patenting by U.S. universities during the 1980s, led by research in the life sciences (Mowery et al., 2004; Owen-Smith, 2003). But many technology

transfer offices (TTOs) were faced with the harsh reality that successful licenses were few and far between. Most TTOs struggled to break even, and profit was highly concentrated among a few universities.

Nevertheless, the revenues garnered by a handful of universities, including MIT, Stanford, and the University of California system, raised questions about missed opportunities at other schools. A 1989 *Economist* article drew attention to Indiana University's 1957 license to Procter & Gamble that gave rise to Crest. The royalty formula for "one of America's most popular toothpastes" was based on the amount of substance utilized, rather than sales of the final product. "Indiana got about \$4m when it might, by some estimates, have \$100m. Given today's slick marketing, such mistakes are unlikely to be repeated in many ivory towers" (*The Economist*, 1989, p. 82).

As the biotechnology and software industries burgeoned, the 1990s witnessed a continuing expansion of university technology transfer programs. At Stanford, a policy was passed in 1994 that any inventions done with university resources had to be submitted for university review. This policy further integrated academic entrepreneurship into the mission of the university. One indication of the growing acceptance of this goal was the scant protest over this broad claim to intellectual property. Entrepreneurship in the academy became regarded as central to economic growth. In 2002, the Bayh-Dole Act was hyped in *The Economist* as "innovation's golden goose," and as "[p]ossibly the most inspired piece of legislation to be enacted in America over the past half-century..." (*Economist*, 2002, p. 3). By the new century, university involvement in commercializing science was widely supported.

Research on entrepreneurship has largely fallen into two camps – an individual focus that emphasizes the motivations, experiences, and attributes of entrepreneurs, and a structural perspective that underscores the circumstances that afford opportunities or access to resources and environments rich in institutional support. In our interviews with prominent scientists who were active inventors and engaged with successful start-up firms, both themes were echoed. For example, one tenured molecular biologist, who was an early entrepreneur and founded a company, posed the question to us: "What motivates people to study a particular disease? Is it money or personal health?" He went on to note that he has had cancer twice, and had many friends die from AIDS. "Look at my research, it deals with HIV and cancer" (Owen-Smith & Powell, 2001a, p. 123). His comments illustrate how a scientist's personal experiences and relationships influence the direction of his science. Similarly, a senior neuro-immunologist observed that peer effects and monetary rewards had become very entangled: "I think this is an

extraordinary place because so many people in your reference group are running around inventing things...” He went on to note that “it would be hard to ignore how fabulously wealthy some of your peers are. You notice the kinds of cars they park in the lot, and your children interact with their children...” (Powell, Owen-Smith, & Colyvas, 2007). These comments highlight how deeply intertwined social and economic motivations are.

A different rationale was offered by an eminent scientist who developed a prototype research device. “We needed a company to make the [technology] so other people could use it.” He argued that this technology would transform the field by dramatically increasing the speed at which research was performed. The proceeds from making this research tool ubiquitous, however, were not intended to return as personal gain, but instead to feed back into his laboratory and replenish the funding for the research program that generated the discovery (Colyvas, 2006). Thus, both the desire to bring inventions into the world of practice and to generate resources for his laboratory fueled this scientist.

Researchers have identified other individual attributes and motivations that have drawn scientists into the world of commerce. Intellectual capital, measured by scientific productivity (“star scientists”), career stage (tenure), and experience (co-publishing with many authors) have been linked to both the proclivity to patent and involvement in start-up firms (Zucker, Darby, & Brewer, 1998; Thursby & Thursby, 2002; Stuart & Ding, 2006). In addition, a scientist’s expertise, personal experience, and tacit knowledge of an invention may be so considerable that she needs to have a hand in any downstream development, thus leading to involvement in a start-up company (Shane, 2004; Lowe, 2006).

From a more structural perspective, researchers have examined why some disciplines and universities have been more conducive to faculty involvement in commercial activities, stressing the importance of university policies, culture, a supportive technology transfer office, and the presence of a medical center (Etzkowitz, 1998; Owen-Smith & Powell, 2001b, 2003; DiGregorio & Shane, 2003). Lach and Schankerman (2004) highlight the role of university-designed incentives that encourage faculty to reap gains from focusing on downstream applications of their research. Stephan et al. (2005) find that there are notable differences in patenting across academic fields, with biomedicine the most active area, followed by the engineering and physical sciences. They report that patenting activity is highly skewed, and significantly related to number of publications.<sup>1</sup>

Others who study academic entrepreneurship have attended to the wider environment in which universities are situated, noting that the growth of

such activity may be propelled by the desire of universities for more resources and autonomy (Slaughter & Leslie, 1997). More critical analysts contend that entrepreneurship reflects industry's growing embrace of, and influence over, university research (Krinsky, 2003; Washburn, 2005). A rich vein of work has looked at MIT, noting how its unique ecosystem has fostered discovery and linked academic inventors to entrepreneurial firms in multiple forms, ranging from the sharing of research tools to advisory board membership to visiting scientists to company founders (Agrawal & Henderson, 2002; Murray, 2002; Shane, 2004). More broadly, the process of "spawning," that is the role of either commercially engaged universities or large corporations in generating a talent pool of entrepreneurial scientists and managers, has been found to be critical to the founding of new firms (Gompers, Lerner, & Scharfstein, 2003; Higgins, 2004; Bercovitz & Feldman, 2005).

We do not downplay the role of university initiatives or individual motivations and attributes in fostering academic entrepreneurship, but we also explore alternative arguments concerning the structure of laboratory life and the research networks in which scientists are involved. We show that faculty engagement with industry varies over time, as different eras reflect divergent levels of acceptance. The context in which research is conducted helps define the range of appropriate opportunities that individual scientists may pursue. By looking at inventive behavior over a 31-year period, we see how a departmental culture and wider university infrastructure supporting commercial involvement began to emerge. As commercialization became more commonplace and acceptable for academic scientists, the meaning of entrepreneurship took a different form. One particularly important aspect of this process was the changing definition and use of research funds, that evolved as entrepreneurship took hold. We begin with a series of possible explanations for individual participation, turn next to structural factors that speak to collective enrollment, and then address how entrepreneurial activity changes over time as new norms of practice ramify across the department.

### *Individual-Level Factors*

The canonical account of scientific advantage stresses the phenomenon of increasing returns, in which those who have early success are subsequently rewarded. This process of cumulative advantage suggests that certain scientists are better positioned to parlay their work into the commercial realm, and mobilize their resources and contacts to generate science that is of

commercial importance (Merton, 1968; Levin & Stephan, 1991). Thus, faculty who are advantaged in the realm of science will be able to convert that position into the world of technology. These scientists have more resources, technicians, students, and postdocs to advance their research enterprise, publish more frequently, and have a larger corpus of science to draw upon. Hence, *faculty with more research funding should have more opportunities to patent, and will invent more frequently.*

Research support alone may not be the whole story, however. The propensity to patent may depend on contacts that influence how a scientist perceives commercial activity. The normative structure of science in the 1970s and 1980s was not altogether hospitable to commercial involvement (Bok, 1982), and the Mertonian ideals of disinterestedness and skepticism cast a broad influence (Merton, 1973). Thus, scientists may have needed contact or exposure to industry in order to be persuaded of the value of patenting. Scientists with industry funding or who consulted with industry may view commercial involvement differently from those who lacked such contacts, and may have more opportunity to be involved in downstream development of basic science. This exposure to the commercial world should heighten the propensity of these scientists to disclose their inventions. Consequently, *faculty with contacts in industry are more likely to be inventors.*

One common explanation for entrepreneurial efforts involves the incentives that attract scientists to engage with industry (Lach & Shankerman, 2004). Viewed in this light, pecuniary motivations pull scientists into the commercial realm. Thus, *the financial rewards of private science prompt faculty to disclose, and provide positive reinforcement to continue to do so as their research develops.* One might expect that scientists are primarily attracted to commercial science for its financial rewards. Furthermore, *those scientists who are successful in garnering significant licensing revenues are most likely to persist in entrepreneurial engagement.*

### *Work Context*

The organization of modern science is strongly shaped by structural factors, including the context of laboratory life, the career ladder of the academy, and patterns of recruitment and reward within departments and, more broadly, inside the university. Most large-scale research activity in the life sciences occurs in laboratories that involve faculty, postdoctoral fellows, graduate students, and research and technical staff. Seen through this organizational lens, disclosing is rarely a solitary act, but shaped by

membership in a laboratory. Thus, we expect *inventors to be less likely to disclose on their own, and more likely to do so through collaboration with other members of their laboratory.*

Furthermore, much of the career structure of contemporary science is based on seniority, in which authority and rewards accrue as one moves through career stages. More senior scientists mobilize research teams to pursue questions and problems that build a program of research. This career structure influences inventing by linking newcomers to established senior scientists who are directors of laboratories. Hence, we anticipate that *new inventors are more likely to invent as part of a team headed by an experienced inventor, rather than with other inexperienced scientists.*

To the extent that career patterns shape the autonomy and discretion of scientists, one might anticipate that the spread of an activity such as inventing occurs through distinct career stages. For example, expansion may occur as structurally equivalent individuals engage in comparable activities because of peer comparison and competition (Lorrain & White, 1971; Burt, 1987). This argument suggests that *inventors are likely to become enrolled in commercialization as others at a comparable career stage do.* Nevertheless, to the extent that commercial involvement with industry is novel, unfamiliar, and frowned upon, one might expect only those scientists who are secure in their status to participate (Philips & Zuckerman, 2001). Scientists seeking to gain tenure or in the early stages of their career would be reluctant to engage in such an activity. Thus, *when commercialization is new or departs from established practices, those most likely to pursue it should be established senior faculty.*

### *Period Effects*

We have argued that academic entrepreneurship developed gradually, rather than abruptly, and evolved in stages. As commercial engagement became more frequent among scientists, the perception of the activity acquired a different tone. Commercializing research results became regarded as legitimate, the activity became institutionalized, and the reputations of engaged scientists were enhanced rather than threatened. In previous work, based on a close reading of Office of Technology Licensing (OTL) archives, we established three time periods that take into account both the institutionalization of technology transfer at Stanford and the larger federal policy changes buttressing the commercialization of science (Colyvas & Powell, 2006). In the early years, from 1970 to 1980, venturing into the unfamiliar

territory of commercial engagement involved risks, notably to one's academic reputation, as entrepreneurship was perceived as possibly eclipsing one's duties as a faculty scientist. Such activity ran counter to many of the norms of science. Decisions to commercialize were characterized largely as an exception rather than the norm. In the middle period, 1981–1993, commercial involvement became more accepted, though it was still a subject of debate and contention. While universities were afforded the legitimacy to transfer technologies, concerns over conflict of interest for individual faculty were amplified as many forms of science were being patented for the first time. By the mid-1990s, the marketization of science became not only routine for both the university and scientists, but was celebrated as a marker of success. Accordingly, our third period runs from 1994 to 2000.

Consequently, we expect that, *in the earliest period, accomplished, high status scientists are more likely to disclose inventions*. These senior scientists have earned their spurs in the world of scientific competition and have well-established laboratories. They are thus less vulnerable to charges that their work has been tainted by involvement with industry or concerned about promotion. In period two, *as commercial involvement became more legitimate, we expect inventive activity to spread first to other senior faculty*. Having obtained tenure, more senior scientists will be susceptible to pursuing commercial opportunities or open to university requests to fulfill national mandates to transfer technologies for public use and benefit. In period three, *we expect commercial involvement to permeate into earlier, pre-tenure career stages*. As controversial cases are adjudicated and success is garnered by others without damage to reputation, the unfettered ability to pursue commercial endeavors will attract early career scientists. Entrepreneurship becomes an identity in the academic context in which doing business with industry signals acumen and success.

High-status scientists, however, are not the only individuals susceptible to commercialization. As mentioned earlier, the research process is highly collaborative and the structure of academic science involves students and postdocs. The extent to which one may participate in commercialization is contingent on research networks and laboratory membership. Students and postdocs are not always in a position to pursue or resist patenting independent of their faculty supervisors. We expect *students and postdocs to be involved in patenting only through the collaboration with high-status faculty*. As more senior faculty become involved in period two, postdocs and students will be more likely to disclose as well. In period three, *we argue that early career scientists will be more likely to disclose independently of a high status collaborator*.

## TECHNOLOGY TRANSFER AT STANFORD UNIVERSITY

Stanford is an auspicious site to observe the emergence of technology transfer as entrepreneurial efforts occurred when commercializing academic science was both new to the university and to the wider setting of the academy. Stanford began a technology transfer program in the summer of 1968, well before federal legislation in the early 1980s mandated such efforts (Colyvas, 2007). Stanford subsequently became one of the most successful technology transfer offices, frequently touted as a model for emulation by many U.S. and foreign universities.

Our focus is on the life sciences where commercial involvement was new. In the late 1970s and early 1980s, the biotechnology field was just emerging. In addition the scientific status of this discipline underwent a transformation in the 1980s, opening up novel opportunities for collaborations with researchers in other basic and clinical fields. Our time periods afford us the ability to observe the early to late stages of an important change, as entrepreneurship developed and became commonplace to university science.

Our case is a single department in the Medical School, albeit among the best funded and most prestigious. Most senior faculty were members of the National Academy of Sciences. It was among the earliest basic science departments in the Medical School, founded in the 1950s by a Nobel Laureate charged with the task of building a first rate basic science program. While small in terms of number of faculty, it was considerably better funded than other basic science departments in the Medical School. The department remained relatively small through the 1980s, making mostly senior hires. In the 1990s, with the appointment of a new chair, it grew quickly, adding junior positions. Even though a small unit, the inventive activity of this department is significant. Over this 31-year period, its members produced 130 patents. While the number of disclosures may seem modest, consider the activity relative to the small size of this department and in comparison to other universities. The level of activity at Duke and Johns Hopkins Universities in the biomedical field in the 1990s was no greater than at Stanford in the 1970s (Bercovitz & Feldman, 2005).

### *Data Sources*

We utilize data derived from electronic and archival sources at the university, supplemented through proprietary search databases, and the World



Wide Web.<sup>2</sup> We focus on the activities of individuals associated with this basic life science department, as well as any co-inventors from other departments or outside the university. The criterion for being considered an inventor is whether an individual has submitted an invention disclosure to the Stanford OTL for consideration to be patented or licensed. In the sample, there are 179 disclosures, or inventions, between 1970 and 2000, and 198 individual inventors. There are 474 incidences of inventing, counting inventors for each individual act of disclosing. Clearly, most disclosures have multiple inventors. For purposes of constructing the sample, we include only those inventions attributed to an inventor while he or she was affiliated with this department.<sup>3</sup> Disclosures by inventors when they held appointments in other Stanford departments (e.g. a faculty member who had been a graduate student or professor in another department) are not included in the sample, nor are inventions by faculty when they were at other universities or working outside the university. We include in our analysis any prior record of disclosing as an indicator of an individual's previous experience.<sup>4</sup>

We coded the attributes of individuals, such as affiliation and job title, at the time of each disclosure, and recorded the revenues generated by each invention that is licensed.<sup>5</sup> As a check for inventive activity that may not have been captured through disclosures, we searched the names of inventors and non-inventors in our sample through the United States Patent and Trademark Office database. We obtained assignment information for each patent, noting whether it was owned by Stanford, an outside academic institution, or a company. While some faculty who disclosed at Stanford also held patents with other organizations, we identified only one case of a Stanford non-inventor with a patent from outside of the university; this occurred when he was a graduate student at another university. In addition, we matched our names against a dataset of biotechnology founders in the Silicon Valley and Boston regions between 1969 and 2002 (Porter, 2004). None of our non-inventors appear as founders, and no new incidents of entrepreneurship among our inventor sample were identified. Finally, to account for the size of the research laboratories and source of support, we collected annual data on faculty members' sponsored and contract research expenditures, collected through the Stanford Office of Sponsored Research. In return for access to these data, we agreed to provide anonymity to the department and its members.

This rich dataset has limitations, however, that reflect how complicated it is to construct a sample of all individuals who engage in entrepreneurial activity. We have rather complete data on those individuals who disclosed inventions. For inventors who are faculty members of the department, we can contrast them to the full population of faculty in the department, which

we were able to construct with university records. We do not have, however, comparable data on the full departmental population of students, technicians, postdoctoral fellows, or other scientific staff, as no complete records exist of who was present at the time. Thus, we know a good deal about those who filed disclosures, have less information about faculty and students in the department who never disclosed, and very little information on other staff in the department who did not invent. When we turn to outside collaborators – whether within the university or in industry – we lack comparable data. Looking at a 31-year period allows us to see how inventive activity changes over time, but adds in the complications of students and staff who depart, faculty who are hired, promoted, or change affiliations. In short, the statuses of our inventors are not constant, and we have to make judgment calls as to how such changes influence commercial engagement.

Because we have a single influential department and a relatively small number of inventors, we present our data in the form of a case rather than with inferential statistics. We do not adjudicate, for example, as to whether a scientist with considerable research funds is more likely to engage with industry on her own, or be pulled into involvement by industrial counterparts that want to exploit her research. Instead, we draw on the archives, specifics of inventions, and departmental context to supplement the data and address the expectations derived from the literature.

## FINDINGS

In broad strokes, entrepreneurship, measured by the number of disclosures and number of inventors, increased over time, albeit in fits and starts. [Figs. 1\(a\) and \(b\)](#) track inventors and disclosures over the period 1970–2000, charting both number and cumulative growth. These figures illustrate the scale of inventing within this department. [Fig. 1\(a\)](#) shows that the overall shape of inventive activity is somewhat bimodal, with an early bump from 1978 to 1982, a subsequent decline, then an increase in 1993 and 1994, with rapid expansion in the late 1990s. The number of inventors exceeds the number of disclosures, especially so in the 1990s, reflecting either multiple authors appearing on an invention or a single prolific inventor with numerous disclosures.

Within the academy, publishing is highly skewed, as a small number of scientists contribute a disproportionate share of the output. It is not surprising then that inventing, especially in the early years, is similarly concentrated. There are relatively few inventors initially, and not until the mid-1990s does the number of inventors in the department exceed 20 annually. [Fig. 1\(b\)](#) plots

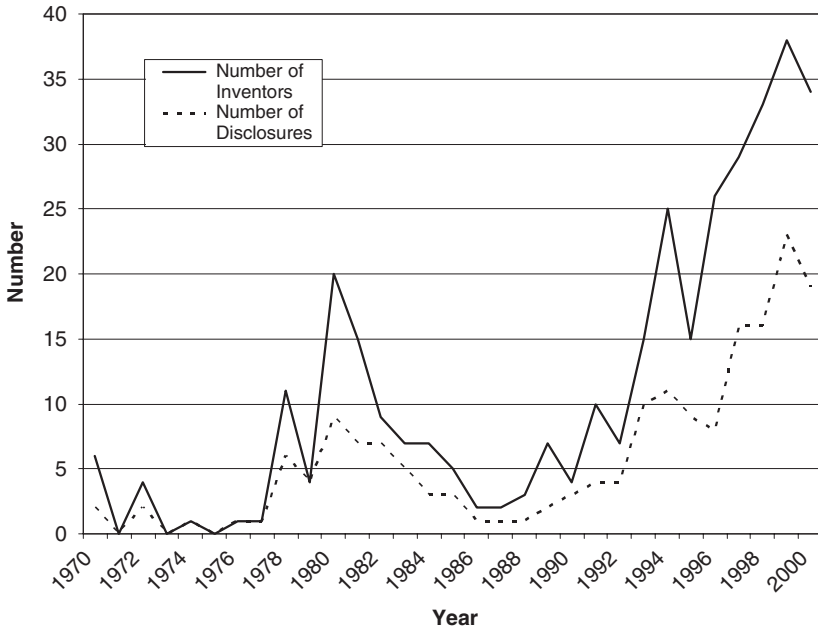


Fig. 1(a). Number of Inventors and Disclosures, 1970–2000.

the cumulative growth of disclosures and inventors. We see that the number of disclosures grew in the early 1980s, then declined through the end of the decade, and rose sharply throughout the 1990s, while the number of inventors mushroomed as well. These figures show that, overall, inventive activity is expanding, and there are distinct stages of development.

One common explanation for academic entrepreneurship relates to the federal policy changes that took place in the 1980s. These changes might be expected to be associated with an increase in disclosure activity. Our data, however, reflect a different trend – a decline through the 1980s, precisely in the wake of legislation that sought to encourage such efforts. This disparity underscores an important distinction between the socio-political form of legitimacy afforded to universities at this time by legislation, compared to a lag in cultural-cognitive legitimacy among individual scientists (Aldrich & Fiol, 1994; Colyvas & Powell, 2006). While technology transfer became a justifiable and politically approved activity, the convention had not yet taken hold in the life sciences community and was still subject to considerable debate and contestation.

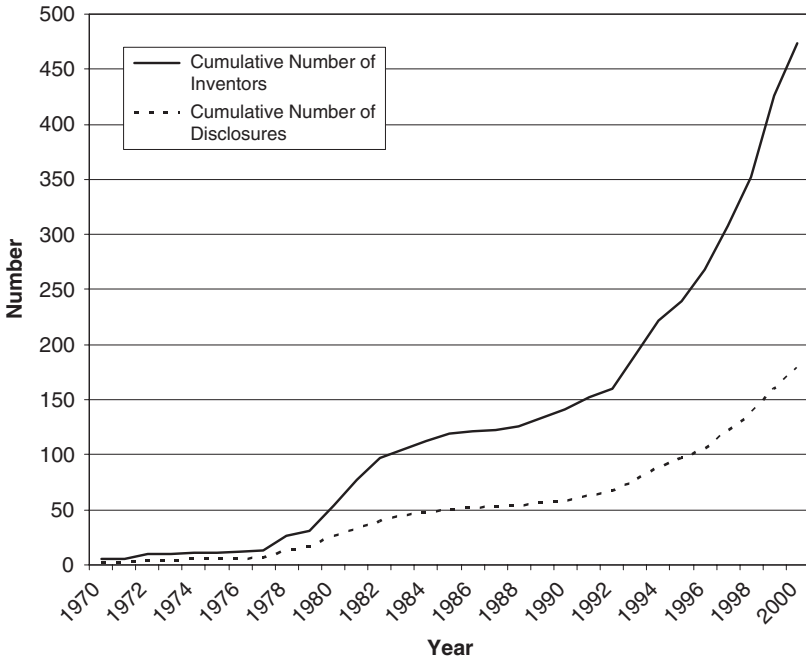


Fig. 1(b). Cumulative Number of Inventors and Disclosures, 1970–2000.

The archival records from the 1980s are replete with examples of such ferment. Although a few professors wanted to become more involved with faculty start-up companies, the rules about conflicts of interest were still being worked out. At Stanford, and other universities, questions about whether faculty could have equity in a start-up, and how much time could be devoted to a company, prompted much discussion between the higher administration, the university’s legal office, and entrepreneurially inclined faculty. Faculty who submitted an invention disclosure and then wished to obtain an exclusive license for their own company were inhibited or required approval from the higher administration. There was some concern that enterprising faculty might be circumventing the university and commercializing their findings elsewhere. Indeed, a close examination of the disclosure data finds that one faculty member who disclosed an invention during this period left soon after to start a company. Another faculty inventor in the department temporarily ceased disclosing inventions to the university in the mid to late 1980s, as issues of conflict of interest were being sorted out.

**Table 1.** Rank at Time of Invention 1970–2000.

	1970–1980	1981–1993	1994–2000	All Years
Faculty	31%	44%	48%	45%
Postdocs or fellows	13%	12%	10%	11%
Students	6%	15%	20%	17%
Scientific or technical personnel	46%	19%	11%	18%
Scientists at other universities	4%	8%	9%	8%
Scientists at companies	0%	2%	2%	2%
Number of individual cases of disclosing	54	137	283	474

This eminent scientist did not, however, apply for patents on his own or with a company during this hiatus. As policies became established in the early 1990s and the campus patenting policy in 1994 was settled, we see broader acceptance of commercial involvement, with more new inventors and repeat disclosing activity by prior inventors, signaling cultural-cognitive legitimacy.

This shift in acceptance is reflected in the numbers in [Table 1](#), which lists inventor rank at the time of a disclosure. We present the data by time period, and the period effects are notable. There are 54 cases of individuals disclosing from 1970 to 1980 (including multiple inventors for each invention and repeat activity of the same inventors), 137 from 1981 to 1993, and 283 in the short seven-year window of 1994–2000. This table captures several key trends. Note that from 1970 to 1980, inventing was primarily done by scientific or technical staff. Less than a third of the inventions listed faculty as inventors. In period two, we see a marked change, as technicians are no longer commonly listed on disclosures, and faculty participation rises. Student involvement more than doubles. Both faculty and student engagement rise again in period three, as technicians' participation continues to decline. Thus, as disclosing inventions expand, it permeates into the academic ranks and travels down the career ladder. Clearly, inventive efforts are no longer the province of the technical staff and become much more common among faculty, and some students.

Interestingly, there is some collaboration with scientists outside the university, but primarily those at other campuses, not with scientists at companies. There is no evidence in the archival records that industry science pulled faculty into inventing. In contrast, as we shall see below, there is much more of a push factor, as a small number of faculty assigned their inventions to companies that they were closely involved in consulting with or founding.

*Attributes*

We turn now to look at the characteristics of faculty in the department. The work context in the life sciences is a laboratory, headed by a faculty member. We present the data by periods, as we find this partitioning most indicative of the changing patterns of involvement. One set of explanations for entrepreneurship stresses the characteristics of scientists, notably their resources, experience, and contact with industry. Table 2 is the first of three tables that provide insight into how these factors influence disclosing. Faculty are represented by letters to preserve anonymity, and arrayed chronologically by the year they joined the department. We list the number of disclosures they make in each time period, their cumulative number of disclosures, and annual research expenditures of each faculty member. Note that average expenditures were calculated by the sum of all contracts and research grants attributable to each individual, divided by the total number of years in which a faculty member had research awards. In some cases, faculty were Principal Investigators (PIs) for research grants that were administered in other units in the university and these data were included in the total calculations. We also ascertained the amount of funding faculty received from industry. We consider the annual research expenditures as an indicator of the scale and prestige of a faculty member's research program, and corporate funding as one measure of contact with industry.

We also list the number of disclosures that were successfully licensed, and the gross revenues generated by these disclosures. The latter sum is an indicator of the commercial success of an invention, and certainly a measure of a faculty member's status as an inventor. But this total figure is not indicative of a faculty member's monetary share. The university takes a 15% cut, and then royalties are split by thirds among school, department, and inventor.<sup>6</sup> With multiple inventors, the share is further subdivided.

Only three faculty out of nine who were in the department over this entire decade were involved in disclosing. Two prominent senior faculty are responsible for the lion's share of the activity. This table does not list students, postdocs, or technicians, but we know that the non-faculty staff that were active in disclosing during this period came largely from these two laboratories. Disclosures by faculty did not occur until fairly late in the 1970s, and only five of the 11 faculty disclosures were successfully licensed by the OTL to companies. With respect to arguments about propensity to disclose, there are several key considerations, including funding, industry contact, and reward. We take up each in turn.

**Table 2.** Invention Disclosure Activity, Research Expenditures, and Returns to Licensing, 1970–1980.

Faculty	Year Joined Department	Number of Disclosures/ Cumulative Total	Year of First Disclosure in Department	Average Annual Research Expenditures	Year of First Industry Funding	Total Experience in Industry Sources	Number of Disclosures Licensed with Income	Gross Revenues from Disclosures <sup>a</sup>
Professor A	1954			\$1,272,793	1975	\$9,087		
Professor B	1960	4/4	1978	\$702,929			2	\$23,064
Professor C	1963			\$142,224				
Assoc. Prof. D	1966			\$93,741				
Professor E	1972			\$197,240				
Assoc. Prof. F	1974			\$52,603				
Asst. Prof. G	1977	1/1	1980	\$173,761				
Professor H <sup>b</sup>	1977	6/8	1978	\$729,047	1977	\$24,333	3	\$106,500
Asst. Prof. I	1980							

<sup>a</sup>These are total dollars received by the university, which is shared among the school, department, and inventor(s) after a 15% overhead is deducted. Note that the inventor 1/3 is split among the total number of inventors. Recall that solo inventors are infrequent (15%).

<sup>b</sup>This faculty member had two disclosures prior to joining the department, the first in 1973.

In this early period, Professors A, B, and H had sizeable research budgets. Professor A, the department chair, had the largest, not surprising given his tenure and stature as a Nobel Laureate. But he never invented. Professors B and H had generous research support, and their laboratories became the initial centers of enterprising efforts in the department. Large research budgets allowed faculty to have more technical staff, support more postdocs and graduate students, and expand the scope and intensity of their research programs. Neither faculty member earned much money during this period, so pecuniary incentives had little force. One might argue that the anticipation of income may have been a factor in motivating faculty to disclose. This claim is weak considering that few faculty disclosed during this period and that the market potential for biological invention had not yet been demonstrated at the university. Professor A's disclosure was one of the first from a life science department and the industry scientist who was involved in the work lamented that there was no current market for the technology. Similarly, Professor H initially declined to disclose his basic research findings. Only after the considerable solicitation by the OTL did this scientist agree, persuaded by the argument that doing so would accelerate application into biomedical therapies and protect the technology from being privatized by industry.

Industry funding does not appear to carry much influence, either. Professor A has some modest funds but does not disclose, and Professor B discloses without any industry contacts. There are, to be sure, other kinds of contact with industry, such as involvement with start-up companies, serving on scientific advisory boards, and consulting relations. The archival records of the OTL point to a sharp distinction drawn between consulting with companies and licensing scientific results. In the 1970s, the former was common, while the latter was unusual (Colyvas & Powell, 2006, 329–37). The department chair, Professor A, consulted regularly with companies as early as 1970, and was receptive to contact with industry. The first two disclosures in the department, made by technicians in 1970, one for a chemical synthesis of a hormone and the other for a device, were both submitted with the encouragement of the department head. Indeed, the leadership of the department was very entrepreneurial. The founder of the department was among the first to bring artificial intelligence and computing to biomedical research, developing a famous venture with a faculty member in the computer science department in the mid-1960s that was commercialized. The next Chair, Professor H, took the reins in 1978 and he had developed a software consumer product in the early 1970s related to biomedicine, that involved the founding of a company to distribute and



provide technical support. Neither effort involved patents, though the latter was administered through the OTL.

Consequently, the evidence on industrial influence is mixed. Based on the relatively scant corporate research support, we do not see signs that disclosing was strongly influenced by industry funding. Nor is there any indication for the faculty that consulting with industry led them to disclose their research findings, as neither Professor B nor H had such a relationship at the time of their first disclosures. None of these faculty appear to be involved in biotechnology ventures during this time, although faculty H's invention drew the attention of a local start-up that hired him to consult, and to which he assigned one patent toward the latter part of the period. But prior contact with industry does appear consequential at the department level, as the earliest chair of this department consulted with local biotechnology companies even though he never patented nor disclosed an invention. The archives and interviews with his contemporaries suggest that Professor A perceived involvement with industry and other disciplines as beneficial to the expansion of science and was credited with making early exceptions to commercializing Professor H's technology possible. While department chair, Professor A recruited Professor H to the department, and his arrival added an established inventor and consultant to the ranks.

We turn to the second period, 1981–1992, when licensing the results of academic research was now encouraged, indeed mandated, by federal legislation. As we mentioned earlier, these policy changes had little immediate impact on this department at Stanford. There are several possible reasons for this. One, Stanford was a 'first-mover' and some faculty were engaged in enterprising efforts well in advance of the federal law. Two, the National Science Foundation and National Institutes of Health had developed institutional patent agreements already in the early 1970s with Stanford and several other universities, making it easy for the OTL to obtain title to life science inventions. Thus, the new legislation changed little procedurally for the life sciences. Three, university restrictions on investing in or licensing to faculty-owned start-ups discouraged the most active form of entrepreneurship. Finally, with respect to this department, the unit remained rather small until 1989, when 11 new faculty members were added rapidly over a four-year period. Among the 11 were two senior faculty who were already prolific inventors.

One policy event, however, may have had some significance when examining the types of technologies being disclosed. Breakthroughs in biological materials such as monoclonal antibodies, coupled with the *Diamond versus Chakrabarty* Supreme Court ruling, opened up a range of scientific artifacts

that could be patented and commercialized. The university became enmeshed in a whole new set of debates around “tangible research property,” and whether biological materials required licenses when distributed to other researchers. One faculty member argued such materials should not be patented and personal income from their dissemination was inappropriate. Others expressed anger that industry was profiting from federally funded and university-derived research (Colyvas, 2007). The policy was settled in the early 1980s, leaving open the question of whether or not one chooses to patent. While biological materials for research purposes would be openly disseminated, licenses would be required for commercial use. As a result, the number of biological materials being disclosed increased markedly from the previous period, as did the frequency of inventing within individual research programs.

This period was significant for the department. Not only did it expand in size, but two notable inventions returned considerable revenues to the department. As is clear from even a cursory look at [Table 3](#), many more faculty were disclosing research results, research funding to the department increased, and revenues from licensing were flowing into the unit and to a small number of senior faculty. We take up these changes in turn.

In the previous period there were two active inventors, both continue their enterprising efforts in the 1980s. Professor B has 17 disclosures between 1981 and 1993, and Professor H has 10. They are joined in 1989 by Professor C who discloses nine times in three years, and Professor O in 1991 who has eight disclosures in 1993 alone. There is now a quartet of senior professors who are very entrepreneurial. Their influence is considerable as other senior faculty disclose for the first time during this period, and several younger faculty get involved as well.

Research budgets mushroomed in the 1980s, as six faculty had annual research expenditures in excess of \$1 million. Nevertheless, two of these faculty – Professors N and T – did not disclose, and Professor E did so only once. Industry funding remained rare, but Professor B worked with a company that was developing his research tool and this contact results in corporate support to his laboratory. All faculty continued to assign their patents to the university, save for Professor H, who assigned six of his nine patents to a local biotech company where he consulted, and Professors N and O who were newcomers to the department and brought with them six and eight previous industry patents.

As in the previous era, contact with industry did not appear to predict first-time engagement in disclosing, but was important at the leadership level and reinforced the appointment of entrepreneurial faculty to the

**Table 3.** Invention Disclosure Activity, Research Expenditures, and Returns to Licensing, 1981–1993.

Faculty	Year Joined Department	Number of Disclosures/ Cumulative Total	Year of First Disclosure in Department	Average Annual Research Expenditures	Year of First Industry Funding	Total Experience in Industry Sources	Number of Disclosures Licensed with Income	Gross Revenues from Disclosures <sup>a</sup>
Professor B	1960	17/21	1978	\$1,277,058	1982	\$259,773	8	\$1,390,775
Professor D	1966			\$122,102				
Professor E	1972	1/1	1989	\$1,025,532				
Professor F	1974			\$104,512				
Asst. Prof. G	1977	1/1	1980	\$137,048				
Professor H	1977	10/18	1978	\$1,026,267	1977		8	\$73,721,969
Asst. Prof. I	1980	1/1	1982	\$147,624				
Asst. Prof. J	1982	1/1	1984	\$268,671				
Professor K	1989	2/2	1992	\$378,996			1	\$10,000
Professor L	1989	9/10	1990	\$305,141			1	\$295,500
Assoc. Prof. M	1990		1990	\$528,672				
Professor N <sup>b</sup>	1990	0/2		\$1,002,190			1	\$110,000
Professor O	1991	8/10	1993	\$1,769,314			6	\$942,551
Assoc. Prof. P	1991			\$249,224				
Professor Q	1991			\$503,707				
Professor R	1992			\$460,142				
Professor S	1993							
Assoc. Prof. T	1993			\$1,668,133				
Assoc. Prof. U <sup>b</sup>	1993	0/2					1	\$723,173

<sup>a</sup>This faculty member had two disclosures as a student in a different department in 1986 and 1987.

<sup>b</sup>This faculty member had two disclosures in 1979 and 1980 prior to joining the department.

department. Professor N arrived during this period to serve as chair of the department after several years as a chief scientist at a biotechnology company. Professor O, his collaborator, joins from another Stanford department. Again, the department Chair's proclivity toward industry is neither reflected in disclosures to the university nor start-up activity as he never founded a company. Like Professor H, Professor O continued to disclose to both the university and patent in industry, suggesting the ability of some faculty to create boundaries between their university and industry work.

Within the department, there were numerous successful licenses – 28 are assigned to companies and generate revenues. This is a remarkable development in several respects. One, financial reward did not flow back into the department or to the inventors quickly. With the most lucrative licenses, those from Professors B and H, it took five years after disclosure before any revenues were received. This lag did not stop them from continuing to disclose, however. Two, this record of success is unusual because technology transfer is so highly uncertain. Consider that of the 179 technologies disclosed in this department, only 30% generated any revenues by 2000. Of those 54, 13 earned more than \$100,000 and only four more than a million. Two of those four were licensed in period two. The department had the good fortune of having two early blockbuster successes, accomplishments that many other departments and universities have never had. For example, in 2000, Stanford earned \$41.2 million in gross revenues from 371 inventions. Only 47 generated revenues over \$100,000, and but seven of these earned more than one million (Stanford OTL, FY2000–2001). This department, then, was very much stamped by the early successes of Professors B and H, and the arrival of new faculty who proved to be enterprising and successful as well (Stanford, 2001).

It is important, however, not to overstate the pecuniary side of this success or underemphasize the resource aspect. Professor B did not accept licensing revenues personally, asking that money be signed over to his laboratory. Professor H initially donated his share back to a research and training fund, also declining any personal gain from his invention. The amount reported for gross revenues in our table is the total received by the university. Recall that after the university takes a 15% cut, the shares are divided equally among the school, department, and inventors. So a very sizable sum of money flowed back to the Medical School and the department. In turn, the department was able to expand by adding new faculty, and laboratories grew much larger too. Here, we see the process of cumulative advantage at work, as both faculty and the department have considerable resources to draw on to advance research programs.

Turning to the third period, 1994–2000, (Table 4) the high-profile faculty that joined in the early 1990s with established track records of engagement with the biotech industry proved to be consequential. Professors O and N both had disclosures at the university as well as patents with external companies, although Professor N does not disclose to the university for 8 years. The spread of entrepreneurship was thus shaped by these new entrants, as well as adoption by incumbents. Both processes are clearly influential. Every faculty member present during this period that was hired between 1960 and 1993 disclosed by 2000. Of those hired between 1995 and 2000, four disclosed during this time period. We see that disclosing inventions has become a routine activity, with Professors B, J, L, N, O, Q, and AA especially active. The ramifications of this activity are apparent in several ways. Nine faculty had annual research budgets exceeding \$800,000 annually, and nine faculty had licenses that brought in more than \$100,000 to the university. The department and school garnered significant income as faculty invention portfolios, such as Professor H's, generated considerable commercial interest.

Two other changes occurred during this recent period. One, entrepreneurial activity is no longer the province of either esteemed professors or non-tenure track scientists. Associate professors and assistant professors became involved, some with considerable success. Here we see again how entrepreneurial activity permeates down the ranks, involving a greater number of faculty at all levels. Two, involvement with industry became more common, and new hires were much more likely to have those contacts. Moreover, much of the contact came in the late 1990s. While only two faculty – Professors O and S – garnered significant industry funding, collaboration with biotechnology companies clearly increased. Professor O's contract with a biotechnology firm was for a project funded by a government agency. Professor S's contract came from an industry–government–academic consortium that coordinated a multi-organizational discovery effort. Consider also that the research results of this enterprise were made available for public dissemination, and the patenting of the results was restricted. Thus, an important transformation began to take place in which industry, government, and the academy were increasingly interlinked at the frontiers of science (Powell & Owen-Smith, 1998; Vallas & Kleinman, 2006). Rather than industry pulling faculty into the world of commerce, it appears industry and government were drawn to the fundamental science conducted in the department.

We supplement the data on faculty funding and licensing with a look at the first-time experiences of inventors. In the database, there are 78 incidences of first invention, all other disclosures are by repeat inventors. We have emphasized how much biological research takes place in the context of

**Table 4.** Invention Disclosure Activity, Research Expenditures, and Returns to Licensing, 1994–2000.

Faculty	Year Joined Department	Number of Disclosures/ Cumulative Total	Year of First Disclosure in Department	Average Annual Research Expenditures	Year of First Industry Funding	Total Experience in Industry Sources	Number of Disclosures Licensed with Income	Gross Revenues from Disclosures
Professor B	1960	14/35	1978	\$1,649,630	1982		14	\$8,108,002
Professor H	1977	2/20	1978	\$1,091,632	1977	\$160,640	5	\$182,210,248
Assoc. Prof. J	1982	12/14	1982	\$270,328	1996	\$10,000	5	\$493,725
Professor K	1989	4/6	1992	\$217,438			1	\$30,612
Professor L <sup>a</sup>	1989	14/24	1989	\$644,284			8	\$250,000
Professor N <sup>a</sup>	1990	8/10	1998	\$2,692,489	1999	\$326,498	4	\$112,495
Professor O <sup>b</sup>	1991	15/25	1993	\$9,648,716	1994	\$926,589	9	\$927,905
Assoc. Prof. P	1991	1/1	1997	\$300,289			1	\$28,500
Professor Q	1991	9/9	1994	\$126,892			4	\$136,000
Professor R	1992	4/4	1994	\$907,094	1995	\$226,445		
Professor S	1993	3/3	1995	\$2,298,594	1999	\$5,322,175	2	\$47,500
Professor T	1993	3/3	1996	\$6,599,653			1	\$47,500
Asst. Prof. U	1993	1/3	1998	\$308,858			1	\$878,250
Asst. Prof. V	1995	5/5	1995	\$147,776			3	\$103,000
Professor W	1995			\$807,072				
Asst. Prof. X	1998			\$188,692	2000	\$2,678		
Asst. Prof. Y	1998	1/1	2000	\$29,754				
Asst. Prof. Z	1998			\$133,993				
Assoc. Prof. AA	1998	13/13	1998	\$885,484	1999	\$323,199	6	\$41,073
Assoc. Prof. AB	1999	1/1	2000	\$645,666	1999	\$246,300		
Asst. Prof. AC	1999			\$24,233				
Asst. Prof. AD	1999			\$205,314				
Asst. Prof. AE	1999			\$224,950				
Asst. Prof. AF	2000							
Asst. Prof. AG	2000			\$201,461				

<sup>a</sup>These faculty members had two disclosures prior to joining the department.

<sup>b</sup>This faculty member had 7 disclosures prior to joining the department.

**Table 5.** First-Time Inventors, 1970–2000<sup>a</sup>.

Characteristics	Frequency (%)
Solo inventor <sup>b</sup>	15
With co-inventor	85
With laboratory PI	56
With experienced co-inventor (two or more prior disclosures)	41
With financially successful co-inventor <sup>c</sup>	31
With higher status co-inventor	58

<sup>a</sup>There are 78 first-time inventors.

<sup>b</sup>Fifty percent of the solo first-time inventors were laboratory PIs.

<sup>c</sup>Defined as had previously earned \$50,000 or more annually as an inventor.

research teams. Hicks and Katz (1996) have shown that publishing now routinely involves multiple authors, and inventorship has some parallels. We see from Table 5 that 85% of new inventors were introduced to disclosing by a colleague, while only 15% were solo inventors the first-time they disclosed. And of those solo inventors, half were the PIs of laboratories. This simple breakdown reinforces two points: one, academic entrepreneurship is seldom a solitary activity, and two, for those who do go it alone, seniority and reputation are critical credentials. We also suggested that newcomers to disclosing were more likely to invent with senior scientists, either the PI in charge of a laboratory or a more veteran scientist with prior experience with disclosing. This expectation that new inventors are introduced to entrepreneurship by more senior scientists is clearly borne out, with 56% disclosing for the first time with the PI of their laboratory, and 41% inventing with a veteran who had previously filed two or more disclosures.

There is less support, however, for the idea that newcomers turn not just to experienced inventors, but to those who have been financially successful. We coded revenue data and noted the point at which an individual received \$50,000 or more in an inventor share from an invention. All subsequent collaborations after the year in which that sum was received were coded as co-inventions with a successful inventor. Only 31% of first-time inventors collaborated with scientists who had derived significant financial gain from their research. Much more salient is academic status. There is a pronounced trend for first-time inventors to attach to more high-status individuals. Fifty eight percent of the new inventors worked with scientists who were higher in rank, suggesting that younger, less experienced scientists were introduced to entrepreneurial efforts by more accomplished mentors, as inventing spreads to those more junior ranks in the academic hierarchy via sponsored

mobility. Thus, socialization and sponsorship are the primary mechanisms by which younger scientists are introduced to entrepreneurship.

Overall, presenting the information on faculty funding and disclosing in three distinct periods illuminates several key trends. In the early period, 1970–1980, the most frequent inventors were scientific and technical personnel. Indeed, while Professor B, who had one of the most active laboratories, was the senior author on all publications, he regarded it inappropriate to consider himself an “inventor,” and reserved that status for his skilled technicians and engineers who developed the research tools (Colyvas, 2007). But the listing of technicians as inventors declines markedly, and faculty involvement increases, especially in the third period. By 2000, 17 of the 25 faculty in the department had disclosed (and examination of more recent records shows that three of the new assistant professors who joined in 1999 and 2000 had disclosed by 2002).

Faculty became much more knowledgeable about the opportunities that commercial activity posed for them. Consider this correspondence in the mid-1990s when a faculty member weighs options over whether to pursue his discovery through the university or on his own: “[Co-inventor] and I had decided to go 50-50 if we included both institutions... With Stanford in, [we] will see about 10 percent of the total.” These professors are not just familiar with patenting their science, but demonstrate acumen in considering the complications of involving other institutions. Experienced scientists had come to understand the complexities of working with multiple universities in addition to licensing to companies, and the ins and outs of brokering commercial ventures. “We both recognize that we could be talking about percentages of nothing, but I don’t think so. We probably should be making a proper deal in the first place.”

By the late 1990s, as more research programs were directed by scientists with both experience and success at inventing, the message conveyed to new members of research teams and new faculty hires was that commercial activity was an appropriate complement to basic science. As entrepreneurship spread among senior participants and generated ample returns, those rewards were no longer viewed as exceptional but as components of routine professorial activity. Moreover, the revenues not only enhanced the inventors’ financial circumstances, but greatly expanded the departmental budget, allowing more faculty and staff to be hired, and students to be funded, thus generating a new better-funded regime of knowledge production. Indeed, part of Professor N’s agreement to leave a successful biotechnology company and assume the chair position included additional billets to hire junior faculty conducting cutting-edge research. A member of the National



Academy of Sciences and experienced in industry and academic science, Professor N returned to the academy to oversee the department's expansion. By the late 1990s, then, entrepreneurial efforts not only became conventional, they generated resources that were mobilized to bring in more talent to the department.

The 1990s also saw the frequent formation of academic-led biotechnology companies. A good deal of the scholarly attention of entrepreneurship researchers has focused on start-ups (Shane, 2004), and the biotechnology industry is well known for the considerable involvement of research faculty (Porter, 2004; Powell, 1996; Zucker et al., 1998). Our focus has been more on the antecedent activities that may eventually culminate in founding a company. In this department, as engagement with industry became a taken-for-granted feature of academic life, involvement in venture formation and start-up activity became commonplace. Examining the archives, patenting data and public records, we identified 10 faculty members involved in founding companies. Prior to the 1990s, such activity was rare. The new ventures formed before 1980 out of this department typically involved the university or an intermediary organization that disseminated application technologies that were not biotechnology related. Between 1980 and 1990, only two individuals started companies, and they left the university. By the early 1990s, conflict of interest policies became standardized. The university permitted exclusive licenses to faculty-owned firms and even came to accept equity from companies as part of licensing agreements. Hence, we see considerably more activity following the adoption of these policies. As Porter (2004) observes, faculty who founded companies in this era did not have to "quit their day jobs."

### *Network Formation*

We turn now to visual representations of the network linkages among individual scientists affiliated with the department and their outside co-inventors between 1970 and 2000. These visualizations afford us the opportunity to expand beyond the faculty and include scientific and technical staff, graduate students, and co-inventors from other departments as well as from outside the university. These representations portray the full population of inventors associated with the department.

In Fig. 2 we visually represent the inventive teams with the nodes reflecting individual scientists and the lines representing linkages through joint invention disclosures. The individuals are coded by shape for career stage

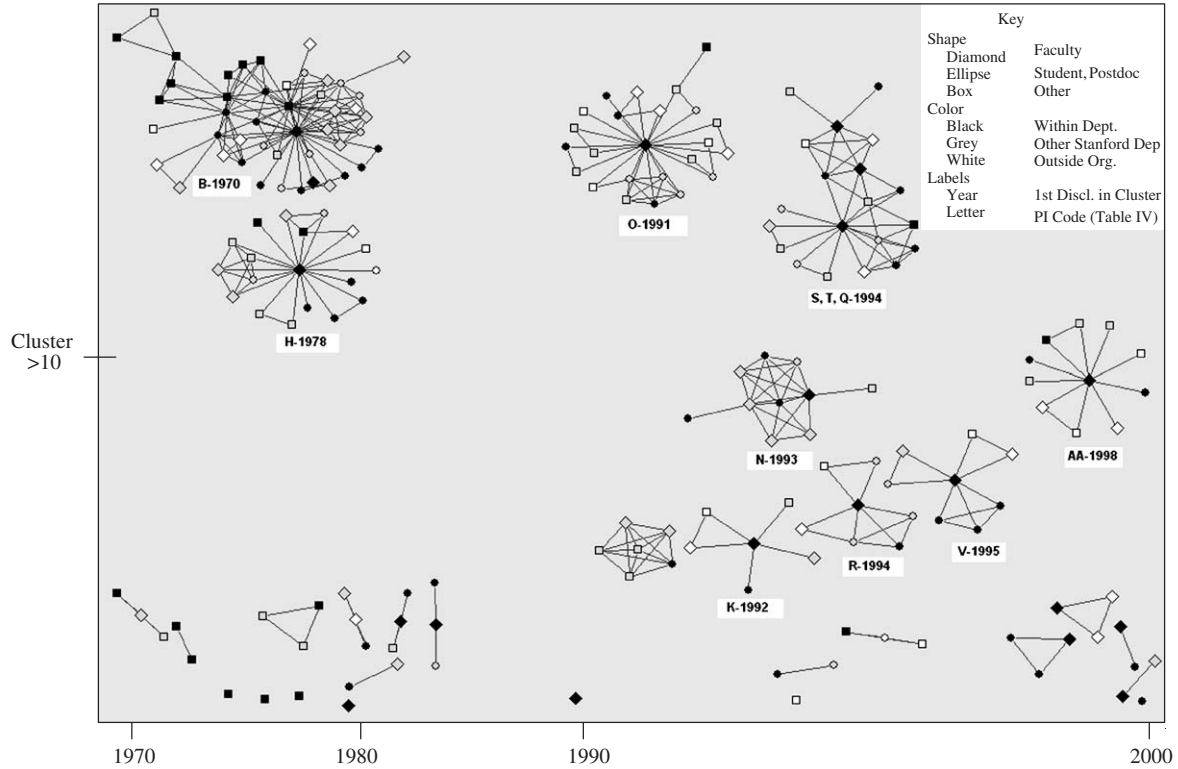


Fig. 2. Scientist Invention Networks, 1970–2000.

and color for affiliation. Diamonds are faculty, ellipses are students or postdoctoral fellows, and boxes reflect other employees such as staff researchers and technicians. Members of our sampled life science department are coded black, while inventors from other parts of the university are gray, and inventors from other universities or companies are white.<sup>7</sup>

Fig. 2 depicts how strongly inventors are clustered around particular research programs. The networks are arrayed by cluster size (vertical axis) and chronology (horizontal axis), based on the year in which a first invention appears in each cluster. For example, in the upper left-hand quadrant of the image, there are two large clusters of inventors with 'founding dates' of 1970 and 1978, respectively. While these network figures include inventions over all years, the placement of the component on the far left of the figure reflects the year of the first invention disclosure by that group. Moving right along the image, there are two more large clusters that emerge in 1991 and 1994, respectively, along with a series of smaller components of 10 or fewer individuals arrayed below. At the very bottom of the figure, the timeline of small clusters consists of inventor teams of three or less, spanning the years of the sample. Owing to the rapid expansion of activity in the 1990s, the chronological scale has been adjusted to condense the first two decades to the left half of the figure, leaving the right half for the highly active decade of the 1990s. Similarly, the vertical placement of the components by size has been adjusted to visualize the largest components in the upper half of the image and the smaller ones in the lower half.

These network pictures vividly portray the strong concentration of inventive activity. Note that there are six components with more than 10 members, and four with more than 18 members. All but two center on a core faculty member in the department. Interestingly, the older clusters, most notably those from 1970 and 1978, have key senior scientists at the center of these large networks. As mentioned earlier, there is limited invention in the decade of the 1980s, as no new large clusters developed during this time. Those inventions that do occur are either in one of the existing clusters founded in 1970 and 1978, or in a few small groups that do not grow beyond teams of two or three. This lack of new cluster formation reinforces the point that most new entrants emerged in the context of existing research programs. Moving to the right of Fig. 2, note how the 1990s reflect a sharp expansion of inventing and a second generation of inventive research programs. These mid-sized components cohere rapidly, rather than remaining small dyads or triads whose ties do not renew. The emergence of these clusters with multiple ties to internal and external collaborators further suggests the integration of entrepreneurship into faculty research programs.

Fig. 3 shows the invention clusters by time period, with the size of the node reflecting prior experience at disclosing. We picture 1980, 1993, and 2000, which captures the five-year windows at the end of each time period. The larger the node, the more disclosures the scientist has. Recall that even until the early 1980s, inventive activity was infrequent, thus the early clusters in the upper left-hand 1980 image are fairly sparse. Inventorship appears sporadic and disconnected, save for the dense clusters that eventually form around Professors B and H in Fig. 2. The largest cluster connects two collaborations with scientists from other departments. Note also the preponderance of technicians, inventing either in teams as in the upper left-hand side of the 1980 image, or on their own in the lower left-hand side. The students and postdocs involved in disclosing are connected to faculty, either within the department or through another Stanford laboratory.

Moving forward, the 1993 image demonstrates how much inventive activity in the previous period was driven by the laboratories of the two prominent faculty, represented by the larger triangles at the center of their inventive clusters. Professor B's cluster in the upper left-hand side of the image has multiple ties, as the group has reconstituted itself with more new inventions. Much of the energy in this second wave of disclosing came from a senior scientist in this laboratory who was promoted to full professor (Professor L) and who developed a research program on her own with nine disclosures in period two and 14 in the third, while continuing ongoing collaborations with Professor B. The small cluster below is Professor H, who is quite experienced, but his group does not expand with either repeat inventions or collaborations among members of the laboratory.

Several comments about the 1993 clusters are in order. Professor B's team continues to be very productive, pursuing work that extends the original innovation from this laboratory. Note the number of larger nodes in this laboratory group, reflecting how many members are now experienced inventors. Many of these collaborations renew themselves or generate new linkages as students and postdocs enter the laboratory. There is a great deal of repeated activity among the participants, suggesting a highly collaborative team, which generates multiple inventions. Some activity reaches outside of the department, but the majority is centered in this laboratory.

Compare this cluster to the contrasting group located just below. What had been a very large hub-and-spoke network organized around Professor H with outside collaborators, is now a triad, as this program did not sustain itself by enrolling new participants or with new collaborations with existing

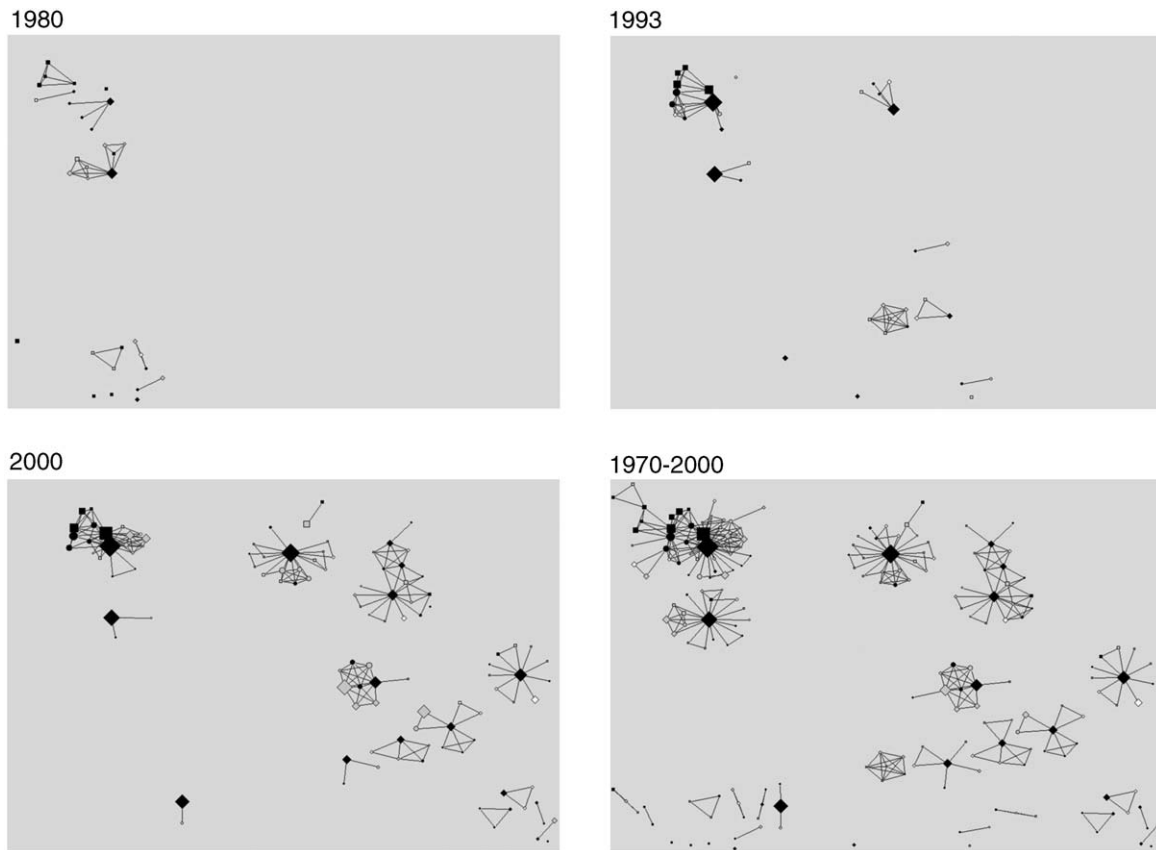


Fig. 3. Scientist Invention Networks with Node Size for Experience.

members. Inventorship in this laboratory relied on a central individual and never expanded to encompass interconnected teams. What the two clusters share, however, is an integration of inventing within the laboratory, as the external co-inventors (coded gray) from the previous era are largely replaced by ties within the laboratory (black) or to faculty outside the university (white). The few gray nodes are students or postdocs affiliated with other basic science departments who collaborate in these laboratories.

Moving to the right of the 1993 image, we see the emergence of a few new small to medium sized clusters. The only cluster in the upper middle image illustrates the appointment of a faculty member, Professor O, to the department who is already an experienced inventor. This professor brings in outside collaborations with two faculty colleagues from another university, depicted by the ties to the two white diamonds. Note again how the student and postdoc inventors in the 1993 image are all connected to a faculty member, either within the department or through work in the laboratories of other Stanford faculty. In the last year of the image, we see for the first time the emergence of a team of two students with no faculty member.

Moving ahead to 2000, notice the growth of new research programs, and an increase in the number of experienced inventors. There are now seven sizable clusters of inventing centered on one or more faculty members. The growth of these prolific teams of faculty, postdocs, students, technical personnel, and outside scientists has become the primary motor for the increase in entrepreneurial activity. Looking at the lower right panel, with the full 31-year period, we see 10 research clusters where considerable inventive activity takes place, with nine of these headed by an experienced inventor. Only one experienced inventor discloses in the context of a very small team, while just one star-shaped cluster does not have an experienced head. A number of the clusters are very sizable laboratories that have produced multiple inventions. Co-invention with other Stanford scientists (gray nodes) is common, reflecting the broadening of life science collaborations and ties to clinicians. Teams of students and postdocs become more common, although embedded in faculty clusters as in those of Professors N and S, T, and O. Note also the solo disclosures of students and postdocs in the lower left of the image and near the cluster of Professors S, T, and O. Finally, it is remarkable just how much the single slice from 2000 captures the broad outline of the summary picture of all years, 1970–2000, suggesting the prevalence of entrepreneurial activity is a relatively recent phenomenon.

The new appointments to the department fueled the expansion of these networks, as the experienced inventors brought industry ties and

demonstrable success in the worlds of both public and private science. Indeed, there is considerable evidence that the department was reshaped by entry as much as adaptation. Recall that both Professors H and N were hired to the department with prior patenting experience to assume the roles of department chair. Professor O joined the department in 1991, bringing a laboratory with students and postdocs, as well as a sizeable \$6.7 million research program. His entry is reflected in the largest cluster in Fig. 3. Professor AA joined in 1998 and developed the fourth largest inventive program of the 1990s. Professor Q, hired in 1991, and Professors S and T, who joined in 1993, came together to form an expansive cluster that links these three tenured faculty's research programs, combining Professor Q's ties to industry and private sector research support with Professor S's interdisciplinary ties and resources. The mid-1990s also reflect the emergence of clusters of entrepreneurship based on the efforts of more junior faculty, such as Professor V, whose more modest annual research expenditures culminated in five disclosures from 1995 to 2000.

## SUMMARY AND IMPLICATIONS

The life sciences underwent a profound intellectual transformation in the 1970s and 1980s, as breakthrough discoveries facilitated new insights into the nature of diseases. In the department we studied, these tools attracted the attention of medical researchers from cardiology to neuroscience to urology, as massive amounts of novel information afforded deeper understanding of the mechanisms leading to diseases. In the context of this scientific revolution, the department expanded and senior faculty received significant federal research support. Given the excitement afoot in this field, it would be surprising indeed if there were not signs of entrepreneurial activity by members of the faculty, as both the newly emerging biotechnology industry and older pharmaceutical and medical instrumentation companies were attracted to these discoveries. Indeed, one sign of the normative power of the older communal, disinterested model of open science is the relative slowness with which entrepreneurship developed in the 1970s and 1980s, despite considerable interest in the work underway in this department by researchers in other departments and scientists in industry. Not until the 1990s did commercially engaged science become widespread.

We account for this general pattern in several steps. Our analysis follows the call of Aldrich and Ruef (2006), and spans multiple levels, connecting

individuals, research networks, the university, and the broader socio-political context. At the individual level, we see that contact with industry in the form of corporate support or revenues was inconsequential until the 1990s. Much more important was the scale of a scientist's research program, notably his or her research budgets. Well-funded scientists with numerous collaborators were more likely to disclose inventions precisely because they had a wider corpus of science and more human capital to draw on. Pecuniary incentives did not loom large in the early years, as the prospect of ample returns from entrepreneurship was unlikely. Most disclosures did not earn any money. By the 1990s, however, these odds changed markedly, suggesting that financial incentives were more an outcome than an input into the development of entrepreneurship. Once commercial opportunities became both visible and legitimate to pursue, industrial involvement comes to be regarded as another core aspect of high-powered science, integrated into career expectations. The pressures to publish, garner grants, fund students, and contribute to the frontiers of science and industry are considerable at a university like Stanford. Based on the disclosure records of this department, we argue that pecuniary considerations were less an incentive and more a part of a broad sweeping change in which public and private science were amalgamated in the 1990s.

To be clear, we are not arguing that financial success was inconsequential. On the contrary, rather than think of revenues as simply money, consider the ways in which commercial rewards reshaped academic science, often unexpectedly. In the case of the successful device from Professor B's laboratory, both the technical needs and opportunities for the laboratory facilitated more engagement with the company in order to develop the invention and eventually mass-produce it. As a result, numerous invention disclosures emerged. Initially, only technicians were inventors, but this group expanded from a small number of technical and research staff to include the principal investigator, co-authors, postdocs, and students. The size of the laboratory grew, fueled by grant money and licensing revenues. The number of inventors in this laboratory increased from 3 in 1970 to 18 in 1984, held steady at 14 in 1995, and rose to 20 in 2000. Many of the inventions included complimentary innovations that were part of the original device, including analysis software, further components of the apparatus, and biological tools and materials that improved the efficacy of the invention. Initially, returns on the invention were utilized to seed a facility that would make the invention available to the entire Stanford community for diagnostic and research purposes, borrowing the department share with the permission of the dean, and combining the funds with the inventor shares as



capital to support this new facility. The financial rewards were employed in different ways over time, but always commensurate with the meanings ascribed to it by the inventors who were steeped in the evolving norms of academic science. One consequence of the commercial success for this group was that laboratory staff had more secure employment, students had access to better research equipment, and the PI became even more productive by the standards of normal science.

Furthermore, as public science was transformed and integrated with private science, so did the constraints and opportunities available to university faculty. Initially, licensing revenues were treated as gifts to replenish the laboratory, then as a means to build a public facility, then as resources to expand the department and the laboratory. While the inventors gained financially, of course, the key to their expansion, we argue, was the broad manner in which these gains were distributed.

Viewed more structurally, in terms of the social organization of laboratory life, the factors that influence inventive behavior were very much tied to the organization of university careers. Newcomers to the disclosure process were unlikely to invent alone; they entered by co-inventing with experienced inventors and/or the principal investigators of the laboratories in which they worked. This process of attachment highlights how much opportunities are shaped, as well as constrained, by whom one collaborates with. As much as newcomers may want to join with financially successful inventors, their ability to do so is limited by where they work. That said, commercial success certainly has many forms of appeal, ranging from resources, laboratory equipment, funding and employment opportunities to personal wealth. But in the context of the contemporary life sciences, such influences operate on young scientists more to shape their choices of which laboratory to go to and what topics to work on (Owen-Smith & Powell, 2001a).

Scientists who are averse to the new blending of public and private science are most concerned that the choice of topics by younger researchers is being shaped by considerations of commercial impact, and that important questions, with no immediate market prospects, are not being explored. Our analysis of the OTL archives and interviews with faculty and students suggest two perspectives on these concerns. To the extent that young scientists join in the context of established teams, then such worries are mitigated as research trajectories are established by the more senior laboratory director. Yet as entrepreneurial involvement becomes more widespread, there is some indication that younger scientists are much more inclined to search for "hot," marketable topics. In some respects, these ramifications are crucial for public health, as more work aimed at specific diseases is being pursued

with urgency. But at the same time, worry about the extent to which academic entrepreneurship makes science more market-driven also seems warranted (Nelson, 2005).

The expansion of entrepreneurial activity was very much conditioned by period effects. The 1970s were an initial era of ferment, and two research programs developed breakthrough technologies that attracted considerable scientific and commercial interest, eventually earning very substantial revenues from successful licenses. But these two laboratories remained the only games in town through most of the 1980s, even though federal legislation and public policy at the time encouraged academic entrepreneurship. The trend changed in the early 1990s, however, as new senior faculty with prior commercial involvement and impressive scientific credentials were brought into the department.

By the late 1990s, the transformation from inventing as a sideline activity pursued by technicians to a core activity by established, high-profile senior scientists was complete. Graduate students began to disclose as well, both with their mentors and occasionally with one another. Younger faculty joined the department and started disclosing inventions, either working solo or with other junior faculty. The network visualizations portray a department partitioned into numerous engaged, highly interactive clusters. By 2000, these inventive teams no longer required an eminent scientist to be at the center of each network. This shift co-occurs with a broader climate change at the university. Entrepreneurship became a venerated activity at Stanford, celebrated by an array of activities on campus and highlighted in numerous university publications. Consider the 2003–2004 Stanford OTL Annual Report “Celebrating Inventors” where prominent faculty inventors were featured and the office proclaims: “our success depends on the researchers whose passion drives the machinery of invention” (Stanford OTL, 2004). Courses such as “Invention 101” became common, fellowships and campus-wide entrepreneurship contests flourished, and university offices offered seminars on “How to be a Stanford Faculty Entrepreneur: Role Models and Resources.”

This shift represents the culmination of a process in which entrepreneurship spread from two early “explorers” to other senior faculty of comparable status and then trickled down the career ladder to become an accepted activity of many life scientists. The identity of scientist entrepreneur became firmly settled and widely embraced, perhaps best summed up in the quip of a UC Berkeley professor that “I have this sense that it’s an almost unwritten rule that you have to start a company to be a successful professor at Stanford” (Abate, 2006).

## NOTES

1. We note below, however, that the norms of publishing and co-authorship did not readily transfer to disclosing and patenting.

2. These data are drawn from a larger project, comprising a systematic comparison of multiple departments at Stanford, that analyzes the development and diffusion of commercial involvement by faculty, students, and staff from 1970 to 2000 (Colyvas, 2007).

3. Of course, co-inventors can come from within or outside the department or university.

4. Note that an extension of this dataset was utilized in an earlier study where we included all inventions of the external scientists that were co-inventors, thus there is a larger total number of inventions in that analysis (Colyvas & Powell, 2006).

5. The main source for these data is the OTL electronic database and filed archives, which together generate a list of invention disclosures and their inventors. A list of names of individuals who had a reported affiliation with the sampled department was retrieved along with their co-inventors. As many faculty have joint appointments, and relevant longitudinal information such as job title or affiliation changes often over the 30-year period, university bulletins and electronic dissertation databases were used to determine whether individuals had a faculty or student affiliation with the department. This list was matched to the OTL list of inventors by hand, using file archives whenever possible to adjudicate among similar or abbreviated names, and organized into a relational database that documented career transitions (i.e. a promotion or departure from the university) at each instance of a new invention disclosure. These data were checked and supplemented through the World Wide Web, using the Google search engine and Google scholar to identify publicly reported affiliations through university websites, CVs and scholarly publications.

6. Initially, Stanford's policy was for the School's share to go to a general university fund. After extensive debate and protest from the Medical School, and by a professor from this department in particular, the policy was changed in the early 1980s to divert the general university share to the Medical School (Colyvas, 2007).

7. The network visualizations were created using Pajek version 1.09 and optimized three times using the Kamada Kawai optimization function. The components were extracted and manually arrayed in the figures to reflect the date of first invention for each node. The nodes and ties were coded for a duration of five years, beginning with the year of disclosure, and the 'generate in time' function in Pajek was utilized to visualize the networks at selected intervals.

## ACKNOWLEDGMENTS

We acknowledge the Columbia-Stanford Consortium on Biomedical Innovation, funded by the Merck Foundation; the States and Markets program at the Santa Fe Institute, supported by the Hewlett Foundation; the Association for Institutional Research; and the Stanford Spencer Research and Training Fellowship program for support for this research. We are

especially grateful to Deborah Strumsky for US patent data and the Office of Technology Licensing at Stanford University for access to their archives. We are also grateful to Kjersten Bunker-Whittington, Michael Lounsbury, Jason Owen-Smith, Kelley Porter, and Martin Ruef for comments on an earlier draft.

## REFERENCES

- Abate, T. (2006). Upstarts and rabble rousers: Stanford fetes 4 decades of computer science pioneers. *San Francisco Chronicle*, March 20. Accessed on June 14, 2006 at <http://www.sfgate.com/cgi-bin/article.cgi?file=/chronicle/archive>
- Agrawal, A., & Henderson, R. (2002). Putting patents in context: Exploring knowledge transfer from MIT. *Management Science*, 48(1), 44–60.
- Aldrich, H. E., & Fiol, C. M. (1994). Fools rush in? The institutional context of industry creation. *Academy of Management Journal*, 19, 645–670.
- Aldrich, H. E., & Ruef, M. (2006). *Organizations evolving*. Thousand Oaks, CA: Sage Publications.
- Bercovitz, J., & Feldman, M. (2005). *Is academic entrepreneurship a team sport?* Working paper, University of Illinois.
- Bok, D. (1982). *Beyond the ivory tower: Social responsibilities of the modern university*. Cambridge, MA: Harvard University Press.
- Burt, R. S. (1987). Social contagion and innovation: Cohesion versus structural equivalence. *American Journal of Sociology*, 92(6), 1287–1335.
- Colyvas, J. A. (2006). *From divergent meanings to common practices: The institutionalization of technology transfer at a research university*. Working paper, Stanford University.
- Colyvas, J. A. (2007). *Institutionalization processes and the commercialization of university research*. Doctoral dissertation, Stanford University School of Education.
- Colyvas, J. A., & Powell, W. W. (2006). Roads to institutionalization: The remaking of boundaries between public and private science. *Research in Organizational Behavior*, 27, 315–363.
- Dickson, D. (1980). Inventorship dispute stalls DNA patent application. *Nature*, 284(3), 388.
- DiGregorio, D., & Shane, S. (2003). Why do some universities generate more start-ups than others? *Research Policy*, 32(2), 209–227.
- Economist*. (1989). *Jewels in the ivory tower*. Jan. 28: 82.
- Economist*. (2002). *Innovation's golden goose*. Dec. 12: 3.
- Etzkowitz, H. (1998). The norms of entrepreneurial science: cognitive effects of the new university–industry linkages. *Research Policy*, 27(8), 823–833.
- Fagerberg, J. (2003). Schumpeter and the revival of evolutionary economics: An appraisal of the literature. *Journal of Evolutionary Economics*, 13, 125–159.
- Gompers, P., Lerner, J., & Scharfstein, D. (2003). *Entrepreneurial spawning: Public corporations and the genesis of new ventures, 1986–1999*. NBER working paper no. 9816.
- Hicks, D. M., & Katz, J. S. (1996). Where is science going? *Science, Technology and Human Values*, 21(4), 379–406.
- Higgins, M. (2004). *Career imprinting: The case of the Baxter boys and biotechnology*. Boston: Harvard Business School Press.

- Hwang, H., & Powell, W. W. (2005). Institutions and entrepreneurship. In: S. Alvarez, R. Agrawal & O. Sorenson (Eds), *Handbook of entrepreneurship research* (pp. 179–210). Kluwer Publishers.
- Krinsky, S. (2003). *Science in the private interest*. Lanham, MD: Rowman and Littlefield.
- Lach, S., & Schankerman, M. (2004). Royalty sharing and technology licensing in universities. *Journal of the European Economic Association*, 2(2–3), 252–264.
- Levin, S., & Stephan, P. (1991). Research productivity over the life cycle: Evidence for academic scientists. *American Economic Review*, 81(1), 114–132.
- Lorrain, F., & White, H. C. (1971). Structural equivalence of individuals in social networks. *Journal of Mathematical Sociology*, 1, 49–80.
- Lowe, R. A. (2006). Who develops a university invention? *Journal of Technology Transfer*, 31(4), 415–429.
- Macy, W. (1968). Discovery to patent to license. *Science*, 162(3858), 1075.
- Merton, R. K. (1968). The Matthew effect in science. *Science*, 159(3810), 56–63.
- Merton, R. K. (1973). *The sociology of science*. Chicago, IL: University of Chicago Press.
- Mowery, D. C., Nelson, R. R., Sampat, B., & Ziedonis, A. (2004). *Ivory tower and industrial innovation*. Stanford, CA: Stanford University Press.
- Murray, F. (2002). Innovation as co-evolution of scientific and technological networks: Exploring tissue engineering. *Research Policy*, 31(8–9), 1389–1403.
- Nelson, R. R., (Ed.) (2005). Basic scientific research. In: *The limits of market organization*. New York, NY: Russell Sage Foundation, 233–258.
- Owen-Smith, J. (2003). From separate systems to a hybrid order: Accumulative advantage across public and private science at research one universities. *Research Policy*, 32, 1081–1104.
- Owen-Smith, J., & Powell, W. W. (2001a). Careers and contradictions: Faculty responses to the transformation of knowledge and its uses in the life sciences. *Research in the Sociology of Work*, 10, 109–140.
- Owen-Smith, J., & Powell, W. W. (2001b). To patent or not: Faculty decisions and institutional success at technology transfer. *Journal of Technology Transfer*, 26(1/2), 99–114.
- Owen-Smith, J., & Powell, W. W. (2003). The expanding role of university patenting in the life sciences. *Research Policy*, 32(9), 1695–1711.
- Philips, D. J., & Zuckerman, E. W. (2001). Middle-status conformity: Theoretical restatement and empirical demonstration in two markets. *American Journal of Sociology*, 107, 379–429.
- Porter, K. (2004). *You can't leave your past behind: The influence of founders' career histories on their firms*. Unpublished Ph.D. dissertation, Department of Management Science and Engineering, Stanford University.
- Portes, A., & Sensenbrenner, J. (1993). Embeddedness and immigration: Notes on the social determinants of economic action. *American Journal of Sociology*, 98, 1320–1351.
- Powell, W. W. (1996). Inter-organizational collaboration in the biotechnology industry. *Journal of Institutional and Theoretical Economics*, 120(1), 197–215.
- Powell, W. W., & Owen-Smith, J. (1998). Universities and the market for intellectual property in the life sciences. *Journal of Policy Analysis and Management*, 17(2), 253–277.
- Powell, W. W., Owen-Smith, J., & Colyvas, J. (2007). Innovation and emulation: Lessons from the experiences of US universities in selling private rights to public knowledge. *Minerva*, forthcoming.
- Schumpeter, J. (1934). *Theory of economic development*. New York: Oxford University Press.

- Shane, S. (2004). *Academic entrepreneurship: University spinoffs and wealth creation*. London: Edward Elgar.
- Slaughter, S., & Leslie, L. (1997). *Academic capitalism: Politics, policies, and the entrepreneurial university*. Baltimore: Johns Hopkins University Press.
- Stanford University. (2001). *Annual report of the Office of Technology Licensing 2000–2001*. Available at <http://otl.stanford.edu/about/resources/otlar01.pdf>
- Stanford University. (2004). *Annual report of the Office of Technology Licensing 2003–2004*. Available at <http://otl.stanford.edu/about/resources/otlar04.pdf>
- Stephan, P., Gormu, S., Sumell, A. J., & Black, G. (2005). "Who's Patenting in the University." *Economics of Innovation and New Technology*, (forthcoming).
- Stuart, T. E., & Ding, W. W. (2006). When do scientists become entrepreneurs? The social structural antecedents of commercial activity in the academic life sciences. *American Journal of Sociology*, 112(2006), 97–144.
- Swedberg, R. (Ed.) (2000). *Entrepreneurship: The social science view*. Oxford: Oxford University Press.
- Thursby, J. G., & Thursby, M. C. (2002). Who is selling the ivory tower? Sources of growth in university licensing. *Management Science*, 48, 90–104.
- Vallas, S., & Kleinman, D. (2006). *Contradiction, convergence, and the knowledge economy*. Working paper, George Mason University.
- Washburn, J. (2005). *University Inc.: The corporate corruption of higher education*. New York: Perseus.
- Zucker, L., Darby, M., & Brewer, M. (1998). Intellectual human capital and the birth of U.S. biotechnology enterprises. *American Economic Review*, 88, 290–306.

# START-UPS IN SCIENCE: ENTREPRENEURS, DIVERSE BACKING, AND NOVELTY OUTSIDE BUSINESS

James A. Evans

## ABSTRACT

*Sociological studies of entrepreneurship focus on social and technical innovations in business. Using an illustration from molecular plant biology and the historical evolution of the term “entrepreneur,” I make a case for the theoretical and methodological importance of studying entrepreneurs and their ventures outside the scope of traditional business. Then, considering the scientific lab as a self-consciously entrepreneurial venture, I use the population of molecular biology labs studying the plant *Arabidopsis thaliana* to demonstrate a relationship less directly measurable among start-ups in business: diverse sources of funding accompany original activities and ideas within a venture. This is not, however, what predicts lab success. Lab size drives success, but hinders originality. Moreover, I show that established institutions in science are usually the ones that become innovations in business.*

---

The Sociology of Entrepreneurship

Research in the Sociology of Organizations, Volume 25, 261–305

Copyright © 2007 by Elsevier Ltd.

All rights of reproduction in any form reserved

ISSN: 0733-558X/doi:10.1016/S0733-558X(06)25008-0

Chris Somerville was one of the first scientists to introduce molecular techniques to the study of plants, he co-founded an international organization to coordinate science using the laboratory plant *Arabidopsis thaliana*, and he currently sits as chairman and CEO of a plant genomics start-up, Mendel biotech. In an interview with Somerville, he separated the substance of his academic and industry research as “totally different.” Moreover, he characterized his commercial concern as “completely separate. It has to be. My lab here at the university is completely open. The work at the company is proprietary. If I see something at the company that would be valuable to a colleague at the university, I don’t say anything.” And yet the process of undertaking them seemed strikingly similar.

In 1997, Somerville and a handful of fellow star biologists charted what they believed to be a commercially promising course: identify functions of the roughly 1,800 transcription factor genes in the model plant, *Arabidopsis*, and then extend those findings and any techniques developed along the way into economically important plants like tomatoes, corn, and trees. Transcription factors are the protein control switches that regulate genetic activity within organisms. In plants, they direct processes as diverse as disease resistance, the ability to withstand freezing or drought, reproduction, plant and fruit size, nitrogen use and other, more complex traits. Mendel appeared in the swell before the first wave of genomics companies. Craig Venter founded Celera, the noted human genomics company a year later. Initially, Mendel founders promoted their plan broadly to eventually partner with the largest company in transgenic agriculture, Monsanto, in addition to Empresas La Moderna (ELM), a Mexican vegetable and fruit firm. In return for cash and future royalties, critical biotech tools and the best “germplasm” or genetic specimens available, Mendel passed along rights to their patented discoveries in large acre grain and vegetable crops. This left the exploitation of specialty markets for themselves. As the management team developed several other corporate relationships in forestry, turf, pharmaceuticals, and other industries, they sought and received a number of Small Business Innovative Research grants from the U.S. Department of Agriculture. Now, with more than a thousand patents issued or pending, Mendel allows its scientists to participate in public science through publishing, but “on their own time.”

The summer after graduating with a biology Ph.D. from Alberta where he studied gene regulation in *E. coli* – the facile microbe model – Somerville (1978) bootstrapped a trip to France with his wife, Shauna. There, they read in the Library of Marie Curie in the mornings and did gedanken experiments over long, Parisian lunches in the afternoon. Somerville realized that



to genetically dissect plants in the way that he had microbes would require something simpler than wheat or corn. That's when he came across an article extolling the virtues of *Arabidopsis thaliana*, a weedy relative of mustard: it grew quickly, reproduced abundantly, and appeared to have a spare genome – it was well-behaved (Redei, 1975). In Bill Ogren's lab at the University of Illinois, Somerville was a productive postdoc. He identified a number of *Arabidopsis* mutants that allowed him to uncover the biochemical pathway through which photorespiration took place. After a brief return to the University of Alberta, Somerville established his lab at Michigan State University and received one of the first Presidential Young Investigator's Awards (1984), which was matched by DuPont, to help finance it.

Somerville used research talks as an opportunity to proselytize *Arabidopsis* as a platform from which to enter the next generation of genetic research in plants. Not surprisingly, researchers committed to crop improvement spurned the idea of an intrinsically useless model. In his own lab, Somerville initiated a wide-ranging program that included investigations of lipid metabolism, plant development, and molecular genetics. He also published work of immediate interest to companies, like research on plant resistance to DuPont's herbicide sulfonylurea. When the *Arabidopsis* community grew sufficiently large, James Watson of DNA double helix fame convened a meeting including Somerville, a team of other top U.S. plant scientists, and Peter Bloch, director of the NSF, to propose the sequencing of *Arabidopsis*' genome. Watson forwarded the idea in hopes that plant findings would prove useful for interpreting the human genome he had championed. The project was funded and with others, Somerville formed the Multi-National and North American Steering Committees for *Arabidopsis* Research to coordinate sequencing efforts. Following a 1987 meeting Somerville organized at Michigan, the new *Arabidopsis* Steering Committees sponsored an annual conference.

When Somerville assumed leadership of the Carnegie Research Institute's Department of Plant Biology at Stanford University in the early 1990s, Somerville's lab continued studies of development, but also began to explore polysaccharides and the cell wall. At Michigan, Somerville had established an email listserver that allowed members of the emerging *Arabidopsis* community to share results and solicit technical assistance. After the move to Stanford, he applied for and received a grant to archive all published and contributed *Arabidopsis* findings. Served through the web with more than 30 million hits a year, this has become a widely used resource in the community. With a string of pioneering discoveries in a plant that had become the major setting for plant research, his plant biology corpus became the

second most cited in the world, just after Elliot Meyerowitz, who serves beside him on the science board of Mendel.

This story renders the entrepreneurial process of institution-formation (Schumpeter, 1934[1911]) in two distinctive environments. In industry, Somerville sought leads on innovative products; in the academy, innovative theory and findings. And yet both involved the creation of semi-autonomous organizations – an academic and an industrial lab. Both involved the cultivation of funding partners from industry and government. Both involved the cultivation of new markets to consume his products. Both involved the creation of a field – and its ancillary supportive organizations – into which Somerville’s own organizations could mature. That the same person built parallel structures in two fields, and that the business venture followed, in critical ways, the academic venture underscores the importance of studying entrepreneurial action beyond the realm of traditional business.

Arguably, this broader notion of entrepreneurship is truer to the etymological history of the word “entrepreneur” and the commonly broad connotation of “entrepreneurial.” Our English entrepreneur comes from the French *entreprendre*, to undertake. An “entreprennoure,” then, initially referred to those undertaking military campaigns. In its first known English usage, William Worcester (1860[1475]) uses it in the *Boke of Noblesse*, a call for Edward IV to renew a campaign against France using the Hundred Years War, figures from Greek and Roman history (and a degenerate term from its own vocabulary) as inspiration: “That most noble centoure Publius Decius, so hardie an entreprennoure in the bataile, whan the Romains were almost overthrow, he avausid hym silfe so ferre in the bataile, to die to th’entent to make the Romains more gret, and felle for his dethe in fighting tille they had the victory” (*Bk. Noblesse*).<sup>1</sup> This entreprennoure, consul of Rome and general in the Battle of Vesuvius, took full risk of profit and loss as a “manly man” directing his warrior “felyshyp” in a surprising, risky, and ultimately doomed maneuver.

William Caxton, the first English printer, uses it again in 1485 in his translation of the French romance of Fierabras<sup>2</sup> into *The Lyf of the Noble and Crysten Prince, Charles the Grete*. Caxton lifts it directly from French with a twist that would prove binding. “Rychard [of normandie – an ally to Charles] went to fore as chyef enterprenour” in a military ruse where 500 French knights disguise themselves as merchants in an effort to pass the Saracen giant, Galafre, who guards the bridge to Balan, the pagan sultan of Spain (Caxton, 1881[1485]). This primes its first modern usage in the entertainment industry (1828): the process of assembling a temporary cast of characters to produce, stage, and direct a performance (like

Rychard's; *entrepreneur*, n.1, Oxford English Dictionary, 2006). The first general usage referring to business beyond the arts appeared in 1852 in Froude's *Life of Carlisle* ("A public set of rooms ... finer than some palaces ... all built by one French gambling entrepreneur," Froude, 1884, p. 107), where its denotative opportunism continued to connote artistry (and Frenchness). Even after the topic of "entrepreneurship" had become a reasonable subject of study in the mid-20th century (the word first entered Webster's dictionary in 1934), Joseph Schumpeter (1943), one of its leading exponents, described its "social function" as central to the economy, but in ways that still suggest extension to broader undertakings: It was ...

to reform or revolutionize the pattern of production by exploiting an invention or, more generally, an untried technological possibility for producing a new commodity or producing an old one in a new way, by opening up a new source of supply of materials or a new outlet for products, by reorganizing industry and so on .... To undertake such new things constitutes a distinct economic function, first, because they lie outside of the routine tasks which everybody understands and, secondly, because the environment resists in many ways that vary, according to social conditions, from simple refusal either to finance or to buy a new thing, to physical attack on the man who tries to produce it. To act with confidence beyond the range of familiar beacons and to overcome that resistance requires aptitudes that are present in only a small fraction of the population and that define the entrepreneurial type as well as the entrepreneurial function. This function does not essentially consist in either inventing anything or otherwise creating the conditions which the enterprise exploits. It consists in getting things done. (1943, p. 132)<sup>3</sup>

Although clearly situated in the realms of finance, production, and sales, it is not hard to imagine other socially important contexts in which entrepreneurs eschew institutionally "familiar beacons" to get things done. In some of these contexts, the specie is still coin: government grants, arts patronage, and philanthropy. In others, the currencies are less liquid: votes, status, and commitment (Parsons, 1968). But because institution-building exhibits similar patterns in business and nonbusiness contexts, and because entrepreneurship in one context often involves the seeding of familiar institutions from another, a sociology of entrepreneurship should examine and compare entrepreneurship across contexts.

Ron Burt's (1992) theory of structural holes, a theory of entrepreneurial opportunities, does this by billing itself broadly as a general theory of competitive action.<sup>4</sup> And yet, all of the data come from business: industries, managers, and firms. Shane's (2003) *General Theory of Entrepreneurship* does the same. It's not that the sociology of entrepreneurship does not extend implications from business to the rest of the social world; it's that it has not as readily received them. Note that this is not true for the sociology of

organizations, where data outside traditional business contexts proved critical to sociological conceptions of business. The rule-bound German state summoned Weber's (1968[1922]) theory of bureaucracy; the role of consultants in seeding mimicry between nonprofits inspired DiMaggio and Powell's (1983) new institutionalism; the population of labor unions anchored one of Hannan and Freeman's (1988) first empirical studies of organizational ecology. The importance of studying entrepreneurial activity across contexts is not only crucial for theoretical generality, but for methodological access. In different contexts, actors consider different things precious, beautiful, and sacred. As such, each setting reveals and obscures distinct features. In this paper, I consider academic laboratories as new ventures.

In chemistry and the laboratory life sciences, students commonly enter a professor's laboratory within their first or second year of graduate training and apprentice there throughout their program. Some laboratories are tiny, walk-in closets, or small segments of a much larger "bench," furnished with piles of cast-off analog instruments and funded by drips of institutional and state money.<sup>5</sup> Others, like Lee Hood's, who invented the first DNA sequencing machines at Caltech, are giant factories hosting scores of researchers working on diverse projects and fueled by investments and gifts from public, nonprofit, and private interests. In some, student apprentices rotate across the projects of the lab, or across functions required to sustain those projects. In others, students specialize early on and gamble on the success of a solo initiative. In either case, the defense of a thesis represents an acknowledgment of the resources students take to their next venture – techniques and publications – usually as a postdoc.

In graduate school and postdoctoral training, apprentice researchers reconcile their interests and abilities with the realities of the job market and begin to sort themselves into or out of research careers – and if research, then in an industry or academic/institute context. One advanced biology student working in an *Arabidopsis* lab framed academic labs as the entrepreneurial option:

I don't really know exactly what [industry] used to be like, but ... the general feeling that I get is there used to be a lot more restrictions on what you could do, and whether or not you'd be able to publish – how much control you had over what projects were actually getting done. So, I think that is what is most attractive about an academic position for me is that you are basically sort of like an entrepreneur and you have control over what you are going to study. It's ... up to you to do good research then, once you write these grants, to do things. But you are on your own; you get to decide what you think is important.

These research choices shape the experience and tools that postdoctorate researchers cultivate as journeymen, or experienced apprentices, in the labs

of others. Another *Arabidopsis* student, from a different institution, explained how students moving toward industry vs. the academy differ:

Postdocs going into industry are TOTALLY different. Its just a totally different game. There was one postdoc in our lab who was going into industry, and he spent all of his time making all of these different vectors ... for the lab – every kind of vector that you could think of. And his resume looked very different than someone looking for a university job. A university scientist will be all about creating a vision, a project that can attract and train graduate students, the ability to get grants .... So yes, techniques, methods, things such that the company can call [my PI] and she can say, “This guy is amazing, he can create anything.”

Postdocs aiming for academic labs develop the vision – the business plan – for a research program that justifies founding an academic laboratory. If they can sell financiers – most often government funding agencies and universities – on their vision and their possession of the skills to effect it, they are given start-up capital to buy the machines, people, and space required to publish initial findings. They then use these to attract more funds and people.

As scientists begin to build their own academic ventures, their activities often shift from hands-on research inside the lab to financing and publicity outside it. Successful ventures put out an innovative product – batches of scientific papers and patents – consumed broadly by their scientific colleagues and their sponsors. Failures either do not produce or their findings are not consumed. They are eventually shutdown from scientific embarrassment or failure to pay rent.

Academic labs expose something that young business ventures often obscure: They leave a published record of their activities – their hypotheses, their experiments, and sometimes, more rarely, their failures. Much of Marxist sociology, industrial sociology, and more recent organization theory can be understood as an extended critique of the over-simplicity and presumed efficiency of economists’ “production function” view of the firm.<sup>6</sup> But that is often all that young, neurotic firms project: a black box with flax forked in one side and spun gold streaming out the other. This fuels the industry of the so-called “trait- and rate”-based entrepreneurship research that focuses on the contribution of social characteristics and contexts to entrepreneurial success (Thornton, 1999), but not on what is “entrepreneurial” about them besides having just begun. Thus, academic labs provide an opportunity to address an issue at the heart of entrepreneurial institution-building. It is interesting that the growing sub-literature on academic entrepreneurs has rarely used them in this way.

Existing scholarship conceives academics as entrepreneurs if they participate in the commercial sector, but not if they act innovatively or effectively

in the academic one. In an early paper – “Entrepreneurs in Academe” – Louis, Blumenthal, Gluck, and Stoto (1989) examine the correlates of “academic entrepreneurship,” defined as academics securing commercial financing (from consulting and industry sponsorship), engaging in commercial activities (suggested by equity in a start-up) or producing commercial outputs (patents).<sup>7</sup> Subsequent research has focused on activities and outputs, casting academic entrepreneurship as virtually synonymous with patenting (Henderson, Jaffe, & Trajtenberg, 1997; Mowery & Zeidonis, 2002; Owen-Smith, 2003; Markeiwicz & DiMinin, 2004),<sup>8</sup> licensing (Thursby & Thursby, 2003), and commercializing inventions through new ventures (Zucker & Darby, 1996; Owen-Smith & Powell, 2001; Ding & Stuart, 2006). More recent research has branched out to examine the consequences as well as the causes of these forms of commercializing science. A recent (2005) NBER conference on “Academic Science and Entrepreneurship” groups its papers under the headings “The Commercialization of University Inventions,” “The Impact of Academic Science on Private Sector Innovation,” and “The Impact of Academic Entrepreneurship on the Academic Enterprise.”<sup>9</sup>

By framing academic entrepreneurship as commercial, existing research traces an important infusion of new ideas, products, and people into the economy. It also tracks commercial customs back into the academy. Less commonly has it exploited variation in the organization, activities and outputs of academic labs on their own terms to open the black box of new ventures.<sup>10</sup> Schumpeter (1934[1911]) characterized innovations as new combinations of preexisting ideas and institutions. In the quotation stated earlier, Schumpeter documents common forms into which firms structure these combinations: they cultivate new sources of supply and exploit new markets, organize business in new ways, and fashion new methods of production or new products (Fagerberg, 2004, p. 6). And yet we see in the Somerville case that even these innovations – all of them applicable to the academic lab context – rarely stand-alone and are themselves combined to form new ventures. In Somerville’s lab, new sources of supply (funds from industry and government) and the exploitation of a new market (turning plant physiologists into consumers of molecular biology) coupled with his continued ability to create new products (original and important experiments using *Arabidopsis*) for consumption. One key to understanding the structure and process of entrepreneurial institution-building is to examine the patterned relationship between entrepreneurial innovations. Of particular interest to sociologists should be the relationship between social innovations and technical ones (Ruef, 2002).

This relationship receives attention in recent work by Hollingsworth, Hollingsworth, and Hage (forthcoming) and Hollingsworth and Hollingsworth (2000) about the organizational determinants of major, path-breaking discoveries in medical science. They argue that interdisciplinarity, strong leadership, and other organizationally integrating features contribute to scientific findings identified post hoc as breakthroughs. In working backwards from breakthroughs to causes – in sampling on the dependent variable – their causes appear necessary but not sufficient. It is not clear, for example, that interdisciplinary, organizationally integrated settings produce more breakthroughs *on average*. Moreover, the notion of a breakthrough depends not only on the context of creation, but on receptivity of the environment and the attribution of priority.<sup>11</sup> Consistent with Hollingsworth and Hage, however, I will argue that certain types of innovative organization and supply lead to more innovative – though not necessarily more successful<sup>12</sup> – activity within academic labs.<sup>13</sup>

Richard Nelson has long argued that a scientific field populated both by commercial and noncommercial interests or approaches will give rise to more innovations than a field captured by one or the other (Nelson, 1981, 1986). This proposition can also be considered *within* organizations. For an entrepreneurial establishment, it suggests that a combination of interests involved in the founding and sponsorship of an enterprise will generate more novel activities and ideas within it. Two primary forces undergird this relationship. The first is that diverse interests bring diverse concerns to the venture, each of which adds distinct and sometimes previously uncombined ingredients. For example, Burt (2004) identifies a connection between the structural diversity of relationships maintained by managers in a global engineering firm and the quality of their ideas. In his analysis, managers with narrow networks produced narrow, “whiny” suggestions for their bosses, while managers with broad networks that bridged disconnected groups – structural holes – generated innovative, promising proposals. Those with more diverse networks synthesized diverse interests into their proposals.

In the domain of invention, Hage and Hollingsworth (2000) (see also Kline & Rosenberg, 1986) argue that settings which foster communication between scientists with diverse scientific and technological interests are more likely to generate radical product or process innovations than those which do not. Radical innovations demand complementary sub-innovations in scientific discovery, technological invention, and product development. Moreover, an innovation in any link of the product development chain will be more successful if it takes into account the distinct concerns and challenges experienced at other links. In the specific context of academic labs,

success is conditioned on obtaining deeper understanding of the natural world through focused experimental development of theory. The infusion of applied and commercial concerns to an academic lab can introduce radical, if not always successful, experimental possibilities. By bringing an interest in solving real-world problems, like bolstering a crop's defense against predators, companies motivate risky experiments, like within-plant generation of selectively toxic proteins, which would not have been considered with only academic theory in mind (Evans, 2006).

Sponsorship ties to academic labs have a character that facilitates this innovation. In a study of new venture creation, Ruef (2002) demonstrates that social ties can constrain as well as enable innovation. Strong, intimate social ties foster the sharing of information, but they also confer control and induce conformity. Weak ties, which involve limited engagement between more distant and diverse actors, facilitate the flow of information without influence (Granovetter, 1973). Ruef found that entrepreneurs with weak ties to the source of their inspiration – to customers or suppliers – were more innovative than those who gained insight through strong ties – through family or friends. Science funding often displays the character of a weak rather than a strong tie because scientists have a cultural mandate to produce disinterested knowledge (Merton, 1942). This is why, for example, it makes news as a breach of professional ethics each time analysts rediscover that pharmaceutical funding modestly influences clinical trials that test their drugs (Council of Scientific Affairs, AMA, 2004; Lexchin, 2004; Montgomery et al., 2004; Götzsche et al., 2006). If scientists appear invested in positive results from their studies, few will believe them. From this position of prescribed indifference, academic scientists gain information and ideas through their “weak ties” to diverse patrons, but rarely bold influence.

The second reason that disparate interests, embodied in diverse sources of funding, promote innovative action within labs is that they indirectly provide scientists with more discretion. By using the same research to speak to the distinct interests of different funders – by being multi-vocal (Padgett & Ansel, 1992) – scientists with institutionally diverse financing can become less accountable. For example, if the U.S. National Institutes of Health (NIH) grants a million dollars to a university lab for stomach cancer research, it expects a commensurable quantity of research from the project. If NIH adds an additional 500 thousand to extend the project, it expects a proportional increase in results. If, however, a private company puts up the additional money to test a new set of assays, the lab can combine the projects, satisfy both patrons, and have money left over with which to explore their own ideas.



James March (1991) has often contended that in periods of resource abundance, companies explore new possibilities, while in periods of scarcity they exploit exiting knowledge. The most common sources of academic funds, government grants, are usually administered through a peer review process that favors incremental amendment to established science (Evans, 2006). Slack funds may enable scientists to follow their more controversial hunches. With new hypotheses proposed by diverse interests and the freedom to test them with unaccountable funds, I propose that scientists with diverse sponsors are more likely to engage in risky, exploratory experiments than those without.

I do not anticipate that diverse funding will, on average, lead to greater success. The novelty that diversity induces will at some times be inspired and at others doomed. More conservative forces like lab size or total lab funding are more likely to predict average success. Greater size may, however, make labs less susceptible to learning from diverse sponsors or engaging in scientifically risky exploration. Levinthal and March (1993) contend and Almeida, Dokko, and Rosenkopf (2003) demonstrate that as start-ups grow they are more likely to turn inward and tend to ignore their informal environment. Academic labs provide an ideal setting in which to examine these relationships: Labs produce scientific papers that document their activities and acknowledge their sponsors and they receive academic citations that define their success. And so it is within the population of *Arabidopsis* labs – not just Chris Somerville’s – that I test them.

## EVIDENCE FROM *ARABIDOPSIS*

Eleven years after the meeting I described in the opening between Somerville, Watson, Bloch, and others in 1989, an international confederation of scientific teams completed the sequence of the *Arabidopsis thaliana* genome, which set it to become the dominant genetic model organism in plant biology and agricultural biotechnology (Walbot, 2000) just as the mouse and *Drosophila* (fly) serve as animal models. Most major universities, many research institutes and government agencies, and all major plant biotech companies perform or fund research on *Arabidopsis*. *Arabidopsis*’s short, haploid genome and abundant seed production make it easy to simultaneously study molecular and classical genetics – linking DNA sequences to plant functions. Fundamental discoveries in *Arabidopsis* have fed back into basic science, revealing previously unknown similarities between plant and animal function. Companies like Mendel are increasingly

transferring insights from *Arabidopsis* into crop plants with social and profitable implications, such as drought-resistant crops, and the plant-manufacture of oils and other pharmaceutical and industrial substances. As such, studying *Arabidopsis* research constitutes more than a single case. *Arabidopsis* provides an elegant platform from which to study funding diversity on entrepreneurial research activities in the world of plant biotechnology.

Moreover, scientific use of *Arabidopsis*, and plant molecular biology in general, are little more than three decades old and have grown up within the new system of intimate connection between firms and universities. Monsanto, Novartis, DuPont, Dow, and other companies in the agricultural, lumber, chemical, food, and pharmaceutical industries became interested in agricultural biotechnology at about the same time that *Arabidopsis* began to emerge as a research tool in the early 1980s. A number of small, dedicated plant biotechnology companies emerged soon after. As such, this setting provides an opportunity to examine the consequences of funding diversity within this multi-sector system over time rather than forcing focus on the evolution from a single-sector system – and change in the meaning of diverse funding – over time.

#### *Publication Data*

To evaluate the effect of funding diversity and lab structure on the originality of scientific activity and ideas within the *Arabidopsis* community, I collected all 18,359 published articles utilizing *Arabidopsis thaliana* between 1907 and 2002 by combining publication data from BIOSIS, PubMed, Medline, AGRICOLA, and SciSearch where “*Arabidopsis*” or “*thaliana*” was mentioned in the title, abstract, or author-provided keywords of an article. I added abstracts from *Arabidopsis* papers presented in plant molecular biology conferences over the past several years with the help of the U.S. Department of Agriculture and the *Arabidopsis* Information Resource (TAIR). I also gathered citation data on the 11,000 *Arabidopsis* articles, which are present in the SciSearch database between 1945 and 2002. SciSearch is produced by Thompson Scientific’s Institute for Scientific Information (ISI). Citations link the 11,000 ISI *Arabidopsis* articles, by citation, to *Arabidopsis* articles within that set and to an additional 45,577 non-*Arabidopsis* articles which cite them.<sup>14</sup> Using this information, Fig. 1 shows the growth in *Arabidopsis* papers and citations to *Arabidopsis* papers worldwide, from 1975 until 2002. In the course of this time period, the number of papers grew exponentially, from 91 and 112 in 1980 and 1985, respectively, to 373 in 1990,

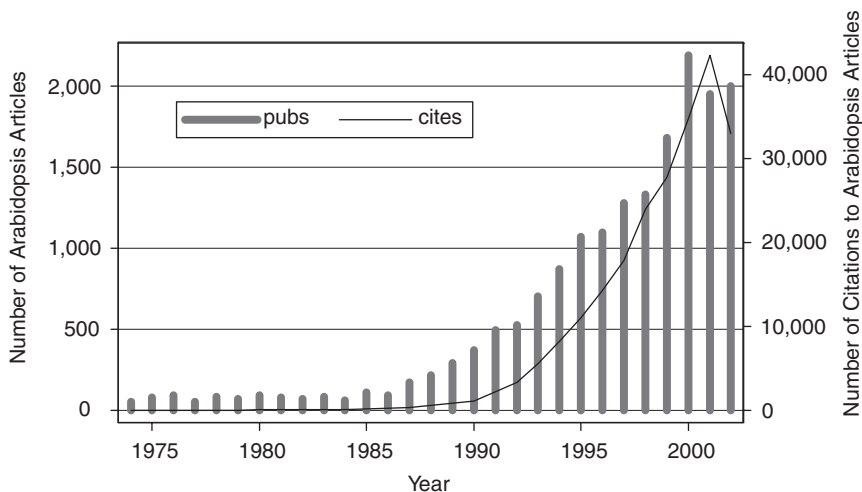


Fig. 1. The Growth of *Arabidopsis* Articles and Citations.

1,070 in 1995, and 2,192 in 2000. Table 1 organizes the *Arabidopsis* papers in SciSearch and their citations by broad, ISI-defined subject area. The table illustrates the variety of fields which *Arabidopsis* articles address, and their relative influence in each. They, of course, cluster in the “plant sciences” (6,512 papers; 18 cites per paper), but also touch on fundamental topics in the “multi-disciplinary sciences” (55 cites per paper) and “cytology” or cell biology (126 cites per paper), and more applied, less academic topics in “agriculture” (5 cites per paper) and “toxicology” (9 cites per paper).

From the titles and abstracts of *Arabidopsis* publications, I extracted an extensive list of scientific terms, corresponding to *Arabidopsis* genes, proteins, species, techniques, biological processes, molecular functions, cellular components, *Arabidopsis* developmental stages, and anatomical locations. Curators from TAIR with PhDs in molecular plant biology hand-coded all *Arabidopsis* genes in the publications and then coded those genes with terms corresponding to their molecular function, the biological process of which they are part, the cellular components they affect, and the stage of development and anatomical position at which they are expressed.<sup>15</sup> For example, their codes indicate that the gene LEAFY is related to flowering as inferred from a mutant phenotype (knocking out the gene and observing flowering mutation).

I contributed to this process by testing for and classifying frequent single- and multi-word concepts in the *Arabidopsis* abstracts,<sup>16</sup> and also by using a

**Table 1.** *Arabidopsis* Publishing Activity: Most Cited Subject Areas<sup>a</sup>.

Subject Area	<i>N</i> Cites <sup>b</sup>	<i>N</i> Papers <sup>c</sup>	Papers/Cites	First Paper
Plant sciences	115,066	6,512	17.67	1974
Biochemistry and molecular biology	92,270	4,319	21.36	1974
Multidisciplinary sciences	45,257	823	54.99	1974
Cell biology	25,503	1,577	16.17	1981
Genetics and heredity	23,669	1,218	19.43	1974
Biology	18,518	684	27.07	1979
Developmental biology	12,419	409	30.36	1979
Biophysics	5,651	494	11.44	1980
Cytology and histology	4,166	33	126.24	1974
Biotechnology and applied microbiology	4,121	416	9.91	1987
Physiology	713	53	13.45	1977
Biology miscellaneous	710	85	8.35	1990
Ecology	620	57	10.88	1983
Microbiology	567	61	9.30	1992
Biochemical research methods	408	89	4.58	1996
Virology	369	35	10.54	1993
Agriculture	345	78	4.42	1985
Agriculture, dairy and animal science	278	30	9.27	1990
Chemistry, analytical	269	28	9.61	1988
Toxicology	253	28	9.04	1979
Reproductive biology	249	30	8.30	1996
Reproductive systems	218	8	27.25	1990
Biometrics	186	8	23.25	1990
Agriculture, soil science	155	34	4.56	1991
Botany	148	9	16.44	1976

<sup>a</sup>Subject areas were developed by the Institute for Scientific Information (ISI), and are represented only for the 13,000 *Arabidopsis* articles found in ISI's database.

<sup>b</sup>*N* cites refers to the total number of scientific citations to *Arabidopsis* articles published within these subject areas.

<sup>c</sup>*N* papers refers to the total number of *Arabidopsis* papers published within this subject areas.

fuzzy search algorithm<sup>17</sup> to match all of the TAIR terms directly to the titles and abstracts of articles in the database. Furthermore, I used a probabilistic algorithm to annotate additional gene and protein names in articles (Chang, Schutze, & Altmann, 2003), and worked with a biology student, Christian Anderson, to build a typology of methods<sup>18</sup> and species that we matched independently into abstracts and titles. This matching process resulted 28,350 unique terms in nearly 400,000 publication-term matches with the 18,359 articles described above. I associated these publications and the terms they contain with the scientists that produce them and the academic research organizations (ROs) where they reside over time. From the term data,

I create and validate measures that capture the originality of a scientists' research activities for each year in which he or she publishes. To address the hypothesis that funding diversity shapes entrepreneurial creativity, I link these scientists – along with their newly minted scores for novelty – to the sources providing funding. Using fixed-effect panel models which statistically compare scientists with themselves over time, I predict how funding diversity from prior years affects the novel research activities in the present one. Then I use negative binomial models to examine the determinants of academic success and commercial relevance. In the following sections, I detail these steps regarding data organization, coding, and analysis.

### *Principal Investigators and Their Labs*

For the purpose of this study, I focus on those scientists who ran their own labs at any time within the period of study. Government grants title such scientists “principal investigators” (PIs), for their role in managing a proposed research program. I established the existence and location of *Arabidopsis* PIs and research labs through two separate sources of information. The first is provided by TAIR, which maintains a database of all individuals and organizations in the *Arabidopsis* community who have ever published *Arabidopsis* papers or used their extensive genetic information databases. TAIR data indicate the location and institutional affiliations of researchers and research labs that use *Arabidopsis*. Between 1974 and 2003, 5,725 PIs used *Arabidopsis* in their research, a third of whom reside in the U.S.

Where labs were not specified in TAIR, the bibliographical data from ISI and BIOSIS electronically lists the institutional locations for the authors of each article in the database. In the field of molecular plant biology, the last author on a research article is almost always (with 90 percent accuracy) a PI – the head of a research lab. Across articles, this rule allowed me to identify most of the remaining PIs, and then verify their status by searching department websites and examining professional association directories. I associated articles with specific universities and research institutes in the same manner described above for PIs. I used this same method to identify the number of nonPIs within each paper from which I derived the average size of each PI's laboratory.

I associated articles with specific universities and research institutes (research organizations – ROs) in the same manner described above for PIs. In total, 3,163 research organizations have sponsored science utilizing

*Arabidopsis*. I associated these ROs with a number of university characteristics from the World Education Database, and used them to classify and control for institution-type in the subsequent models. Table 2 lists the 25 most influential *Arabidopsis* scientists, ranked by article citations. The list highlights the mix of institutions, from elite private universities like Caltech, Stanford, and Harvard (i.e., Massachusetts General Hospital) to public, land-grant institutions like Michigan State, Washington State, and Agricultural Universit  Wageningen, a Dutch land-grant equivalent. Research institutes figure prominently, including the Salk Institute, Rockefeller University, and Britain’s John Innes. Fewer are government agencies like Japan’s RIKEN Genomic Sciences Center and France’s Institut Scientifique Vegetales (CNRS). Only one company researcher, John Ryals from Ciba Geigy, makes the list.

### *Novel Activities*

Following Schumpeter’s insight that novelty occurs through combination, I derive my measure of novel activities from the combinatorial novelty of biological terms linked to *Arabidopsis* articles (i.e., genes, proteins, species, methods, biological processes, molecular functions, cellular components, *Arabidopsis* developmental stages, and anatomical locations). To capture novel research activities, I measured the novel introduction and combination of (1) methods; (2) biological processes, molecular functions, and cellular components; and (3) *Arabidopsis* genes. When scientists introduce new methods or originally combine existing methods into a novel procedure, they generate a fresh stream of exploratory insights into the natural world. As Kuhn (1961) notes, the produce of new methods rarely squares directly with older findings . New findings may eventually fuel a fresh scientific paradigm, but initially leads the academic community to regard them with skepticism. The codes for biological processes, molecular functions, and cellular components stem from the Gene Ontology (GO). The GO is a logical hierarchy of scientific terms, from general to specific, developed by thousands of molecular biology labs to facilitate communication between scientists working on the same biological phenomena in different settings. When scientists combine these biological phenomena in novel ways, they propose theoretically unanticipated processes. Such investigations are risky, with low average and highly varying prospects for success. Scientists have not combined those elements in print for a reason: theory and unpublished experimental failures suggest no meaningful relationship between them.

**Table 2.** *Arabidopsis* Publishing Activity: Most Cited Authors, 1974–2002.

Author	Institution of Longest Duration	<i>N</i> Cites <sup>a</sup>	<i>N</i> Papers <sup>b</sup>	Cites/Papers	First Paper
Meyerowitz, Elliot M.	Caltech, Division of Biology	9,684	133	72.81	1984
Somerville, Chris R.	Michigan State University, Department of Energy Plant Research Lab	6,838	165	41.44	1979
Van Montagu, Marc	State University of Ghent, Vlaams Interuniversity Institute for Biotechnology, Department of Plant Genetics, Belgium	5,585	220	25.38	1979
Feldmann, Kenneth A.	University of Arizona, Department of Plant Science	4,621	92	50.23	1981
Ausubel, Fred M.	Massachusetts General Hospital, Department of Molecular Biology	4,385	84	52.20	1986
Chory, Joanne	Salk Institute for Biological Studies, Howard Hughes Med Institute	4,151	100	41.51	1986
Koornneef, Martin	Agricultural Universite Wageningen, Graduate School of Experimental Plant Science, Netherlands	3,902	111	35.15	1977
Yanofsky, Martin F.	University of California, San Diego, Department of Biology	3,663	51	71.82	1988
Quail, Peter H.	University of California, Berkeley, Department of Plant & Microbial Biology	3,319	69	48.10	1989
Chua, Nam Hai	Rockefeller University, Plant Molecular Biology Lab	3,123	99	31.55	1990
Shinozaki, Kazuo	RIKEN Genomic Sciences Center, Plant Molecular Biology Lab, Japan	3,038	163	18.64	1992
Davis, Ronald W.	Stanford University, School of Medicine, Department of Biochemistry	2,993	29	103.21	1988
Dean, Caroline	John Innes Institute, Department of Cell & Developmental Biology, UK	2,959	81	36.53	1991
Bowman, John L.	Monash University, Department of Biological Sciences, Australia	2,808	30	93.60	1987

**Table 2.** (Continued)

Author	Institution of Longest Duration	<i>N</i> Cites <sup>a</sup>	<i>N</i> Papers <sup>b</sup>	Cites/Papers	First Paper
Ecker, Joe	University of Pennsylvania, Institute for Plant Sciences	2,681	63	42.56	1988
Goodman, H. Maurice	Massachusetts General Hospital, Department of Molecular Biology	2,651	74	35.82	1972
Deng, Xing-Wang	Yale University, Department of Molecular, Cellular & Developmental Biology	2,631	90	29.23	1991
Inze, Dirk	State University of Ghent, Vlaams Interuniversity Institute for Biotechnology, Department of Molecular Genetics, Belgium	2,617	130	20.13	1986
Estelle, Mark	University of Indiana, Department of Biology	2,537	65	39.03	1985
Yamaguchi-Shinozaki, Kazuko	RIKEN Genomic Sciences Center, Plant Molecular Biology Lab, Japan	2,464	60	41.07	1992
Giraudat, Jerome	Institut Scientifique Vegetales, CNRS, France	2,434	41	59.37	1989
Cashmore, Anthony R.	University of Pennsylvania, Institute for Plant Sciences, Department of Biology	2,359	52	45.37	1988
Jurgens, Gerd	University of Tubingen, Germany	2,195	70	31.36	1991
Ryals, John	Ciba Geigy Corporation, Agricultural Biotechnology Research Unit	2,161	31	69.71	1992
Browse, John	Washington State University, Institute of Biological Chemistry	2,152	77	27.95	1985

<sup>a</sup>*N* cites refers to the total number of scientific citations to *Arabidopsis* articles published by this author.

<sup>b</sup>*N* papers refers to the total number of *Arabidopsis* papers published by this author.



But because unexpected, the success of such novel compounds can open dramatic new doors for science. When scientists discover new genes, or combine known genes in investigations of novel pathways, they engage in a more modest form of exploration that can lead to new, unexpected findings.

For the purposes of this analysis, I consider novel introductions and combinations of biological terms as a function of the number of times that pairs of biological terms have ever been used together before and the number of times they have ever been used before at all. Specifically, I measure novelty within each publication as

$$\frac{\sum_{i=1}^n \sum_{j=i+1}^n \left[ 1 - \left( \frac{\sum_{ij=1}^N (\chi_i \cap \chi_j)}{\min \left( \sum_{i=1}^N \chi_i \wedge \sum_{j=1}^N \chi_j \right)} \right) \right]}{[n/2]}$$

where  $i$  and  $j$  are terms in article  $x$ ,  $n$  the total number of terms in  $x$ , and  $N$  the total number of articles within the entire field of *Arabidopsis* published prior to article  $x$ . Hence, novelty is composed of one minus the frequency with which any two terms have been used together in the previous literature divided by the number of times they could have been combined – the frequency of the least frequent term. This measure is averaged for the pairwise combinations of all terms within an article  $x$ . If a term, such as a gene, is mentioned for the first time in an article, it is given the weight of a completely novel combination even if no other relevant terms are mentioned. If I coded no relevant terms from which to calculate a score, I gave no score to the paper. In order to nudge these measures to normality for parametric analysis, I took the Box–Cox transformation of each.<sup>19</sup>

For example, consider the abstract for the “Shatterproof” article illustrated in Fig. 2 by Yanofsky et al. Only two genes, SHP1 and SHP2, are mentioned in the article. If these genes were mentioned together in five previous articles, and the least frequent (SHP2) was mentioned in seven total articles, then the gene novelty score for the article would equal 0.287. I generated a score for each publication and then associated the score with the PIs and ROs that produced it.

I tested the validity of these originality measures against a database of expert rankings, entitled “Faculty of 1000.” In this database, prominent life scientists nominate their favorite new articles in the field, score them, and classify them as “confirmation of hypothesis,” “new finding,” “contentious finding,” “methodological innovation,” “new hypothesis,” or some combination of these. The categories may be understood as anchoring a continuum

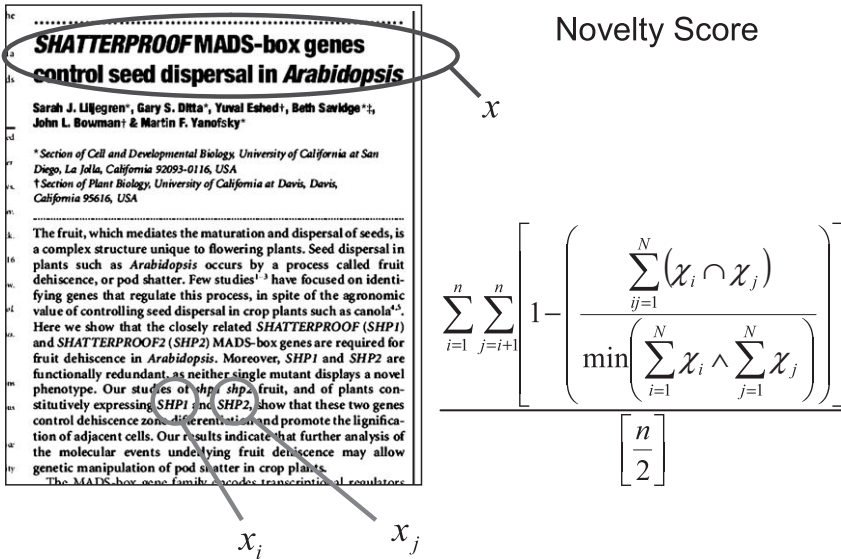


Fig. 2. Faculty of 1000 Database: Novel Techs Correlates with Tech Advance (0.64,  $p < 0.05$ ) and New Hypothesis (0.77,  $p < 0.01$ ).

of risk, from the expected developments of normal science (“confirmation of hypothesis”) to the less expected developments (“new finding”) to the groundbreaking (“contentious,” “methodological innovation,” and “new hypothesis”). Three hundred and ninety-one articles in Faculty of 1000 use *Arabidopsis* and are located in my database, and nearly 300 have novelty scores of some type. Table 3 lists pairwise correlations between Faculty of 1000 classifications and my corrected measures of original scientific activities. My technical novelty measure correlates at 0.65 ( $p < 0.05$ ) with the Faculty of 1000 “technical advance” code and at 0.78 ( $p < 0.01$ ) with the “interesting hypothesis” code. My measure of GO term novelty (novel combinations of biological processes, molecular functions, and cellular components), fails to positively vary with the “interesting hypothesis” code, but nevertheless correlates at 0.13 ( $p < 0.05$ ) with “technical advance.” My gene novelty correlates most highly (0.11) with the “new finding” score, but below accepted thresholds of statistical significance ( $p = 0.21$ ). In sum, my measures correspond most closely with the expected Faculty of 1000 indicators, although they more efficiently indicate the most speculative (“interesting hypothesis” and “technical advance”) than the more modestly original (“new finding”).

**Table 3.** Correlation Coefficients of Faculty1000<sup>a</sup> Nominations with Novelty Scores.

Faculty of 1000 Nominations	“Important Confirmation”	“New Finding”	“Technical Advance”	“Interesting Hypothesis”
Gene novelty	-0.11	0.11	-0.04	-0.06
GO term novelty	-0.10 <sup>‡</sup>	0.03	0.13*	-0.15*
Methodological novelty	0.19	-0.34	0.64*	0.78**

Note:  $N = 12-243$ .

\*\*Pairwise coefficients significant at 1%.

\*Pairwise coefficients significant at 5%.

<sup>‡</sup>Pairwise coefficients significant at 20%.

<sup>a</sup>Faculty of 1000 paper scoring and nominations viewable at <http://www.f1000biology.com>.

### Success

I measured the concept of success in two ways. In the first, I examine the number of paper citations that a lab’s own papers receive over time. Academic departments and deans routinely use this measure to evaluate lab contributions. In a second, more oblique measure, I consider the lab’s commercial relevance by counting the patent citations that a lab’s papers receive over time. This second measure will allow us to determine whether the factors driving novelty or those driving success push institutions from academic science into business.

### Funding Diversity

In order to document funding received by academic labs, I scanned the acknowledgments of all *Arabidopsis* papers published between 1975 and 2000 in the SciSearch database (8,400 documents) and parsed these acknowledgments for the company collaborations and funding they mention. Fig. 3 provides an example for the *Nature* article “Shatterproof MADS-box genes control seed disbursal in *Arabidopsis*,” coauthored by PIs John Bowman, Martin Yanofsky, and Beth Savidge from UC San Diego and UC Davis. Because of the funding mention in the acknowledgment, I code all three as having a tie to Monsanto, the NSF, NIH, and the University of California Bio-Star program in the year 2000. I then classified funds by type, including government funds (e.g., NIH, NSF), company funds (e.g., Monsanto), nonprofit funds (e.g., Rockefeller Foundation), and university funds (e.g., California Bio-Star). I measure a scientific lab’s funding

**SHATTERPROOF MADS-box genes control seed dispersal in *Arabidopsis***

Sarah J. Liljegren\*, Gary S. Ditta\*, Yuval Eshed†, Beth Savidge\*, John L. Bowman† & Martin F. Yanofsky\*

\* Section of Cell and Developmental Biology, University of California at San Diego, La Jolla, California 92093-0116, USA  
 † Section of Plant Biology, University of California at Davis, Davis, California 95616, USA

...ration and dispersal of seeds, is ...  
 ...ing plants. Seed dispersal in ...  
 ... occurs by a process called fruit ...  
 ... dehiscence, or pod shatter. Few studies<sup>1-3</sup> have focused on identifying genes that regulate this process, in spite of the agronomic value of controlling seed dispersal in crop plants such as canola<sup>4,5</sup>. Here we show that the closely related SHATTERPROOF (SHP1) and SHATTERPROOF2 (SHP2) MADS-box genes are required for fruit dehiscence in *Arabidopsis*. Moreover, SHP1 and SHP2 are functionally redundant, as neither single mutant displays a novel phenotype. Our studies of *shp1 shp2* fruit, and of plants constitutively expressing SHP1 and SHP2, show that these two genes control dehiscence zone differentiation and promote the lignification of adjacent cells. Our results indicate that further analysis of the molecular genetic n...  
 The M...  
 P2  
 tic

**Acknowledgements**

We thank C. Ferrandiz, A. Sessions, S. Kempin, A. Pinyopich, E. Alvarez-Buylla, J. Spence and N. Harris for helpful discussions; K. Feldmann and his lab for the gift of DNA and seeds from his T-DNA insertional collection; R. Martienssen for providing GT140 seed; and S. Guimil, T. Khammuangkhone, C. Chien, H. Cartwright and A. Roeder for assistance with the *shp1* and *shp2* mutagenesis screens. This work was supported by grants from the National Science Foundation, the National Institutes of Health, Monsanto Company and the University of California BioSTAR programme.

Correspondence and requests for materials should be addressed to M.F.Y. (e-mail: marty@ucdavis.edu).

30. Marquez, M. A., Sidorow, L. N., Lee, I. S. & Weigel, D. LEAF1 expression and flower initiation in *Arabidopsis*. *Development* 124, 3835–3844 (1997).

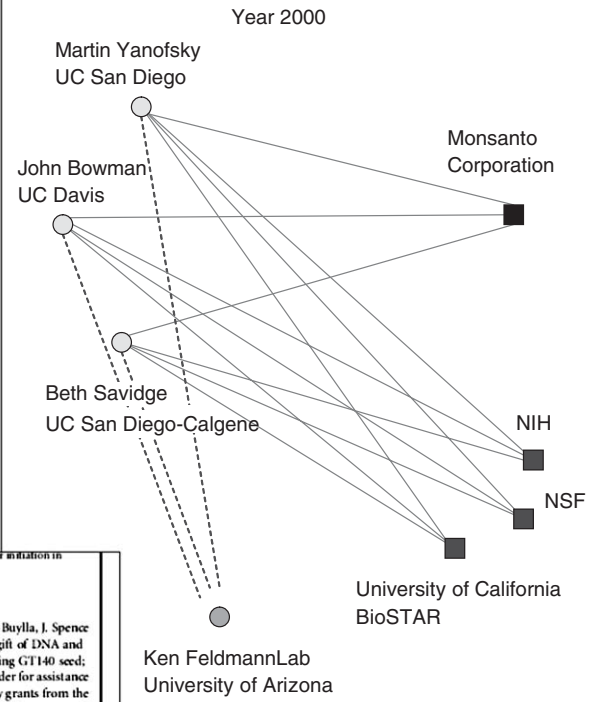


Fig. 3. Lab-Level Funding Ties.

diversity by the computing one minus the Herfindahl index of its research relationships, by type. The Herfindahl index was initially used to measure the relative concentration of market share within industries and equals  $\sum_{i=1}^n (s_i^2)$ , where  $s_i$  represents the percentage of a lab's funding from any one source type. For the Bowman, Yanofsky, Savidge paper, its funding diversity sums its government funding ( $0.5^2$ ), university funding ( $0.25^2$ ), and corporate funding ( $0.25^2$ ), then subtracts this total from one, leaving their labs – if they produced only this paper in 2000 – with funding diversity scores of 0.625.

### Controls

I controlled for several qualities of the scientists, their organizational location, and their scientific subfields that could contribute to a speculative or confirmatory stance toward research. Among scientist-level factors, I controlled for the length of time that a PI had ever worked with *Arabidopsis*, which I call “tenure” in the tables and subsequent discussion. Scientists specialize, suggesting that their research will appear somewhat less original over time. I also controlled for the square of tenure to capture the diminishing effects of tenure over time. The originality of PIs' research will almost inevitably drop more after their first year of research than just before their last.

I also control for the commercial relevance of the particular regions of *Arabidopsis* science that labs explore. If researchers know that certain sets of experiments, which may appear theoretically speculative or original, are relevant to commercial products, this and not novel thinking per se could account for their choices. In order to establish this, I first grouped all of the *Arabidopsis* publications into a set of 17 subfields by clustering all 18,563 articles on the 28,350 unique scientific terms that linked them together.<sup>20</sup> I used the principal component of the matrix of articles by terms to divisively partition the data into substantively meaningful subfields (Boley, 1997). With a team of biologists, I examined these clusters to select the appropriate number, reunited related clusters separated by gross initial divisions, and characterized their scientific content. Subfields vary in size, from the large cluster studying organogenesis, the scientific area concerned with plant development and the differentiation of cells into distinct organ tissues, to the much smaller cluster of commercial disease resistance.

I then identified all of the patent citations to publications within each cluster and flagged those clusters with a higher-than-median patent citation per publication as potentially commercial. Note that less commercial fields

like “photosynthesis” examine one of the most biologically significant and distinctive aspects of plant metabolism, while more commercial fields like “terpenes” examine a class of protein substances that are biologically inessential, but have historically been very useful as the basis for industrial substances such as turpentine and rubber. Interviews confirmed that the subfields listed as more and less commercial correspond with assessments by firms. One scientist at a major chemical company clarified:

Plant development is a big field in academic fields. How do plants develop? How do leaves? How is the structure formed? [organogenesis cluster; nonapplied] That is not really something that ChemCo has been very interested in, because we don't see it as having direct applicability, whereas if someone is interested in biochemical pathways [receptors and phosphorylation cascades clusters; both applied], and metabolism [nutrient uptake; applied], and how those kinds of things are controlled – how gene expression is controlled – those kinds of things, we're much more interested in.

I include dummy variables to control for RO types including government agencies (e.g., Agricultural Research Service at the USDA), nonprofit research institutes (e.g., Cold Spring Harbor), hospitals (e.g., Scripps Clinic), universities with agricultural schools or influences (e.g., Texas A&M – “Agricultural & Mechanical”), all compared against the reference category of nonagriculturally influenced universities (e.g., Harvard). I also included dummies for regions where these ROs were located. In addition, I controlled for organizational factors that might directly influence creativity in research. I include an indicator of the diversity of subjects studied using *Arabidopsis* within the organization, calculated as one minus the Herfindahl index of research from each scientific cluster within that organization. I include a related indicator of the diversity of departments studying *Arabidopsis*, computed as one minus the Herfindahl index of research from each department studying *Arabidopsis* within that organization. As scientific labs interact with other *Arabidopsis* researchers in their organizations who study different phenomena or work in different departments with different resources and concerns, their subfield and department diversity measures rise and they may be more likely to link insights from one area to another.

Finally, I controlled for the diversity of organization-level funding. Every public university and research institute in the sample, by definition, receives government support, and most of the nonprofit ones do as well. As such, I captured organization-level funding diversity indirectly, by measuring the number of industry funding ties it received in that year. With the complete list of companies that co-author *Arabidopsis* papers, invent patents that cite *Arabidopsis* papers, and receive mention in *Arabidopsis* paper acknowledgments, I organized a group of students to search the news databases of

Lexis–Nexis and related databases for instances where each company name co-occurred in the same article with “institute” or “university.” We then recorded the funding relationships linking these companies with the organizations in my sample. For example, a 1998 news article in the San Francisco Chronicle mentions a \$25 million relationship between Novartis and UC Berkeley’s Department of Plant and Microbial Biology from 1998 until 2003. For each of those years, I coded one organization-level industry tie. In addition to explicit mentions of funding, I coded an industry tie when licensing and marketing relationships were mentioned.

### *Modeling Strategy*

I use panel models to take advantage of the dynamic nature of my data. In order to evaluate my hypothesis about diverse funding and novel activities, I associate scientists’ industry ties with their subsequent scientific activity. As a conservative test of my primary hypothesis, I compare researchers in periods when their funding is more diverse to themselves in periods when it is less so. Hausman tests also favor a fixed-effects modeling strategy by highlighting significant differences between the fixed- and random-effects models. This choice does not, however, preclude the use of organization variables, as a substantial proportion of *Arabidopsis* scientists move through organizations and subfields across their careers.

The unit of analysis is PI-publication-years. For example, if Martin Yanofsky, the head of a large lab at UC San Diego published one article in 1984, three articles in 1986, and two articles in 1987, he would receive three separate entries in the regression model corresponding to 1984, 1986, and 1987. For 1986 and 1987, when he published more than one article, I take the mean values of Yanofsky’s novelty and independent variables across all articles published that year. This yields unbalanced panels with some gaps corresponding to years that scientists do not publish any *Arabidopsis* articles.

The basic model, specifying fixed effects and estimated using OLS, is

$$y_{it} = \alpha_{it} + x_{it}\beta + v_i + \varepsilon_{it}, \quad y \sim N()$$

where time  $t$  is measured in years, each  $i$  a PI or RO, and  $v_i + \varepsilon_{it}$  the residual with  $v_i$  representing the PI-specific component that differs only between and not within scientists.

In order to examine the factors affecting venture success, I used conditional, fixed-effect negative binomial models to account for the overdispersion

of paper and patent citations (Hausman, Hall, & Grilliches, 1984).<sup>21</sup> In these models, fixed effects refers not to the  $x\beta$  term in the model but to the dispersion parameter, forcing it to be the same within labs, but allowing it to take on any value across them. The negative binomial model is a generalization of the Poisson model that allows its dispersion parameter to vary by group, but remains constant within group thereby accounting for persistent lab-level effects.

## DESCRIPTIVE STATISTICS

Table 4 lists descriptive statistics for the variables described in previous sections. Because some *Arabidopsis* papers – and so some PI-publication-years – do not have enough information to compute certain types of novelty or persistence scores, the sample used in some models is larger than in others. As a result, the table includes variable means, standard deviations, minimums, and maximums for all PI-publication-years in the largest, most inclusive analysis. It also includes the maximum deviation from the mean variable value in the largest sample and its value in any subsample analyzed in the paper. These maximum deviations highlight that subsamples are very similar.

Means for the transformed scientific novelty scores range from  $-0.237$  for methodological novelty to  $-0.066$  for gene novelty, which correspond to raw scores of 0.729 and 0.919, respectively. From the sheer number of *Arabidopsis* genes, it was much easier for scientists to discover and originally combine them than it was for them to introduce or combine new methods. Scientific papers received an average of 34.35 citations per paper, but with a standard deviation of 45.11 and maximum of 748, some received many more than others. Patent citations to papers spread even more unevenly, with a mean of 1.4, a standard deviation of 5.26, and a maximum of 153.

Forty-one percent of the scientific activity took place in the U.S., while somewhat smaller proportions – 25 and 17 percent – took place in Continental Europe, Britain, and its (former) commonwealths – Canada, Australia, and New Zealand. A substantial amount of *Arabidopsis* research activity, 16 percent, occurred in Asia. The remaining, residual activity took place in South America, the Middle East, and Africa. Of the research activity examined in the study, 39 percent took place in universities hosting agricultural schools or departments (e.g., UC Davis) and 38 percent did not (e.g., The University of Chicago). Nearly 12 percent occurred in nonprofit research institutes (e.g., Cold Spring Harbor) and 7 percent in government laboratories



**Table 4.** Descriptive Statistics of Variables Used in Largest Analysis.

	Mean	Standard Deviation	Maximum Deviation in Analyses*	Minimum	Maximum
<b>Novelty</b>					
Genes ((Gene Novelty <sup>6.779</sup> -1)/6.779)	-0.064	0.062	n/a	-0.148	0
GO terms ((Term Novelty <sup>2.418</sup> -1)/2.418)	-0.215	0.087	n/a	-0.438	0
Methods ((Method Novelty <sup>1.906</sup> -1)/1.906)	-0.237	0.149	n/a	-0.513	0
<b>Citations</b>					
Paper citations	34.354	45.110	n/a	0	748
Patent citations	1.400	5.263	n/a	0	153
<b>Region (0/1)</b>					
U.S.	0.408		0.026	0	1
Europe	0.250		0.024	0	1
Asia	0.160		0.092	0	1
Britain and Commonwealths (e.g., Australia, Canada)	0.166		0.034	0	1
<b>Institutional context</b>					
Hospital	0.005		0.006	0	1
Research institute	0.119		0.004	0	1
Company	0.036		0.011	0	1
Government	0.072		0.006	0	1
University w/o agriculture	0.375		0.002	0	1
University with agriculture	0.393		0.001	0	1
Subfield diversity	0.800	0.232	0.013	0	0.910
Department diversity	0.278		0.011	0	1
Organization-level funding diversity (industry support)	0.126	0.409	0.021	0	4
<b>Lab context</b>					
Tenure	2.336	3.130	0.446	0	26
Tenure <sup>2</sup>	15.257	37.546	3.170	0	676
Commercial subfield (0/1)	0.366		0.016	0	1
Count of ties	1.610	1.863	0.222	0	22
Ln number in lab	1.374	0.689	0.170	0	3.401
Funding diversity	0.170	0.237	0.020	0	0.750

Note: N = 4,593.

\*Maximum deviation from mean for any covariate in the analysis of any dependent variable in this paper.

(e.g., Agricultural Research Service of the USDA). Much smaller amounts occurred in companies (3.6 percent) and hospitals (0.5 percent).

*Arabidopsis* research activities within these organizations commonly cut across multiple subfields, generating a high average subfield diversity index (0.8) with a low standard deviation (0.164). The diversity of departments performing *Arabidopsis* research within these organizations is much lower (0.278), and varies more widely (s.d. 0.232). Only 12.6 percent of the ROs received press mentioning company sponsorship, indicating the cultivation of diverse sponsors.

Many authors were relatively new users of *Arabidopsis*, with an average of 2.3 years experience publishing about the plant. The standard deviation of 3.13 years, however, suggests that others – scientists like Elliot Meyerowitz, Chris Somerville, and Marc Van Montague from Table 2 – persisted much longer. Thirty-seven percent of the *Arabidopsis* research activity took place in commercially relevant subfields – in areas with a history of stimulating patentable inventions.

Laboratories ranged in size from one publishing author – a solo PI – to 30, with an average of just under four authors. When logged, these numbers translate into 0, 3.40, and 1.37 found in the table. Funding diversity varied widely, with a mean of 0.17 and a standard deviation of 0.24. Many investigators took most of their money from a single source – from governments or nonprofit foundations. Fewer added industry funding into the mix.

Table 5 lists the correlations of the central covariates of the study. While gene and GO term novelty correlate positively (0.11,  $p < 0.01$ ), neither post a significant relationship with methodological novelty. Similarly, paper and patent citations correlate highly with one another (0.41,  $p < 0.01$ ) and at a lower level with gene and GO term novelty (0.03–0.08,  $p < 0.05$ ), but neither maintain a patterned relationship with methodological novelty. Scientific activity within the U.S. negatively relates with gene novelty ( $-0.05$ ,  $p < 0.01$ ), but positively relates with GO term and methodological novelty (0.04 and 0.13,  $p < 0.01$ ). This suggests that activity in the U.S., relative to other parts of the world, is methodologically more original, producing more speculative results (Evans, 2006). European, Asian, and Anglo *Arabidopsis* scientific activity more highly correlates with gene novelty, the indicator of moderate originality that results in new, sometimes unexpected findings, and negatively correlate with GO term novelty. Asian research activity is the antipodes of the U.S. with a strong, negative relationship with methodological novelty.

Note the positive relationship between companies and the original research activities they host. Theory, the focus of academic science, acts to improve the productivity of experiment by reducing the vast number of

possible combinations to the most expected ones. In its attempt to restrict these intelligently, however, theory is also a constraint, holding scientists back from under-theorized combinatorial possibilities (Evans, 2006). Companies, with a focus on products, lack this restraint.

In terms of research success, universities with agriculture and government labs relate negatively with future citations, while research institutes, companies, and hospitals, often seen as the most applied organizations, produce the most well received science. The company effect is highest, and may result from not needing to publish less important papers (Powell & Owen-Smith, 1998). Not surprisingly, companies as research hosts maintain the only significantly positive relationship with future patent citations (0.10,  $p < 0.01$ ).

Tenure with *Arabidopsis* correlates positively with methodological novelty (0.08,  $p < 0.01$ ), suggesting that researchers who persist with the plant tend to move through new technologies in studying particular scientific phenomena. While research activity within a commercially relevant subfield maintains no significant bivariate relationships with originality, it relates positively with both academic and commercial success (0.06 and 0.07,  $p < 0.01$ ). This underscores an influential point made by Ronald Stokes (1997) that scientists may successfully search for fundamental principles in commercially relevant settings.

The presence of organization-level funding diversity – reports of industry sponsorship from the business press – correlates negatively with both gene and GO term novelty. Perhaps companies, in pursuing organization-level relationships with academic science, select institutional partners based on their prior success. As expected, funding diversity within the lab posts a significant, positive correlation with methodological novelty (0.07,  $p < 0.01$ ), although its relationships with gene and GO term novelty are small and insignificant. Lab size has the opposite effect on methodological novelty, suggesting that PIs can sustain creative research activities within artisanal workshops more easily than in scaled-up research factories. Funding diversity also has a small, positive relationship with academic success (0.03,  $p < 0.01$ ). Lab size has an even stronger effect (0.06,  $p < 0.01$ ), suggesting that even if less novel, larger stables of publishing scientists attract more attention. Not included in the table or models is a measure of the total number of funding ties maintained by a lab. Because this measure correlates highly with funding diversity (0.57,  $p < 0.01$ ), it produced multi-collinearity in the models and inflated the standard errors. For conceptual clarity, and to identify the independent effect of funding diversity, it was dropped from the analysis.

Other correlations add additional color to this portrait of scientific activity. Correlations among countries suggest the flows of scientists across national

**Table 5.** Correlation Coefficients of Selected Variables<sup>a</sup>.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Gene novelty	1.00**									
(2) GO term novelty	0.11**	1.00**								
(3) Method novelty	0.05	0.04	1.00**							
(4) Paper citations	0.03*	0.05**	0.10	1.00**						
(5) Patent citations	0.08**	0.05**	0.05	0.41**	1.00**					
(6) U.S.	-0.05**	0.04**	0.13**	0.13**	0.08**	1.00**				
(7) Europe	0.05**	-0.02†	0.00	-0.06**	-0.03**	-0.48**	1.00**			
(8) Asia	0.03*	-0.02†	-0.15**	-0.11**	-0.04**	-0.36**	-0.25**	1.00**		
(9) Britain, Australia, New Zealand	0.01**	-0.01	-0.01	0.00	-0.03*	-0.37**	-0.26**	-0.19**	1.00**	
(10) Hospital	-0.02	0.00	0.03	0.03**	0.00	0.09**	-0.05**	-0.03**	-0.03**	1.00**
(11) Research institute	0.00	-0.03**	-0.01	0.05**	0.01	-0.18**	-0.01	0.02	0.22**	-0.03*
(12) Company	0.05**	0.04**	0.02	0.08**	0.10**	0.08**	0.01	-0.05**	-0.07**	-0.01
(13) Government	0.00	0.04**	-0.03	-0.04**	-0.02	-0.12**	0.09**	0.04**	0.01	-0.02†
(14) University w/o agriculture	-0.01	-0.02†	0.04	0.01	0.00	-0.13**	0.22**	-0.09**	0.01	-0.05**
(15) University with agriculture	-0.01	0.01	-0.04	-0.06**	-0.02†	0.27**	-0.27**	0.07**	-0.12**	-0.05**
(16) Subfield diversity	-0.01	0.02†	0.01	0.08**	0.03**	0.13**	-0.12**	-0.03**	0.01	0.01
(17) Department diversity	0.05**	0.01	-0.02	-0.05**	-0.01	0.12**	-0.02†	0.00	-0.12**	0.06**
(18) Organization funding diversity	-0.03*	-0.03**	0.00	-0.01	0.01	0.21**	-0.14**	-0.06**	-0.04**	0.02*
(19) Tenure	-0.12	-0.04	0.08**	0.00	-0.01	0.10**	-0.03*	-0.06**	-0.04**	0.04**
(20) Tenure <sup>2</sup>	-0.10	-0.01	0.05†	0.00	-0.01	0.09**	-0.01	-0.05**	-0.04**	0.03**
(21) Commercial subfield	0.05	0.02	0.00	0.06**	0.07**	0.00	0.02†	-0.01	-0.03*	0.01
(22) Funding diversity	-0.06	-0.04	0.07*	0.03**	0.01	-0.03†	0.01	0.08**	-0.04**	0.00
(23) Ln number in lab	-0.07	-0.06	-0.05*	0.06**	0.01	-0.07**	0.03**	0.08**	-0.01	0.03

<sup>a</sup>Pairwise Pearson correlations, *N*'s from 746 to 8873.

\*\*Significant at 1%.

\*Significant at 5%.

†Significant at 10%.

---

(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)
1.00**												
-0.06**	1.00**											
-0.08**	-0.05**	1.00**										
-0.27**	-0.14**	-0.22**	1.00**									
-0.24**	-0.13**	-0.20**	-0.61**	1.00**								
-0.03**	-0.11**	0.00	-0.12**	0.19**	1.00**							
-0.25**	-0.15**	-0.07**	-0.17**	0.42**	0.33**	1.00**						
-0.04**	-0.03**	0.01	-0.10**	0.14**	0.09**	0.06**	1.00**					
-0.01	-0.05**	-0.04**	-0.03*	0.04**	0.13**	0.02 <sup>†</sup>	0.15**	1.00**				
0.00	-0.04*	-0.02 <sup>†</sup>	-0.03*	0.03**	0.06**	0.00	0.12**	0.88**	1.00**			
0.01	0.05**	-0.01	-0.02 <sup>†</sup>	0.00	0.05**	0.04**	0.03*	0.00	-0.01	1.00**		
0.03*	0.01	-0.01	0.03*	-0.07**	0.04**	-0.03*	0.03*	0.10**	0.08**	0.02	1.00**	
0.06**	0.03**	-0.03*	-0.02 <sup>†</sup>	-0.04**	0.07**	0.02	0.06**	0.20**	0.12**	0.06**	0.17**	1.00**

---

borders. Scientific activity in North America, Europe, and Britain is most likely to move to and from Asia ( $-0.36$ ,  $-25$ , and  $-0.19$ ,  $p < 0.01$ ), suggesting the frequency with which Asian scientists train in and travel between Western labs and their home institutions. The positive correlation between the U.S. and hospitals ( $0.09$ ,  $p < 0.01$ ) suggests the importance of the research hospital in the American system, and its relative unimportance in other parts of the world. An even stronger, positive association between the U.S. and agriculturally influenced universities recalls the Morrill Act of 1862 that sponsored land-grant institutions in nearly every state. Together, these relationships paint a picture of American exceptionalism in research that sponsors strong ties between academic knowledge and focused practical needs (Rosenberg & Nelson, 1994). Positive correlations linking Europe and Asia with government labs suggest the importance of government funding *and management* of research in those regions. The organization-level diversity correlations highlight that while the U.S. houses more diverse kinds of research, types of departments, and flavors of funding than other countries, U.S. institutions are also individually more diverse on each dimension than their foreign counterparts.

Interestingly, hospitals tend to correlate most highly with organization-level diversity scores. As institutions focused ultimately on medical innovation, and housing a diverse array of life science departments, they are the location of the most even distribution of different kinds of bioscience. Not surprisingly, companies correlate positively with commercially relevant research ( $0.05$ ,  $p < 0.01$ ), and their PIs tend to be dilettantes, flirting with but not persisting in *Arabidopsis* research ( $-0.05$ ,  $p < 0.01$ ). Nonprofit and government research institutes maintain the highest correlations with funding diversity and lab size, while university labs are smaller and can get by with fewer and less diverse sources of such funds.

Commercially oriented subfields – like commercial disease resistance – post high correlations with all measures of organization-level diversity. Lab-level funding diversity also correlates positively with institution-level funding diversity ( $0.03$ ,  $p < 0.01$ ), although the relationship is small, suggesting that companies use a different calculus for entering institution than lab-level relationships. Funding diversity and lab size, themselves correlated at  $0.17$  ( $p < 0.01$ ), associate positively with the duration that scientists have researched with *Arabidopsis*.

## RESULTS

Table 6 lists results for the linear models predicting experimental novelty. Note that because these are fixed-effects models comparing scientists to

**Table 6.** Fixed-Effects Models of Lab Characteristics on Novelty.

	Gene Novelty		Term Novelty		Method Novelty	
Region (U.S. baseline)						
Europe	-0.023 (0.019)	-0.008 (0.021)	-0.015 (0.016)	-0.004 (0.019)	-0.504 (0.277) <sup>†</sup>	-0.900 (0.357)*
Asia	0.014 (0.020)	0.004 (0.024)	-0.038 (0.017)*	-0.048 (0.022)*	0.106 (0.088)	0.211 (0.109) <sup>†</sup>
Britain and commonwealths (e.g., Australia, Canada)	-0.011 (0.021)	-0.007 (0.026)	0.018 (0.017)	0.037 (0.021) <sup>†</sup>	0.000 (0.000)	0.000 (0.000)
Institutional context (university w/o agriculture baseline)						
Hospital	0.003 (0.029)	-0.000 (0.032)	0.011 (0.032)	0.017 (0.033)	-0.216 (0.362)	-1.088 (0.957)
Research institute	0.007 (0.012)	-0.006 (0.015)	-0.009 (0.011)	-0.006 (0.014)	-0.029 (0.087)	0.050 (0.095)
Company	0.012 (0.017)	0.021 (0.020)	0.014 (0.015)	0.008 (0.018)	-0.068 (0.106)	-0.054 (0.134)
Government	0.018 (0.015)	-0.005 (0.019)	0.018 (0.014)	0.033 (0.017) <sup>†</sup>	0.059 (0.089)	0.077 (0.101)
University with agriculture	-0.004 (0.010)	0.002 (0.012)	-0.013 (0.009)	-0.001 (0.011)	0.004 (0.072)	0.019 (0.079)
Subfield diversity	-0.024 (0.027)	-0.062 (0.036) <sup>†</sup>	0.009 (0.022)	-0.012 (0.025)	-0.054 (0.187)	-0.059 (0.223)
Department diversity	0.024 (0.018)	0.008 (0.022)	0.005 (0.016)	0.006 (0.019)	0.069 (0.127)	0.218 (0.151)
Industry support	0.004 (0.004)	0.001 (0.004)	0.003 (0.004)	0.000 (0.004)	-0.002 (0.022)	0.006 (0.024)
Lab context						
Tenure	-0.007 (0.001)**	-0.006 (0.001)**	-0.007 (0.001)**	-0.007 (0.001)**	-0.001 (0.007)	-0.000 (0.008)

*Table 6. (Continued)*

	Gene Novelty		Term Novelty		Method Novelty	
Tenure <sup>2</sup>	0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.001)	0.000 (0.001)
Commercial subfield	0.009 (0.004)*	0.010 (0.005) <sup>†</sup>	0.002 (0.004)	0.003 (0.004)	-0.014 (0.025)	-0.014 (0.029)
Funding diversity		0.008 (0.008)		-0.003 (0.007)		0.083 (0.043)*
Ln number in lab		-0.014 (0.004)**		-0.008 (0.003)**		-0.008 (0.022)
Constant	-0.034 (0.024)	0.019 (0.032)	-0.198 (0.020)**	-0.178 (0.024)**	-0.115 (0.160)	-0.113 (0.211)
Observations	2,571	2,571	4,593	4,593	869	869
Number of pi_id	1,425	1,425	2,364	2,364	625	625
R <sup>2</sup>	0.081	0.096	0.055	0.075	0.026	0.073

*Note:* Standard errors in parentheses.

One-tailed tests for funding diversity and lab size and two-tailed tests for other coefficients.

\*\*Significant at 1%.

\*Significant at 5%.

<sup>†</sup>Significant at 10%.



themselves, the coefficients reflect only those cases in which scientists changed over time. For each form of originality, the table presents nested models with and without the lab-level variables for lab size and funding diversity. In each case, the addition of lab diversity and size substantially increases the *R*-squared of the models, from 25 percent in the case of gene novelty to over 200 percent for method novelty. For scientists whose labs shifted between regions, regional differences had no effect on gene novelty, my indicator of moderate originality. Scientific activity in Asia, compared with work performed in the U.S. by the same PIs, was somewhat less original in the use and combination of GO terms ( $-0.048$ ,  $p < 0.01$ ). Activity in Europe was substantially less original than the U.S. in the use and combination of methods ( $-0.9$ ,  $p < 0.01$ ), my indicator of radical innovation.

Institution types and characteristics, as they differed within the experience of individual PIs, had little effect on research originality. Governments incited a bit more GO-term originality than other research locations (0.033,  $p < 0.10$ ), while institutions housing diverse subfields of *Arabidopsis* research negatively influenced the originality with which research labs discovered and combined genes experimentally. In all models, tenure has a negative effect on originality, suggesting that the most original work takes place early in a scientist's career. Once a lab begins on a particular scientific path, even if it is relatively well trodden by others, it rarely changes to another. The pressures to achieve "brand recognition" required by tenure, and the skill and technology investments required by any branch of molecular biology makes a scientists' later work less original than her earlier work, if only because she has already done some version of it.

Research in more commercial subfields increases labs' originality in gene exploration, presumably in order to isolate and patent economically important gene products. In all three models, lab size posts a negative effect on novelty, foreshadowed by correlations in Table 5. When computed through the transformations of gene novelty and lab size, this means that if labs increase in size by one standard deviation (two publishing scientists), the likelihood that papers they produce will combine genes in original ways decreases by 15 percent and GO terms by 2 percent. Whether large labs are the cause of normal science, or the consequence of an unmeasured strategy to perform it, they are associated with a mode of production that is intellectually less entrepreneurial.

Funding diversity posts a statistically insignificant influence on gene and GO-term originality, but a positive effect on methodological originality (0.083,  $p < 0.05$ ).<sup>22</sup> Labs with funding diversity at one standard deviation above the mean are nearly 4 percent more likely to introduce and combine

methods in new ways, holding all other variables at their means. At two standard deviations above the mean, they are 7 percent more likely, and so on. This suggests that as funding sources become more institutionally diverse within labs, those labs experiment with new techniques, which introduces radical new hypotheses into the area.

Table 7 lists the covariate estimates for models predicting paper and patent citations to a lab's scientific corpus, indicating academic success and commercial relevance. As in Table 6, for each indicator, estimates for two models are listed with and without the lab characteristics of funding diversity and size. Significant increases in the Log-likelihood of models that take lab size and funding diversity into account suggest their collective importance. These results suggest, as expected, that the determinants of success differ sharply from those for originality. They further demonstrate that similar factors drive both academic success and commercial relevance.

Interestingly, research activity housed in hospitals, nonprofit research institutes, and companies garner the most citations. This may be partly because these institutions have fewer, less consistent requirements to publish. Hence, they will more likely put out only the best findings, or mete them out in more substantial papers, garnering more citations per paper. Interestingly, scientists who move between companies and other types of institutions produce papers less frequently cited by patents when they work at companies ( $-0.723$ ,  $p < 0.10$ ). This suggests some generality to Somerville's attitude that Mendel scientists could only publish work without commercial potential.

Subfield diversity has a positive effect on paper citations ( $0.315$ ,  $p < 0.10$ ). As diversity in the subjects studied by neighboring, similar scientists increases, labs are able to do work that is more relevant to those fields. Diversity of departments using *Arabidopsis* within a lab's institution has a negative effect, insignificant for paper citations, but mildly significant for citations from patents ( $-0.708$ ,  $p < 0.10$ ). Organization-level funding diversity also exerts a negative influence on commercial relevance ( $-0.186$ ,  $p < 0.10$ ). Together, this paints a portrait of universities and noncommercial research institutes as producing the most commercially relevant published science. It also suggests a possible censorship effect: research activities in companies or institutions receiving publicized industry support may leave unpublished their most commercially relevant research.

The tenure and tenure-squared effects indicate that scientists' earlier work gets cited more than their later work, although the effect tapers over time. This is partly a function of right censoring – earlier work has more time to be cited. Research activity in commercially relevant subfields studying things

**Table 7.** Fixed-Effects Models of Funding Diversity on Citations.

	Paper Citations (Academic Success)		Patent Citations (Commercial Relevance)	
Region (U.S. baseline)				
Europe	0.144 (0.061)*	-0.157 (0.074)*	-0.446 (0.180)*	-0.458 (0.224)*
Asia	0.163 (0.069)*	-0.073 (0.085)	-0.174 (0.227)	-0.276 (0.286)
Britain and Commonwealths (e.g., Australia, Canada)	0.130 (0.066)*	-0.020 (0.083)	-0.334 (0.206)	-0.202 (0.259)
Institutional context (university w/o agriculture baseline)				
Hospital	0.250 (0.252)	0.501 (0.237)*	-1.026 (0.657)	0.207 (0.820)
Research institute	0.100 (0.068)	0.156 (0.084)†	-0.099 (0.199)	-0.149 (0.259)
Company	0.253 (0.102)*	0.309 (0.141)*	-0.222 (0.230)	-0.723 (0.409)†
Government	0.025 (0.083)	0.147 (0.111)	-0.231 (0.269)	-0.123 (0.405)
University with agriculture	-0.095 (0.052)†	-0.001 (0.062)	-0.281 (0.148)†	-0.087 (0.187)
Subfield diversity	0.332 (0.153)*	0.315 (0.189)†	0.207 (0.440)	0.164 (0.593)
Department diversity	0.085 (0.095)	-0.147 (0.114)	-0.398 (0.279)	-0.708 (0.363)†
Industry support	0.016 (0.034)	0.015 (0.035)	-0.293 (0.103)**	-0.186 (0.111)†
Lab context				
Tenure	-0.114 (0.008)**	-0.159 (0.009)**	-0.264 (0.023)**	-0.316 (0.029)**
Tenure <sup>2</sup>	0.004 (0.001)**	0.005 (0.001)**	0.008 (0.002)**	0.010 (0.002)**
Commercial subfield	0.147 (0.030)**	0.150 (0.035)**	0.466 (0.084)**	0.269 (0.107)*
Number funding rels.		0.040 (0.062)		0.106 (0.182)
Ln number in lab		0.095 (0.025)**		0.284 (0.077)**
Constant	0.108 (0.133)	0.651 (0.169)**	-0.129 (0.383)	-0.115 (0.518)
<i>N</i>	3,534	3,534	1,956	1,956
PIs	1,015	1,015	424	424
Log-likelihood	-14,957.834	-10,050.925	-2,064.543	-1,445.392

Note: Standard errors in parentheses.

\*\*Significant at 1%.

\*Significant at 5%.

†Significant at 10%.

like receptors, phosphorylation cascades, and nutrient uptake attracts attention in academic science as well as in the realm of industrial development.

In stark contrast with its effect on originality, lab size exerts a modest positive influence on both academic success and commercial relevance (0.095 and 0.284,  $p < 0.01$ ). More publishing scientists in a lab's stable can muscle out more research, and increase the overall citation-yield, even if there are lower average paper citations. The coefficient for funding diversity, which played a strong role in generating research novelty, is weakly positive in both models, but does not differ statistically from zero in either.

## DISCUSSION

Model results provide selective support for the proposition that diverse sponsorship fosters innovative practice in academic labs. Funding diversity has a positive relationship with originality in combining scientific methods and generating unexpected hypotheses, although it does not necessarily affect more moderate forms of experimental novelty. New insights introduced by diverse patrons may stimulate labs to experiment with new procedures. Alternately, labs may parry funds from diverse patrons toward the purchase of new equipment, requiring them to puzzle over unanticipated findings. Insofar as diverse insights should affect all forms of novelty, while parried funding might be expected to impact only expensive methodological innovation, my findings provide more support for the latter mechanism. Methodological innovation may be a multi-vocal strategy that enables scientists to address the demands of commercial patrons for new tools and problem-solving insight, while simultaneously generating fresh findings to address concerns of mature science. In short, diverse funding more likely facilitates than inspires academic entrepreneurs to creativity.

My models compare prior funding and present research activity in order to isolate a causal relationship between the two. The results, however, could reflect other unmeasured factors. Creative scientists – like Somerville – may seek diverse funding to do creative science. Or even more likely, diverse funding could co-evolve with creativity. Diverse patrons may enable and ultimately reinforce a disposition of creativity. More extensive data on scientists' historical creativity – possibly including educational background or research beyond the scope of their *Arabidopsis* oeuvre – can be examined to disentangle this relationship.

Lab size, in contrast, significantly discouraged original activity in the realm of genes and GO terms. The coefficient for lab size was also negative,

though not significant, in the methodological novelty model. When PIs direct smaller labs, their scientific activity is more original; when they direct larger ones, it is less. Smaller labs may simply be more agile. Scientists who run them may more easily organize and reorganize their research activities in response to their own ideas – ideas independent of the prevailing scientific frontier. Larger labs, while doing some original work, hang more heavily on established patterns of experiment in the field, if only because so many of their researchers come from other established labs. Even if these large labs begin with original practices, they more quickly reproduce them in quantity. Like factories, they are more effective at scaling up innovations than consistently creating them.

My findings also confirm sharp differences between the drivers of originality and those of success. While lab size is a liability in the generation of novel practices, it is a boon in pumping out papers and garnering citations – the coin of the realm in academic science. Lab size also fuels patent citations to labs' papers, an indicator of their commercial relevance. Funding diversity posts a positive, but statistically insignificant effect in these equations, suggesting its reduced importance as a direct determinant of academic or commercial success. In promoting original practices and unexpected findings, funding diversity may have a relationship with the variance and not the mean of success: labs with more diverse institutional funding may have higher success *and* failure rates, with a similar overall average of success (March, 1991). In this way, Stinchcombe's (1965) "liability of newness" may apply not just to young but to original labs whose practices often fail to produce digestible findings, but when they do succeed, they do so spectacularly. Small labs with diverse funding may be better, on average, at exploring new scientific possibilities. Large ones are clearly better at exploiting them.

Extending these insights to the realm of new business ventures outside laboratory science, my findings suggest that under certain conditions, innovative sources of supply provide fertile ground for the cultivation of innovative processes and products. A resource base representing diverse interests may facilitate originality if entrepreneurs are able to create processes and products that speak to them all. This allows them the freedom to explore with excess funding parried from projects that overlap. When entrepreneurs are unable to address the diverse interests of sponsors with a single effort, or when their sponsors, though diverse (e.g., venture capitalists, banks, large firms) organize amongst themselves, firms may experience greater constraint than if funded by a single source. Moreover, unlike academics, business owners are not shielded from the direct influence of their

sponsors. As a result, sponsors with controlling investments may constrain entrepreneurs from experimenting with the insights they provide. Future research in business contexts will be necessary to support these inferences.

This research also provides a new, direct example of the inertness of larger organizations. Larger enterprises simply produce more. If they began with a promising innovation, and the business environment remained stable, their prospects for success are bright. But if either of these conditions are not met – if their initial innovation was not promising or the environment shifts – their ability to innovate again reduces as their size increases. This research also highlights that the same qualities that determine success in academic science determine originality in business. The same things that drove publication citations also drove patent citations. Similarly, Somerville's familiarity with established ideas in research – the conservation of biological control among transcription factor genes – allowed him to innovate in business to found one of the earliest genomics companies. This demonstrates that entrepreneurship in one context can often be primed by established institutions in others. Any sociology of entrepreneurship that hopes to credibly recount the creation story for new innovations will need to examine not only the direction, but the timing of flows from one entrepreneurial context to another.

More broadly, my findings underline the methodological importance of comparing entrepreneurial contexts beyond the realm of traditional markets for goods and services. To gain insight into the relationship between innovation in financing and research activity, I examined academic science because labs disclose a published record of their activities, and a historical record of all prior activity exists alongside it in previous issues of the same journals. New companies frequently use secrecy to keep their work from being outpaced by bigger, more established firms (Leibeskind, 2000; Cohen, Nelson, & Walsh, 2000), which makes it hard for entrepreneurship research to penetrate the black box of production.<sup>23</sup> In his study of occupations, the sociologist E. C. Hughes noted the difficulty of penetrating the work of professionals:

Prestige is so much a matter of symbols, and even pretensions – however well merited – there goes with prestige a tendency to preserve a front which hides the inside of things; a front of names, of indiscretion, of secrecy (much of it necessary secrecy). On the other hand, in things of less prestige, the core may be more easy of access. (Hughes, 1971, p. 342)

Entrepreneurial businesses may put profits over pretension, but entrepreneurial contexts where less is at stake could provide sociologists with a

clearer window into those where more is. Of course, the comparisons must be focused: different contexts produce insights with only partial relevance to one another. But sociology's most enduring contribution to the interdisciplinary study of entrepreneurship may be its museum of social innovations that can stand alongside entrepreneurship in the market.

## NOTES

1. Worcester hastens to add in the margin that though Publius Decius was a "manly man," and, as such "ys more to be commended" than a merely "hardye man," his "willefull deth ... ys not aftyr christen lawes commended." In short, do not do this at home.

2. The book also appears to contain Vincente de Beauvais' *Speculum Historiale*.

3. Ironically, this description, at the beginning of a chapter entitled "Crumbling Walls" in *Capitalism, Socialism and Democracy*, prefaces Schumpeter's infamous prophesy about the demise of the entrepreneur and his "function" in the wake of big business.

4. Burt's (1982) first book is actually titled *Toward a structural theory of action: Network models of social structure, perception, and action*. Its theory is essentially the same as *Structural Holes*, except it did not couple the information and control benefits of brokering those holes.

5. My sister, a biochemistry professor in a small, Midwestern liberal arts college knows a biochemist at a California junior college that has to offer his lab up for cooking classes in the evening.

6. Microeconomics uses a production function to characterize the minimum inputs required to produce designated quantities of output in a firm or other production process. This assumes technical efficiency – it ignores the engineering and managerial problems associated with production within the firm – in order to highlight the economic problem of allocative efficiency.

7. The article does look at one noncommercial, distinctly academic "entrepreneurial" feature: success in capturing large government grants. This interest is not sustained in subsequent research.

8. Also relevant here is the work of economist Zvi Griliches and his students – Bronwyn Hall, James Adams, Ariel Pakes (as also Henderson, Jaffe, and Trajtenberg).

9. Papers from this conference will appear in a forthcoming issue of the *Journal of Economic Behavior and Organization*.

10. Bruno Latour (1987) explicitly tries to open the black box of academic labs, noting the diversity of their funding, activities, networks, and outputs. He does not, however, consider or systematically examine variation in these features.

11. March and Sutton (1997) make a similar point about the instability of organizational performance as a dependent variable in studies of organization.

12. I expect, although I do not test it in this paper, that these novel, scientific explorations will have a higher variance rather than a higher average rate of success

(March, 1991). They will succeed and fail more dramatically than those of conservatively organized labs.

13. Note that this claim is more specific than the popular but problematic notion that diversity facilitates innovative, productive action (Edelman, Fuller, & Mara-Drita, 2001; Reagans & Zuckerman, 2001).

14. Thompson Scientific generously gave me use of the entire Science Citation Index for the purpose of this project.

15. Codes for molecular function, biological process, and cellular component (13,976) used to code the *Arabidopsis* genes were developed in a joint effort among curators of several model organisms to characterize genes independently of their host organisms, entitled the Gene Ontology (GO) project. Development and anatomical annotations were developed by TAIR specifically for *Arabidopsis*.

16. A multi-word concept is established if a word frequently co-occurs with another word or string of words (e.g., Northern Blot; Polymerase Chain Reaction).

17. I used a BLAST (Basic Local Alignment Search Tool)-like algorithm that searches through windows of text-characters looking for substring matches.

18. We drew on the methodological classification scheme used in the BIOSIS database.

19. The Box and Cox (1964) transform,  $y^\lambda - 1/\lambda$ , selects  $\lambda$  such that the skewness of the new variable equals 0.

20. I tried an alternate process of “finding” subfields in the data by assigning a series of random variables to each publication which I conceptualized as “latent concepts,” and then simulating an influence process by which the latent concepts in each article influenced those they cited and those that cited them (Moody, 2001). This process created subfields that were more integrated by author and institution, but were much less distinguishable by their content.

21. Overdispersion refers to the case where the variance of a quantity divided by its mean is high.

22. One-tailed *t*-tests are used to evaluate the significance of hypothesized coefficients.

23. See Ruef (2002) for a notable exception.

## REFERENCES

- Almeida, P., Dokko, G., & Rosenkopf, L. (2003). Startup size and the mechanisms of external learning: Increasing opportunity and decreasing ability? *Research Policy*, 32, 301–315.
- Boley, D. (1997). *Principle Direction Divisive Partitioning*. <http://www.users.cs.umn.edu/~boley/Distribution/PDDP2.html>, accessed May 15, 2002.
- Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society*, 26, 211–243.
- Burt, R. (1982). *Toward a structural theory of action: Network models of social structure, perception, and action*. New York: Academic Press.
- Burt, R. (1992). *Structural holes: The social structure of competition*. Cambridge, MA: Harvard University Press.
- Burt, R. (2004). Structural holes and good ideas. *American Journal of Sociology*, 110(2), 350–399.



- Caxton, W. (1881[1485]). *The Lyf of the Noble and Crysten Prince, Charles the Grete*. (Translated by W. Caxton, with introduction, notes and glossary by S. J. H. Herrtage.) London: N. Trubner & Co.
- Chang, J., Schutz, H., & Altman, R. (2003). *GAPSCORE: Finding gene and protein names one word at a time*. Stanford Working Paper.
- Cohen, W. M., Nelson, R. R., & Walsh, J. P. (2000). *Protecting their intellectual assets: Appropriability conditions and why U.S. manufacturing firms patent (or not)*. NBER Working Paper 7552.
- Council of Scientific Affairs, American Medical Association (AMA). (2004). Influence of funding source on outcome, validity, and reliability of pharmaceutical research. Report presented as *CSA Report 10* at the 2004 AMA annual meeting.
- DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organization fields. *American Sociological Review*, 48, 147–160.
- Ding, W., & Stuart, T. (2006). When do scientists become entrepreneurs? The social structural antecedents of commercial activity in the academic life sciences. *American Journal of Sociology*.
- Edelman, L. B., Fuller, S. R., & Mara-Drita, I. (2001). Diversity rhetoric and the managerialization of law. *American Journal of Sociology*, 106, 1589–1641.
- Evans, J. (2006). *Industry collaboration and theory in academic science*. Working Paper.
- Fagerberg, J. (2004). Innovation: A guide to the literature. In: J. Fagerberg, D. C. Mowever & R. R. Nelson (Eds), *The Oxford handbook of innovation*. Oxford: Oxford University Press.
- Froude, J. A. (1884). *Thomas Carlyle: A history of his life in London, 1834–81*. New York: Scribner.
- Gøtzsche, P. C., Hröbjartsson, A., Johansen, H. K., Haahr, M. T., Altman, D. G., & Chan, A.-W. (2006). Constraints on publication rights in industry-initiated clinical trials. *JAMA*, 295, 1645–1646.
- Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology*, 78, 1360–1380.
- Hage, J., & Hollingsworth, J. R. (2000). A strategy for the analysis of idea innovation networks and institutions. *Organization Studies*, 21(5), 971–1004.
- Hannan, M., & Freeman, J. (1988). The ecology of organizational mortality: American labor unions, 1836–1985. *The American Journal of Sociology*, 94, 25–52.
- Hausman, J., Hall, B. H., & Grilliches, Z. (1984). Econometric models for count data with an application to the patents–R&D relationship. *Econometrica*, 52(4), 909–938.
- Henderson, R., Jaffe, A., & Trajtenberg, M. (1997). University versus corporate patents: A window on the basicness of invention. *Economics of Innovation and New Technology*, 5(1), 19–50.
- Hollingsworth, J. R., & Hollingsworth, E. J. (2000). Major discoveries and biomedical research organizations: Perspectives on interdisciplinarity, nurturing, leadership, and integrated structures and cultures. In: P. Weingart, & N. Stehr (Eds), *Practicing interdisciplinarity* (pp. 420–438). Toronto: University of Toronto Press.
- Hollingsworth, J. R., Hollingsworth, E. J., & Hage, J. (forthcoming). *Organizations and performance in bio-medical science*. Cambridge: Cambridge University Press.
- Hughes, E. C. *The Sociological Eye: Selected Papers*. Chicago: Aldine-Atherton.
- Kline, S., & Rosenberg, N. (1986). An overview of innovation. In: R. Landau & N. Rosenberg (Eds), *The positive sum strategy* (pp. 275–305). Washington, DC: National Academy Press.

- Kuhn, T. (1961). The function of measurement in modern physical science. *Isis*, 52, 161–190.
- Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Stratford, UK: Open University Press.
- Leibeskind, J. P. (2000). *Property rights, secrecy, and the boundaries of the firm: Theory and evidence*. Working Paper.
- Levinthal, D. A., & March, J. G. (1993). The myopia of learning. *Strategic Management Journal*, 14(Winter), 95–112.
- Lexchin, J. R. (2004). Implications of pharmaceutical industry funding on clinical research. *The Annals of Pharmacotherapy*, 39(1), 194–197.
- Louis, K. S., Blumenthal, D., Gluck, M. E., & Stoto, M. A. (1989). Entrepreneurs in academe: Exploration of behaviors among life scientists. *Administrative Science Quarterly*, 34(1), 110–131.
- March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, 2, 71–87.
- March, J. G., & Sutton, R. I. (1997). Organization performance as a dependent variable. *Organization Science*, 8(6), 698–706.
- Markiewicz, K. R., & DiMinin, A. (2004). *Commercializing the laboratory: The relationship between faculty patenting and publishing*. Working Paper.
- Merton, R. K. (1942). Science and technology in a democratic order. *Journal of Legal and Political Sociology*, 1, 115–126.
- Montgomery, J. H., Byerly, M., Carmody, T., Li, B., Miller, D. R., Varghese, F., & Holland, R. (2004). An analysis of the effect of funding source in randomized clinical trials of second generation antipsychotics for the treatment of schizophrenia. *Controlled Clinical Trials*, 25(6), 598–612.
- Moody, J. (2001). Peer influence groups: Identifying dense clusters in large networks. *Social Networks*, 23, 261–283.
- Mowery, D., & Ziedonis, A. (2002). Academic patent quality and quantity before and after the Bayh–Dole act in the United States. *Research Policy*, 31(3), 399–418.
- Nelson, R. R. (1981). Assessing private enterprise: An exegesis of tangled doctrine. *The Bell Journal of Economics*, 12, 93–111.
- Nelson, R. R. (1986). R&D, innovation and public policy: Institutions support technical advance in industry. *American Economic Review*, 76, 186–189.
- Owen-Smith, J. (2003). From separate systems to a hybrid order: Accumulative advantage across public and private science at research one universities. *Research Policy*, 32(6), 1081–1104.
- Owen-Smith, J., & Powell, W. W. (2001). To patent or not: Faculty decisions and institutional success at technology transfer. *Journal of Technology Transfer*, 26(1), 99–114.
- Oxford English Dictionary (OED). (2006). In: J. A. Simpson & E. S. C. Weiner (Eds) (2nd ed.). Oxford: Clarendon Press (1989). OED Online Oxford University Press, 10 April 2006. <http://dictionary.oed.com/cgi/entry/50076367>.
- Padgett, J., & Ansell, C. (1992). Robust action and the rise of the medicine. *American Journal of Sociology*, 98, 1259–1330.
- Parsons, T. (1968). On the concept of value commitments. *Sociological Inquiry*, 38, 135–160.
- Powell, W. W., & Owen-Smith, J. (1998). Universities and the market for intellectual property in the life sciences. *Journal of Policy Analysis and Management*, 17(2), 253–277.
- Reagans, R., & Zuckerman, E. (2001). Networks, diversity and productivity: The social capital of corporate R&D teams. *Organization Science*, 12(4), 502–517.

- Redei, G. P. (1975). *Arabidopsis* as a genetic tool. *Annual Review of Genetics*, 9, 111–127.
- Rosenberg, N., & Nelson, R. R. (1994). American universities and technical advance in industry. *Research Policy*, 23, 323–348.
- Ruef, M. (2002). Strong ties, weak ties and islands: Structural and cultural predictors of organizational innovation. *Industrial and Corporate Change*, 11(3), 427–449.
- Schumpeter, J. (1934[1911]). *Theory of economic development*. Cambridge, MA: Harvard University Press.
- Schumpeter, J. (1943). *Capitalism, socialism and democracy*. New York: McGraw-Hill.
- Shane, S. (2003). *A general theory of entrepreneurship: The individual-opportunity nexus*. Cheltenham, UK: Edward Elgar.
- Somerville, C. (1978). *E. coli mutants defective in the conversion of pppGpp to ppGpp*. Unpublished doctoral dissertation. University of Alberta, Alberta.
- Stinchcombe, A. L. (1965). Social structure and organizations. In: J. G. March (Ed.), *Handbook of organizations*. Chicago: Rand McNally & Co.
- Stokes, D. (1997). *Pasteur's quadrant: Basic science and technological innovation*. Washington, DC: Brookings Institution Press.
- Thornton, P. (1999). The sociology of entrepreneurship. *Annual Review of Sociology*, 25, 19–46.
- Thursby, J., & Thursby, M. (2003). University licensing and the Bayh–Dole act. *Science*, 301(22), 1052.
- Walbot, V. (2000). A green chapter in the book of life. *Nature*, 408, 794–795.
- Weber, M. (1968[1922]). *Economy and society*. New York: Bedminster Press Inc.
- Worcester, W. (1860[1475]). *The Boke of Noblesse: Addressed to King Edward the Fourth on his invasion of France in 1475*. (With an introduction by John Gough Nichols, F. S. A. Printed for the Roxburghe club.) London: J. B. Nichols and Sons.
- Zucker, L., & Darby, M. R. (1996). Star scientists and institutional transformation: Patterns of invention and innovation in the formation of the biotechnology industry. *Proceedings of the National Academy of Sciences*, 93(23), 709–716 (With Michael R. Darby).

# TURNING IDENTITY INTO FORM: THE CAUSE AND CONSEQUENCE FOR KAISER PERMANENTE OF BECOMING AN HMO

Carol A. Caronna

## ABSTRACT

*How do organizations act as entrepreneurs and what are the outcomes of their innovations? This paper intersects two broad areas of organizational research: the sociology of entrepreneurship and the study of organizational forms. A case study of Kaiser Permanente's role as an institutional entrepreneur in the creation of the health maintenance organization form illuminates the benefits and pitfalls of institutional entrepreneurship – in this case, the act of turning identity into form. Examining organizations as institutional entrepreneurs also raises questions and challenges for future research about both entrepreneurs and models of organizing.*

In the typical story of entrepreneurship, an individual, or perhaps a small team of persons, founds a new organization “to accomplish things they cannot do on their own” (Aldrich, 1999, p. 75). Drawing from social networks, external investments, and societal status, as well as personal knowledge, capital, experience, skill, and hope, founders intentionally create an entity that is both measurable and real. As the sociology of entrepreneurship

---

**The Sociology of Entrepreneurship**

**Research in the Sociology of Organizations, Volume 25, 309–336**

**Copyright © 2007 by Elsevier Ltd.**

**All rights of reproduction in any form reserved**

**ISSN: 0733-558X/doi:10.1016/S0733-558X(06)25009-2**

points out, the process of formation may be complex and messy, the environment may be harsh and unforgiving, and most likely the new venture will fail, but the outcome of entrepreneurial activity always is marked by the existence of some entity, no matter how fleeting.

In the context of such a story, what does it mean to describe the founding of organizational forms as the outcome of entrepreneurship? An organizational form, which exists as a cognitive “building block” or “blueprint” for new or extant organizations to follow (Scott, 2001), is an intangible entity with symbolic value as a cultural object (Carroll & Hannan, 2000). Institutional entrepreneurs (DiMaggio, 1988) influence form creation through “efforts to identify political opportunities, frame issues and problems, and mobilize constituencies ... to infuse new beliefs, norms, and values into social structures” (Rao, Morrill, & Zald, 2000, pp. 238–239). Although general processes of entrepreneurship may be similar, the differences between actors and new ventures at varying levels of analysis raise a number of challenges for studies of institutional entrepreneurs.

Researchers generally can measure when an organization is “born” and “dies.” But when is a form created, when is it measurable, and when is it real? In addition, who or what constitute its founders? Conceivably the founder(s) of a form could range from the micro to the macro – individuals, organizations, professional associations, or the state (e.g. Hoffman, 1999; Greenwood, Suddaby, & Hinings, 2002; Maguire, Hardy, & Lawrence, 2004). Multiple founders could be working toward the same goal of form creation, but not necessarily in concert; “teams” of founders may be more virtual and retrospective than real. The contribution of founders may be symbolic as well as economic and social; a particular organization’s presence in the institutional environment itself may be a resource, without any direct or intentional activity on the organization’s part. Founders may drop in and out of the creation process as their resources allow and their efforts are or are not influential, especially if the process is lengthy. In short, can theories of entrepreneurship account for the complexities of form creation, and what is the value for organizational theory of framing form creation as entrepreneurial activity?

In this paper, I address these questions by examining the role of Kaiser Permanente (KP), one of the nation’s largest and oldest prepaid group practices, in the establishment of the health maintenance organization (HMO) form. First I provide a brief theoretical framework to clarify key terms. Then I present an analysis of KP’s entrepreneurial activity and relationship with the organizational field divided into three time periods, guided by previous work on institutional eras in the U.S. health care field

(Scott, Ruef, Mendel, & Caronna, 2000), and focusing on the dynamic interrelationship between identity and institutions in the process of entrepreneurship (see Lounsbury & Glynn, 2001). Based on this case study, I close the paper with a discussion of when and how organizations can affect change in the institutional environment, under what conditions organizational identities can influence organizational forms, and the consequences – both intended and unintended – for institutional entrepreneurs of creating new forms while trying to sustain old identities.

## IDENTITY, FORM, AND FIELD

### *Organizational Identity*

Studies of entrepreneurship pay a great deal of attention to the identity of individual founders. Characteristics of interest range from the ascribed, such as race/ethnicity and gender, to the achieved, such as immigrant status, professional background, and job experience (Aldrich, 1999; Ruef, Aldrich, & Carter, 2003). Dobrev and Barnett (2005) propose that “the founder’s identity is tightly linked to that of the organization and to its innovative endeavors” (Dobrev & Barnett, 2005, p. 435). Unlike static views of identity that assume individuals maintain their personal characteristics over time, their conception of identity relates to the changing role, from charismatic to bureaucratic, of a founder vis-à-vis his or her organization, the dynamic influence of external evaluations on a founder’s self-conception, and the irony that a founder’s “(search) for an identity ... tends to be denied by the very organization that he builds” (Dobrev & Barnett, 2005, p. 446).

In parallel, conceptions of organizational identity range from characteristics to relationships. Consistent with the interpretive approach to organizational studies (Aldrich, 1999), Albert and Whetten (1985) define organizational identity as the features of an organization that its members perceive to be central, enduring, and distinctive. It is considered “an important and collectively held frame invoked by members to both interpret and take action; that is, to make sense of their world” (Golden-Biddle & Rao, 1997, p. 594; see also Dutton & Dukerich, 1991; Gioia & Thomas, 1996; Whetten & Godfrey, 1998).

By contrast, organizational ecologists argue organizational identity should be based on the constraints on an organization rather than the constancy of a set of features: “an identity constrains what an organization would/could do and what is expected and not expected of it” (Carroll & Hannan, 2000,

p. 68; see also Baron, 2004). An organization's identity is deeply embedded in its social context, such that "social codes (comprised of sets of social rules and signals) ... specify the features that an organization can legitimately possess" (Carroll & Hannan, 2000, p. 68). In the long view, as insiders and outsiders enforce dynamic social codes on an organization, the features that an organization can legitimately possess can change dramatically (see Scott et al., 2000). Thus, organizations can and generally have many identities, including nested identities identifying sharper and broader interpretations of its social constraints. In addition, the focus, resonance, and authenticity of an organization's identity can change over time (Baron, 2004).

For this analysis, I attempt to strike a pragmatic compromise between the two views of organizational identity. I rely on KP's own presentation of its central, enduring, and distinctive features to create a generally stable sense of identity over time, while exploring how internal and external evaluators interpreted these features differently in different time periods. The changing meaning of KP's identity and its relation to its environment led to varying assessments of KP's legitimacy and illegitimacy, resulting in diverse sanctions aimed at virtually the same characteristics (Caronna, 2000). As the social code around it changed, KP's central and enduring features became more and less distinctive, more and less legitimate, and more and less influential in shaping the organization's entrepreneurial activity.

### *Organizational Form*

Organizational forms have been described as genetic structures, blueprints (Hannan & Freeman, 1989), patterns, archetypes, and templates (Greenwood & Hinings, 1993; Scott et al., 2000) that "convey an image of organizations similar enough to be called a population but different enough to display unmistakable variety" (Aldrich, 1999, p. 40). Once conceived of in straightforward terms, recent work has argued that forms "are historically specific and change over time" (Carroll & Hannan, 2000, p. 79). The form is best thought of as an emerging and evolving process, not a static entity (Carroll & Hannan, 2000; Ruef, 2000). Carroll and Hannan (2000) relate organizational forms to their conception of identity: "whereas identities might be established primarily or even exclusively by evaluations of insiders, forms are cultural objects" (Carroll & Hannan, 2000, p. 73). Like identity, social codes (both signals and rules of conduct) and their constraints on features are central in defining forms. But as a cultural object, an organizational form also is "a recognizable pattern that takes on rule-like standing" (Carroll & Hannan,

2000, p. 67). Once organizational members, external constituencies, and society at large can enforce this pattern, the form itself becomes a social code by which organizations' identities are formed and judged.

### *Organizational Field*

The contexts in which organizations and forms emerge greatly influence their development and, ultimately, success or failure. Ruef (2000) argues that examining the institutional environment (Meyer & Rowan, 1977; DiMaggio & Powell, 1983; Scott, 2001) is key to understanding the development of organizational forms because of the influence on entrepreneurial activity of populations of existing forms, legitimacy claims and assessments, and social demands. The institutional environment defines what is appropriate and expected through its "regulative, normative, and cultural-cognitive structures that operate to provide coherence, meaning, and stability to a field" (Scott et al., 2000, p. 20). Institutional environments encompass organization fields: "those organizations that, in the aggregate, constitute a recognized area of institutional life: key suppliers, resource and product consumers, regulatory agencies, and other organizations that produce similar services or products" (DiMaggio & Powell, 1983, p. 148).

Organizational fields generally are subject to and shaped by three types of elements: regulative, normative, and cultural-cognitive (Scott, 2001). Regulatory elements include rules backed by sanctions and enforcement agencies that govern commerce and exchange relations (Scott et al., 2000). Normative elements refer to informal and diffuse rule systems that structure expectations and systems of mutual obligations. The cultural-cognitive aspects of the field encompass the beliefs, orienting frames, logics, and scripts that influence actors' constructions of meaning. They also include models of organizing, or sets of expectations for how organizations that provide certain functions or services should be structured. Cognitive categories of social actors and ways of acting constitute much of the order of social life (Scott, 2001).

One important way the institutional environment affects organizations is through expectations of isomorphism: pressures to conform to socially defined cultures in order to be considered legitimate (Meyer & Rowan, 1977; Greenwood & Hinings, 1996). Suchman defines legitimacy as "a generalized perception or assumption that the actions of an entity are desirable, proper, or appropriate within some socially constructed system of norms, values, beliefs, and definitions" (Suchman, 1995, p. 574). Internal and external



audiences confer legitimacy if an organization conforms to a standard or model (Ruef & Scott, 1998), such as an organizational form. The benefits of legitimacy include stability of resources (Suchman, 1995), stronger identification of members (Scott & Lane, 2000), and an increased likelihood of survival (Ruef & Scott, 1998).

As the dynamic and contextual definitions of identity, form, and field imply, entrepreneurial activity aimed at creating a new organizational form necessarily becomes an interplay between organizations and environments. Just as a form is social code and is subject to social code (Carroll & Hannan, 2000), the creation of a new model of organizing reflects back on the entrepreneurial organization and its identity. A new form establishes new criteria to measure and assess the legitimacy of an organization, which can affect in both intentional and unintentional ways the perception of the legitimacy of the founding organization. The case of KP and the HMO form reveals the consequences – both positive and negative – of these dynamics.

## **KAISER PERMANENTE AS INSTITUTIONAL ENTREPRENEUR**

### *The Era of Professional Dominance, 1930–1965*

The organization known today as KP<sup>1</sup> was formed by Sidney Garfield M.D. in the 1930s to provide on-site medical services for Kaiser Industries workers at geographically isolated construction sites (Hendricks, 1991; Smillie, 1991). For pragmatic reasons, the services were financed by voluntary payroll deductions of five cents per week per worker, with no additional charges for services (Cutting, 1971; Heiner, 1989). As the only medical practitioners within miles, the physicians and other staff members worked together as a team. The success of these programs, including the care of nearly one hundred thousand Kaiser Shipyards workers during World War II, led Garfield and a few other doctors to open their health plan to the public in 1945 in California, Oregon, and Washington state. At that time, KP was one of a small number of health care organizations labeled “prepaid group practices” (Starr, 1982). These organizations, considered today to be prototypical HMOs, included the Ross-Loos Clinic of Los Angeles, founded in 1929, the Group Health Association of Washington, DC, founded in 1937, the Group Health Cooperative of Puget Sound, founded in 1947, and the Health Insurance Plan of Greater New York, founded in 1947 (Luft, 1987).

KP's organizational identity, in terms of its central, enduring, and distinctive characteristics (Albert & Whetten, 1985), is captured by a generally stable set of core principles. Repeatedly published, discussed, and analyzed by KP leaders over the decades, any individual acting on behalf of KP arguably would know "who" the organization was and what it stood for. These principles,<sup>2</sup> which have been called the "Kaiser formula" (Somers, 1971a) and the "genetic code" (Cutting, 1986; Smillie, 1991) are: (1) prepayment by members; (2) physician group practice; (3) the integration of KP-owned medical facilities, including hospitals, medical offices, laboratories, and pharmacies; (4) comprehensive and preventive medical care; (5) voluntary enrollment in the medical plan with "dual choice" – enrollees must have a choice of insurance providers; (6) physician responsibility for medical decisions; and (7) the nonprofit status of the health plan and hospitals.<sup>3</sup>

This neutral list of principles does little to capture the controversial nature of KP's identity during the era of professional dominance. Conceived of in terms of its social context and constraints imposed upon it (Carroll & Hannan, 2000; Baron, 2004), KP's identity was wholly illegitimate. Between World War I and 1965, professional providers, supported by state authority, had emerged as the strongest and most effectively organized profession in U.S. history and the main agents of governance in the U.S. health care field (Freidson, 1970; Starr, 1982; Abbott, 1988; Scott et al., 2000). With membership peaking at over 70 percent of active physicians in the U.S. during the 1940s and 1950s (American Medical Association [AMA], 1997), the AMA was the professional body with the most authority in the health care field. It served as an advocacy and lobbying body overseeing legislative activity, controlled access to hospital privileges, patient referrals, and malpractice insurance, and enforced norms against advertising, fee-splitting, and corporate medicine. The majority of physicians were solo practitioners, and the AMA fought directly and powerfully against detractors and deviants as it actively constructed the health care agenda (Starr, 1982).

During this era, the relationship between KP and organized medicine ran the gamut from "active hostility to armed neutrality" (Foster, 1989, p. 223). Permanente physicians routinely were denied membership in medical societies, privileges at local hospitals, and board certification for specialties (Foster, 1989; Smillie, 1991; Hendricks, 1991). In 1953, a California Medical Association editorial claimed that "'Permanente saved money at the expense of proper patient care,' and not only 'destroyed' the doctor-patient relationship but also 'robbed' the patient of his 'freedom' as a 'captive of

the plan” (Hendricks, 1991, p. 461). At the AMA’s annual convention in 1954, doctors from New York introduced a resolution designed to “condemn the restricting of a patient’s choice of doctor to the members of a group or panel as a violation of ‘the right of free choice’” (*Time Magazine*, 1954, p. 37).

The survival of KP and its ability to sustain a deviant identity within this institutional environment stemmed from several resources, most importantly Henry J. Kaiser. Born in 1882, Kaiser’s career as an industrialist began in 1909 with the founding of a paving, sand, and gravel company. A millionaire by age 48, he was the national president of the Associated General Contractors in the 1930s. During World War II, Kaiser gained national fame from the relentless production of his shipyards. In 1942, he was invited to testify at a U.S. Senate Committee hearing in Washington, DC, where he called for more doctors and more medical care. In the 1940s, he set up a public relations group to handle his growing fame and established a Washington office (Foster, 1989). He became friends with President Franklin D. Roosevelt and was reputedly the only industrialist who could get along with Secretary of the Interior Harold Ickes (Heiner, 1989).

In 1944, Kaiser was promoted as a presidential candidate and seriously considered as Roosevelt’s running mate (Foster, 1989). By 1945, his “fame rivaled that of war heroes like MacArthur and Eisenhower. He became ... a renowned public hero” (Foster, 1989, p. 112), and a Roper poll that year showed that “the public believed Kaiser had done more to help the president win the war than any other civilian” (Adams, 1997, p. 9). When Garfield in 1942 sought the very first facilities loan for KP, Henry Kaiser accompanied him to Bank of America headquarters in San Francisco where they appealed directly to founder A. P. Giannini (Heiner, 1989). Both were surprised when Giannini flatly turned them down, arguing that a hospital was a poor investment, but Giannini relented as a personal favor as long as Kaiser would guarantee the loan, which he immediately did. The fact that Kaiser helped the health plan fund and build its own facilities solved one of KP’s problems stemming from its illegitimacy; the fact that medical societies denied hospital privileges to Permanente physicians became irrelevant to the health plan.

One of the enduring stories about KP’s founding involves a meeting in 1938 between founder Sidney Garfield, M.D. and Henry Kaiser. During a visit to the Grand Coulee Dam site, Kaiser singled out Garfield for attention. He wanted to learn about the medical plan and toured its hospital and facilities. Impressed with what he saw, Kaiser reputedly said to Garfield, ““Young man, if your idea is half as good as you say it is, it’s not only good

for this project, it's good for the entire country'" (Heiner, 1989, p. 75). In an interview many years later, Garfield recalled that

[Kaiser said] 'if you're any good, you're going to have a great deal of competition, and that'll help you do your job. You're going to be copied. Therefore, your particular job must be to make sure your model is the very best in the whole country and remains worthy of being copied' ... He never lost that vision. For the nearly thirty years after that meeting, until he died, Henry Kaiser had a missionary zeal in his efforts to strengthen and broaden the health program. (Heiner, 1989, p. 76)

Henry Kaiser thus believed KP's identity should remain intact despite pressure from the AMA. The physicians in the Permanente Medical Group agreed with Henry Kaiser: they believed "prepaid group practice was the best health care financing and delivery system in sight" (Fleming, 1997, p. 71), in part because their plan "provide(d) so much care to the people at a cost they [could] so easily pay" (Foster, 1989, p. 216). Several early leaders recalled that Permanente physicians "were all very strong people and dedicated to this method of practice" (Link, 1986, p. 31) and that "idealistically, [new physician recruits] were told they would be pioneering a new method of care ... the key ... men and women we got came to us, not because of salary ... but because they believed in [us]" (Kay, 1987, pp. viii, 48).

In the 1950s and 1960s, KP's struggles to legitimate its identity were aided by several actors in its institutional environment. In the 1950s, positive coverage of prepaid group practice appeared in national magazines (*Time Magazine*, 1953; Velie, 1953). Attention from the government came as the numbers of prepaid group practices increased. In 1959, the Federal Employees Health Benefits Act mandated that federal employees have the option of a prepaid group practice if one was available. KP economist Avram Yedidia reflected that the act was "the first time Congress had acted to legitimize the existence of health service programs such as Kaiser's and other prepaid group practice medical programs" (Yedidia, 1987, p. 46). Most importantly, the AMA's independently commissioned Larson report, finished in 1959, supported the freedom to choose a system of medical care as an acceptable variant of the tradition of freedom of choice of physician (Starr, 1982; Smillie, 1991), reported that there was no evidence of lay interference in medical decisions in prepaid group practices (Starr, 1982), and stated that "the care delivered by group practice prepayment plans was on a level with that provided in the community" (KFMCP, 1973, p. 10). Combined with "a long string of legal defeats," "the Larson report ended official sponsorship of reprisals against prepaid group practice" (Starr, 1982, p. 327).

*The Era of Federal Involvement, 1966–1982*

In 1965, watershed legislation passed as part of President Lyndon B. Johnson's Great Society program instituted Medicare and Medicaid programs, which provided federal funding to reimburse medical care provided for the nation's elderly and poor (Marmor, 1970; Starr, 1982). Medicare-type programs proposed in the late 1950s and early 1960s had been defeated by organized medicine and Republican majorities in Congress (Starr, 1982). The Democratic sweep in 1964's presidential and Congressional elections ushered in a legislature and executive office more sympathetic to health care reform, but lawmakers were careful to appease the AMA in order to improve the new program's chances of success (Starr, 1982). Instead of establishing direct relations between the government bureaucracy and providers, the federal government reimbursed treatments for Medicare/Medicaid patients through intermediary insurance carriers (Starr, 1982). Even the strongest opponents of Medicare quickly discovered its fee-for-service reimbursement structure was a "bonanza" (Starr, 1982, p. 370). National health expenditures ballooned, "from 4.4 percent of the federal budget in 1965 to 11.3 percent in 1972" (Starr, 1982, p. 399).

Without intending to usurp the authority of the AMA and organized medicine, new institutional logics symbolized by Medicare opened the door to increasing government involvement in the health care field. Medicare/Medicaid legislation had strongly emphasized the importance of equity (Rao et al., 2000); American policy makers argued that every American had a right to receive health care when it was required (see Kennedy, 1972). This belief in equity was part of and reinforced the perspective that the government was the only social institution that could ensure the equitable distribution of health care. In the early 1970s, the federal government became more involved in licensing, health planning, rate setting, and market building (Scott et al., 2000). At the same time, both the AMA's membership and governance authority waned (Starr, 1982; Campion, 1984; Krause, 1997), decreasing centralized professional opposition to the government. As the government, driven by a "crisis" mentality (Starr, 1982), struggled to contain rising health care costs and regulate the field, it searched for new models of organizing for the health care field. With "doctors, hospitals, and insurance companies ... now completely on the defensive, trying to hold back a tide of disaffection" (Starr, 1982, p. 383), "the socialized medicine of one era [was to become] become the corporate reform of the next" (Starr, 1982, p. 396).

As medical costs escalated following the passage of Medicare/Medicaid, KP and other similarly cost-effective programs began to receive widespread

public attention. KP in particular was the focus of government visits and glowing press coverage, which enabled and reinforced KP's role as a leader of prepaid group practices. In 1967, a report of the National Advisory Committee on Health Manpower recommended the use of prepayment mechanisms, such as those utilized by KP, over fee-for-service medicine to accomplish the "mass delivery of medical care as a human right" (Williams, 1971, p. ii). The commission reported that

the quality of care provided by Kaiser is equivalent, if not superior to, that available in most communities. Permanent physicians use standard medical practices and procedures. Patient satisfactions is indicated by the overall flow of patients into Kaiser from competing health plans under the dual choice available to all Kaiser subscribers... The staff study group concluded that the majority of savings achieved by Kaiser result primarily from effective control over the nature of medical care that is provided and over the place where care is given. (Smillie, 1991, p. 219)

Several of the Commission's specific recommendations reinforced aspects of KP's structure and strategy: develop outpatient services, allow prepaid programs to share in savings from effective utilization of Medicare/Medicaid resources, reduce hospital utilization, encourage health insurance organizations to share savings with doctors, and develop peer review (Smillie, 1991). Around the same time, Cecil Cutting was invited to discuss KP in front of the American Association of Medical Clinics (which became the American Group Practice Association in 1974), the AMA Congress on the Socio-Economics of Medical Care, and the White House Conference on Medical Care Costs (Smillie, 1991). Perhaps most important symbolically, in 1967 the president of the AMA, Charles M. Hudson, visited The Permanente Medical Group. KP's organizational identity, considered illegitimate for years, was now resonant with the changed values, beliefs, and goals of American health care.

To deal with its new attention and relations with the health care field, in 1967 KP created the KP Committee to respond to requests to merge with other HMOs and provide information to others forming prepaid plans and group practices (Cutting, 1986; Smillie, 1991). In 1969, KP joined the Group Health Association of America, which a number of prepaid group practices founded in 1958. KP leaders were "convinced that it was...in the best interests of the program to have [a] relationship, [and] influence over, the national association which purported to speak for group practice prepayment plans" (Smillie, 1991, p. 234) and quickly assumed a leadership role. Dr. Keene recalled:

up until the early 1960s, none of us went anywhere, or talked to anybody about anything. Mr. Kaiser Sr. testified before a few committees of Congress about the

organization of medical care, and I went with him on a couple of occasions, and Sid [Garfield] had gone with him. But we weren't trying to evangelize or get involved in politics or crusades of any kind. (Keene, 1986, p. 132)

By 1969, though, KP leaders were “united in their belief in prepaid group practice medical care; [and] equally united in their desire to see it spread” (KFMCP, 1969, p. 10). The organization, with a broad and deep collection of resources at its disposal, was poised to begin its new phase as an institutional entrepreneur.

Just as social and economic contexts, structures of existing organizations, and dominant institutional and market forces matter for understanding the behavior and success of individual entrepreneurs (Dobrev & Barnett, 2005), KP's entrepreneurial actions and motivations were embedded in the politics, beliefs, values, and institutional logics characteristic of the early 1970s. Minnesota physician Paul M. Ellwood, Jr. coined the term HMO in 1970 to broadly label prepaid group practices and other comprehensive health care plans “whose common feature would be contractual acceptance of responsibility to provide a comprehensive range of services to members in exchange for prepaid premiums” (Brown, 1983, p. 207; see also Rao et al., 2000). At that time, there were only about 3 million HMO enrollees in the U.S. (Miller & Luft, 1994) and about 37 HMOs in 14 states (Gruber, Shadle, & Polich, 1988). KP was considered at the forefront of this new model of organizing health care (Williams, 1971; Somers, 1971a, 1971b), and Ellwood stated that, to ground his theories of HMOs in reality, he “had to go to Kaiser Permanente to find out how to do things” (KPMCP, 1987, p. 17). Indeed,

The record of the Kaiser Health Foundation suggested it was possible to provide high quality prepaid health care at 20 to 40 percent lower cost than fee-for-service medicine. Advocates of the ‘health team’ approach hoped that nurse practitioners, physicians’ assistants, and other ‘physician extenders’ could improve access and efficiency. High rates of surgery and hospitalization suggested that more careful peer review might significantly reduce expenses by discouraging unnecessary procedures. Extensive duplication of facilities and equipment suggested that effective health planning could yield notable benefits and savings. (Starr, 1982, p. 383)

In 1971, President Richard Nixon announced a new national health strategy: the Health Maintenance Organization (Somers, 1971b; Starr, 1982). With increasing requests to assist newly forming prepaid group practices, KP curbed its expansion efforts in order to consult to new HMOs. One KP leader recalled

the consensus within the Kaiser Permanente Committee was that direct expansion should be discontinued and that the Kaiser Permanente organization should expend its

evangelical zeal by providing consultation to other sponsors that were willing to undertake development of prepaid group practice plans. (Fleming, 1997, p. 83)

In March 1971, KP held a symposium, sponsored by the Association of American Medical Colleges, the Commonwealth Fund, and KP, to describe and discuss its organization. There were 250 attendees, including 31 deans of medical schools; 19 Permanente physicians were featured speakers; and their presentations were published by the Commonwealth Fund (Somers, 1971b). Topics included the KP founding story, basic philosophy, and organizational structure; business aspects of the health plan; medical groups and physician profiles; special programs in research, education, and development; the expansion of KP into new areas; and the relationship between prepaid group practice and academic medicine (Somers, 1971b). KP's organizational form was presented as "*one valid solution* to some long-standing problems" (Keene, 1971, p. 4, emphasis in original), not a panacea. Speakers emphasized the way KP was a partnership of business and the medical profession, which recognized the "social responsibility of business" to serve the public (Trefethen, 1971). According to KP, "medical care [was] three things – a *profession*, a *business*, and a *social responsibility*" (Williams, 1971, p. 8, emphasis in original).

To encourage the development of more HMOs, the federal government passed the Health Maintenance Organization Act in 1973. The act subsidized the development of HMOs with the objective to have 1,700 HMOs established to serve 40 million Americans by 1976 (Starr, 1982; Morrison & Luft, 1990). The act also required that employers with 25 or more employees offer a federally qualified HMO as a health insurance option (Health Maintenance Organization Act of 1973, 1974). The organizational form HMO included the following characteristics: (1) "the HMO assumes a contractual responsibility to provide ... health services ... (to) a population defined by enrollment in the plan"; (2) "subscriber enrollment is voluntary"; (3) "the consumer pays a fixed annual or monthly payment that is independent of the use of services"; and (4) "the HMO assumes at least part of the financial risk or gain in the provision of services" (Luft, 1987, p. 2).

These features were shared by "three major functional parts to an HMO that may even be legally distinct organizations: (1) the 'plan' that contracts with enrollees, (2) the physician group that provides medical services, and (3) the hospital that provides inpatient services" (Luft, 1987, p. 9). There could be considerable variations in the relationships between these parts. The HMO definitions only required at least one of the three groups to bear at least part of the overall financial risk of health care. Plans, physicians,



and hospitals could be tightly or loosely connected, with exclusive or multiple contracts, and widely varying services. There also were no restrictions on using fee-for-service reimbursement in addition to capitation, nor were there requirements for amounts of prepayments (Luft, 1987).

In fact, the HMO Act specifically defined two types of HMOs that the government would subsidize: prepaid group practices and individual practice associations (IPAs). The two types differed in their organization of physicians, particularly the connections to and the amount of involvement with specific HMOs. Prepaid group practice HMOs would either employ physicians or contract with multispecialty groups to provide care for their members. IPAs would have a more traditional physician/insurer relationship, with solo practitioners and small groups of physicians contracting independently with HMO insurers to provide care (Christianson, Sanchez, Wholey, & Shadle, 1991). These two types of HMOs spoke to two different audiences: the IPA reassured organized medicine that physicians could maintain autonomy from health plans, while the prepaid group practice HMO legitimized a model of organizing that had been on the fringes of the health care field (Caronna, 2004). These differences in subform and audience created a context in which many variations of the HMO could develop.

After the HMO Act was passed in 1973, KP leaders were concerned that the definition of HMO

left out the vital facilities and full-time group practice Garfield insisted were integral to success. There were further critical deviations from the Kaiser Permanente model. An HMO could be 'funded privately, publicly or be a combination of both; it may be for-profit or non-profit.' Doctors could practice full-time or part-time with the HMO and could be salaried or paid fee-for-service. By the end of 1973 there were over sixty so-called HMOs with eight million subscribers in a patchwork combination of prepaid group and fee-for-service rate systems, some with their own hospitals, some centered only around Independent Practice Associations. These definitions of the HMO bore little resemblance to the pragmatic operating principles developed by early Kaiser Permanente organizers, or to the social medicine philosophy of the ideologues among Permanent physicians. (Hendricks, 1993, pp. 215–216)

KP leaders joined with other prepaid group practices to form a "Consensus Group" that lobbied for amendments to the HMO Act. The group included representatives from KP, the Group Health Association of America, the American Group Practice Association, the American Association of the Foundation of Medical Care, the Health Insurance Association of America, and Blue Cross (Smillie, 1991). The Act was amended in 1976 to relieve some of the "onerous" requirements and also allow subsidies to for-profit HMOs (Christianson et al., 1991). Almost immediately, KP received federal

qualification, which was announced at ceremony in New York City in 1977 presided over by Joseph Califano, Secretary of Health, Education, and Welfare. At the same ceremony, Lady Bird Johnson presented the Lyndon Baines Johnson Foundation Award to Sidney Garfield for his humanitarian achievements. This attention on the national stage perhaps best symbolizes the sea change in KP's legitimacy, brought on primarily by change in the institutional environment rather than within the organization itself.

In the mid-1970s, KP made more direct efforts to further the HMO population's growth. Under the leadership of CEO James A. Vohs, in 1976 Kaiser Permanente Advisory Services (KPAS) was founded to provide extensive consultations to interested organizations (Cutting, 1986). Several retired Permanente physicians, including Cecil Cutting, traveled the country to spread the KP model. They recommended that plans "sooner or later must come to grips with the complete Kaiser model and must combine organization, delivery, and financing of ambulatory and hospital patient services in one total system" (Williams, 1971, p. ii; see also Collen, 1988). One leader recalled that

KPAS was explicitly on an evangelical mission – namely to identify potential strong sponsors for prepaid group practice and to encourage them and assist them in establishing such plans. The posture remained this way until late in the 1970s. (Fleming, 1997, p. 83)

As KP took a lead role in the formation of the HMO population, its leaders tried to reproduce itself in other organizations and in the organizational form itself. Because of its size, age, experience, and resources, KP was able to take an active and influential role in the new venture. But HMO development made little progress in the 1970s. By 1980, there were 9 million enrollees in HMOs (Miller & Luft, 1994) and about 200 HMOs, mostly locally sponsored and limited geographically (Christianson et al., 1991). Kaiser Foundation Health Plans accounted for almost half of the HMO enrollment (Gruber et al., 1988). It would have been difficult to predict, at the onset of the 1980s, that the health care field was about to experience another major shift in institutional logics that would bring KP's legitimacy yet again into question. At the end of the era of federal involvement, KP was at the height of its influence as an institutional entrepreneur and had no plans to dampen its leadership role in the health care field.

### *The Era of Managed Care, 1983–present*

Although KP's leaders eagerly embraced the new organizational form defined by the HMO Act, within two decades they were motivated to

differentiate themselves from this form and other HMOs. A combination of transformations of the institutional environment and the evolution of the HMO population away from KP's model changed KP's role within these contexts. A major consequence of entrepreneurial activity for this prototypical HMO was a rejection of the very label it once sought to create and claim.

In the 1970s, neither professional nor public regulatory controls were able to stem the rising costs of medical care. This failure, combined with the deregulation of many industries during President Ronald Reagan's administration, paved the way for the expanded role of market forces in the health care sector (Starr, 1982; Scott et al., 2000). These changes in governance and logics were accompanied by an influx of for-profit health care organizations into a previously nonprofit dominated field. Other changes occurred among providers, as increasing numbers of hospitals joined health care systems and physicians joined group practices or other similar arrangements. Businesses began forming purchasing alliances to increase their power in negotiations with insurance companies (Bergthold, 1990). As insurance companies had to compete for business, often by lowering rates, they began to pressure providers to reduce costs. Prospective payment and other legislation in the early 1980s encouraged competition among providers and encouraged providers to reduce unnecessary services, shorten the length of hospital stays, and provide medical care for less than the set rate in order to profit (Scott et al., 2000).

Within this context of increasing market controls and corporate logics and encouragements from prospective payment, the HMO industry grew rapidly. Unlike HMOs in the 1970s, which tended to be small and localized, HMOs in the 1980s tended to be large, national organizations (Gruber et al., 1988). There also was a shift in the types of HMOs that developed, such that the term "managed care" began to be used to more accurately describe insurance plans and medical organizations with roots in the HMO movement. In the early 1980s, IPAs and a new type of HMO, the Network model,<sup>4</sup> predominated HMO growth.

Network HMOs and IPAs were more flexible than traditional HMO models and were easier and less expensive to organize. Because they offered more freedom and flexibility for members and preserved traditional arrangements between patients and doctors while still controlling costs, they quickly overtook prepaid group practice models in numbers (Gruber et al., 1988). Prepaid group practice models continued to have higher enrollments than IPAs and Network HMOs, though, and these enrollment differences

buffered traditional HMOs, including KP, from the effects of increasing competition for several years (Gruber et al., 1988).

In the early 1980s, KP leaders confidently noted that many organizations were adopting their approaches to organizing and financing health care (KPMCP, 1987) and KP took opportunities to further the development of prepaid group practice. A “Statement of Expansion Policy” in 1983 identified a program objective of “encourag(ing) and assist(ing) in making group practice prepayment health care coverage more widely available in the U.S.” (Statement of Expansion Policy of the Kaiser Permanente Medical Care Program, 1983, p. 1). But a year later, KPAS was dismantled because

some members of the Kaiser Permanente Committee became concerned that ‘we were educating the competition.’ They felt that Kaiser Permanente Advisory Services should discontinue its evangelism and devote its attention to program-oriented activities with the door being open to continue providing assistance to plans that we were philosophically comfortable with. Potential candidates were not explicitly restricted to nonprofit plans but certainly did not encompass the wheeler-dealers like Maxicare. (Fleming, 1997, p. 84)

In 1986, 59 percent of HMOs were for profit, and “with the exception of Kaiser Foundation Health Plans, investor-owned HMOs dominated the industry” (Gruber et al., 1988, p. 204). Intense competition caused the development of hybrid forms of HMOs offering more flexibility, the rapid growth of Preferred Provider Organizations (PPOs) (Morrison & Luft, 1990), and point-of-service (POS) plans (Miller & Luft, 1994). As KP’s leaders realized their newfound need to compete with very different types of HMOs than they had intended to be created, entrepreneurial efforts were ended in favor of self-reflection and change. As one Permanente physician leader stated in a 1987 interview:

On the national level, I think we’re in a position now where we can no longer modify the environment that we live in like we did with the government relations activity. The competition environment is going to develop on its own. There’s lots of money now for HMO development. It’s private money, investor money. New ideas and new concepts in the delivery of care are arising very rapidly, and they’re being marketed very effectively. Kaiser Permanente is going to be at a disadvantage, I think, unless they become more flexible in their approach to the organization of the delivery system itself .... (Smillie, 1987, p. 131)

In addition to the demands of intensified competition, a weakening of members’ commitment to KP’s identity created concern within the organization. The more KP was accepted by organized medicine, and the larger it

got (Cutting, 1986), the less its members felt an urgent need to protect its uniqueness. In 1986, Former CEO Clifford Keene reflected:

Certainly there isn't the atmosphere and perhaps the attitude of the frontiersman. We all had the personal perception that the medical world outside was against what we were doing. We had the perception on the inside that what we were doing was a good thing, and that it would be successful. We were determined to make it successful, and we were determined to make other doctors respect us. I don't think that evangelistic zeal is necessary now. No Kaiser person has to prove to anybody else that the Kaiser thing is workable or that it delivers good care or that it's a respectable, fine organization. So in that sense it's lost that pioneering spirit. (Keene, 1986, pp. 136–137)

By the late 1980s, KP seemed to lack the resources needed to shape the population of HMOs and the health care field. Dissatisfied with the market-driven direction new HMOs were taking, KP's leaders traded entrepreneurialism for retrenchment and disassociation, to differentiate their organization from the HMO form. An article in KP's 1987 annual report read:

As an established nonprofit prepaid group practice plan, Kaiser Permanente has resisted being classified with other types of managed care plans, even under the HMO label. "We have absolutely nothing in common with some of them, particularly those that offer a laundry list of different plans and arrangements in efforts to gain market share," says Bruce Sams, M.D., Medical Director in the Northern California Region. (KPMCP, 1987, p. 6)

In the 1990s, KP faced a number of challenges, including stagnating membership growth, downsizing and reorganization, the introduction of components contrary to its original principles, its first national level financial loss, and the growing sense that "its relative importance to the industry by both measures [enrollment and number of HMOs] [had] declined substantially" (Christianson et al., 1991, p. 22; see also Caronna, 2000). Its organizational identity lacked resonance with changes in the field and lacked the sharp focus it once had, when forced by circumstances outside its control to maintain its own facilities and contract only with Permanente physicians. In 1994, CEO David Lawrence provided further evidence of the disassociation strategy:

I think there is a real difference between managed care and Kaiser Permanente and organizations like ours ... what distinguishes Kaiser Permanente is that ... we are concerned about the way that health care is organized ... we have balanced the points of view of physicians and management ... deliberately in the form in which we have organized as a partnership between medicine and traditional management ... we are really focusing very heavily on evidence-based, clinical decision making ... these factors make us very different from other forms of managed care. In fact, we are so much different that I myself fight the term "managed care". (Stenger, 1994, p. 66)

Comments on KP's differences from most managed care organizations were threaded throughout documents, interviews, and annual reports. In 1988, CEO James Vohs described KP as a prepaid group practice instead of as an HMO: "the group practice form of prepaid care that we have developed over time is the most effective way to organize and finance health care delivery ... and we should never ever abandon [that] form" (Traska, 1988, p. 109).

KP's nonprofit status was first articulated as an operating principle in 1987 (KPMCP, 1987), in an attempt to differentiate the organization from the world of for-profit HMOs and emphasize its authenticity (Baron, 2004). In the 1990s, KP's nonprofit status communicated its social mission, its charitable care, and its community service. Reflecting this new emphasis, most annual reports in the 1990s contained some kind of community service report, and two annual reports were entirely dedicated to KP's charitable efforts (KPMCP, 1990, 1996). Notably, community service had not been given so much attention in the annual reports since the 1960s when Henry Kaiser wrote that community service was an "obligation" of private enterprise (KPMCP, 1961; Caronna, 2000). A page called "Who We Are" on the KP website in 2000 stated:

Kaiser Permanente is America's largest not-for-profit health maintenance organization ... As a not-for-profit organization, we are driven by the needs of our members and our social obligation to provide benefit for the communities in which we operate, rather than the needs of shareholders. (KPMCP, 2000c)

In response to competitor United Health Care's decision, in 1999, to place medical decisions in the hands of doctors, Francis Crosson, Executive Director of the Permanente Federation, wrote: "For 54 years, Permanente physicians, without interference, have made all medical decisions at Kaiser Permanente" (Crosson, 1999). The KP website explained further:

Doctor(s) can order any tests, medications, medical procedures or referrals they need without approval from someone in the health plan. Physicians have full authority to make health care decisions with their patients. Physicians also take the lead in determining all medical care policies for Kaiser Permanente. (KPMCP, 2000b)

In sum, at the turn of the twenty-first century, KP leaders differentiated their organization from the majority of HMOs by focusing on KP's "strong social purpose, physician responsibility for clinical care, and ... enduring partnership between [the] Health Plan and medical groups" (KPMCP, 2000a). Long gone were the days when KP actually could influence the structure, strategy, and values of other HMOs, even though other HMOs in theory were conforming to the model of organizing KP helped create.

## DISCUSSION AND CONCLUSION

This case study of KP and the HMO organizational form reveals both promising aspects and potential limitations of our understandings of institutional entrepreneurship. Importantly, it illustrates the dynamic nature of the relationship between organizational identity, organizational form creation, and the institutional environment. In the era of professional dominance, KP was imprinted with principles counter to prevailing institutional logics, which led to struggles to obtain legitimacy from professional providers and pressure to change its distinctive identity. In the era of federal involvement, shifts in the institutional environment facilitated KP's emergence as an institutional entrepreneur – its once illegitimate identity now resonated with new macro-level logics, beliefs, and goals. In the context of newfound legitimacy and national attention, KP capitalized on its organizational and institutional resources to influence the creation and revision of the HMO form, in part using its own organizational identity to help define the HMO. In the era of managed care, KP reacted to the proliferation and diffusion of the HMO form as very few new HMOs were created in KP's likeness. KP found its identity challenged and reinterpreted by actors in its institutional environment, and although its legitimacy did not diminish, its authenticity as a nonprofit organization was questioned. In each era, KP's abilities to sustain and organize principles and influence the institutional environment depended on the fit between identity and institutions.

One promising avenue of research on institutional entrepreneurs illuminated by this analysis involves studying founders of new ventures. Aldrich (1999, p. 77, *emphasis in original*) points out that researchers “typically identify founders only *after* [they] have already identified their organizations.” Locating an individual entrepreneur before he or she has created any measurable outcome is a difficult research problem. At a higher level of analysis, though, it is possible to identify organizations as potential entrepreneurs before they develop organizational forms, to study failed attempts of organizations to establish new organizational forms, and to locate organizations that are innovative but do not take part in form development. For example, Clark's (1970) study of distinctive colleges described how Reed College deviated from standard forms but did not attempt to become a model for other organizations. Being able to identify organizational founders before they engage in entrepreneurial activity has the potential to advance theory and research. The nascent individual entrepreneur “begin(s) with almost nothing but [his or her] intentions” (Aldrich, 1999, p. 112), but

the nascent institutional entrepreneur begins with the resources of an existing organization.

There also is a potential to exploit parallels between individual theories of entrepreneurship and the experiences of institutional entrepreneurs. Dobrev and Barnett's (2005) theory of individual entrepreneurs predicts that a founder's commitment to his or her new venture will decline as the organization ages, increasing the chance that the founder "will leave the organization in search of another context where he can regain his identity of a founder" (Dobrev & Barnett, 2005, p. 436). Part of the push to leave the organization stems from the founder's loss of control over the new venture, which leads to a process of differentiation between the founder and organization. In KP's case, as the HMO form evolved, KP's leaders felt they looked less and less like an HMO. In reaction to negative evaluations of HMOs, KP's leaders attempted to distance themselves from other organizations with the same form. The model HMO tried to claim that it was not an HMO after all. This experience fits with Dobrev and Barnett's expectations.

The application of an individual-level theory to an organizational entity is not entirely smooth, however. It is relatively easy to know when an individual has left an organization, but how does one know when an organization has abandoned an organizational form? For KP, "leaving" the HMO population was much more symbolic than real. Leaders could and did make claims that their organization did not fit the HMO form, but for all intents and purposes, KP was and still is an HMO. If an organization maintains its structure but interprets its identity differently, has it "left" its new venture? If external evaluators buy the new interpretation and regard the organization as different, does that count as leaving? Does an organization have to attempt to found a new organizational form in order to "regain [its] identity" (Dobrev & Barnett, 2005, p. 436)? In addition, KP was decades old when its leaders began to reject the HMO label. How might leaders of a younger, more flexible organization react to a new venture that takes off in unintended directions? It seems highly probable that a younger organization would adapt to the form as it evolves and accept a modified organizational identity, since rejecting the form could lead to a loss of legitimacy.

There are other issues beyond questions of identity when comparing individual and institutional entrepreneurs. What motivates the entrepreneur seems like a crucial difference across levels of analysis. Individuals create organizations to accomplish goals, spread innovations, and get rich, but why do organizations construct a form? In KP's case, leaders had an altruistic sense that they were spreading a superior model of health care. The



Kaiser Permanente Advisory Committee freely gave away the organization's "formula," hoping to inspire other organizations to adopt their principles. What kind of individual entrepreneurs happily and proudly distribute their innovations, without compensation? In addition, KP's leaders took on the HMO movement in order to increase the legitimacy of the prepaid group practice. By establishing a legitimate organizational form in their likeness, the new form could then reflect legitimacy back to the organization. The parallel would be an individual establishing an organization just like him-/herself in order to become more legitimate as a person, which does not make a lot of sense. Organizations and organizational forms analytically are much more alike than individuals and organizations, which also complicates potential research.

Another problem is the complexity of identifying the founders of an organizational form. This paper presents the experience of just one of the founders of the HMO form, but KP was by no means alone in its efforts. Other entities that could be considered founders include the Group Health Cooperative of Puget Sound, the Health Insurance Plan of Greater New York, and other prototypical prepaid group practices, the Group Health Association of America, the American Medical Association, the Johnson Administration, the Nixon Administration, the Congress, various policy makers, advocates, and think tank members, not to mention Paul Ellwood himself (see [Rao et al., 2000](#) for a version of HMO development focused on Ellwood). It should be noted that not all of these founders were actively and intentionally involved in codifying the form, but each had some degree of influence on how the form was defined and developed. With so many different players, agendas, and audiences, it is much more difficult to track the entirety of the entrepreneurial effort involved in organizational form creation and the consequences of form creation for the entrepreneurs.

Institutional entrepreneurship raises the issue that there are really two entrepreneurial stories underlying form creation, and potentially many more. An organization that founds an organizational form was itself founded at some point. An individual entrepreneur created the organization in a particular historical moment, and even if that particular founder is long gone, the entrepreneurial activity taken on by an organization is in part a reflection of that founder and founding story. Do analyses of organizational form creation need to begin with analyses of organizational foundings? Examining institutional entrepreneurship does not eliminate the individual level, just as stories of individual entrepreneurship cannot ignore institutions.

The idea that organizational form creation is a type of entrepreneurship raises additional issues about the forms themselves, for example, when does

a form originate and who or what creates it? Ruef (2000) argues that “once health care professionals became widely aware of HMOs as a distinctive organizational arrangement, the form had been in the process of emergence for at least half a century” (p. 668). In my analysis, given KP’s experience, in the 1940s and 1950s prepaid group practices were too illegitimate and their leaders sanctioned too much to consider this a period of form emergence. If form creation takes entrepreneurial effort, KP and other prepaid group practices were too busy surviving in this era to intentionally help develop a model of organizing. But Rao et al. argue that HMO development was a product of a spin-off movement of the health care consumer movement of the 1960s involving Ellwood and his “activist network” (Rao et al., 2000, p. 266). This account overlooks the efforts of the Group Health Association of America, which organized prior to the consumer movement, and the importance of the simple presence of prepaid group practices since the 1930s as a building block in the institutional environment. Differences in analyses of HMO form creation illustrate the complexity of pinpointing when a form emerges, when the process begins, and where the form exists.

Ruef’s (2000, p. 668) insightful point that the “cognitive recognizability of the HMO form among the general public did not become firmly entrenched until a number of years [after the HMO Act]” raises additional issues. If a form is not a social code (Carroll & Hannan, 2000), in that there are no sanctions if an organization deviates from the form, then is the form really a form? The HMO Act of 1973 codified a definition of HMO that included two subtypes, intending the definitions to become blueprints for new organizations. Since the federal government had to include subsidies for individuals to create HMOs, and since so few individuals actually took the subsidies and succeeded, in the 1970s were the codified definitions of HMO and IPA actually organizational forms? “Code enforcement” in terms of punishing organizations that deviated from managed care developed in the 1980s, not for reasons of social legitimacy but economic competitiveness. Thus, did the organizational form of HMO really emerge in the mid-1980s, in response to prospective payment legislation? By then, was the HMO form so different from KP that KP was not a founder after all? These are questions this paper cannot answer, but seem crucial for advancing our understanding of organizational forms.

A final issue to consider is the special nature of KP. Very few organizations begin with the backing of someone like Henry Kaiser, in terms of economic capital, social capital, and pure entrepreneurial spirit. KP managed to survive and grow for decades despite vigorous persecution from organized medicine. Although external evaluators tried to constrain KP’s

identity, the organization managed to find solutions to get around these constraints (for example, when hospitals denied Permanente physicians privileges, KP built its own hospitals). In contradiction to the claims of institutional theorists, the lack of legitimacy did not deprive KP of sufficient resources to ensure its failure. In contradiction to the claims of organizational ecologists, the features that made up its identity existed in spite of the constraints of the field. This experience may not be typical, yet the existence of an organization like KP reveals that our assumptions about identity and legitimacy, from both institutional theory and organizational ecology perspectives, need to more carefully consider how entrepreneurs negotiate the dynamic interrelationship between identities and institutions in order to sustain illegitimate identities and institutionalize innovative organizational forms.

## NOTES

1. In 1945, the organization that provided health insurance was called the Permanente Health Plan, named after the Permanente Creek in the Santa Cruz Mountains of California (Keene, 1971). The physicians called their group practice the Permanente Medical Group. In 1952, Henry Kaiser changed the name of the health plan to the Kaiser Foundation Health Plan and Hospitals to memorialize his late wife Bess (Heiner, 1989); the physicians retained the name Permanente.

2. Originally, the written principles included only the first four listed. Dual choice was formally added in 1952, physician responsibility in 1971, and nonprofit status in 1987.

3. The Permanente Medical Groups, which are separate legal entities from the Kaiser Foundation Health Plan and Hospitals, are for-profit corporations.

4. Traditional HMOs include the Staff HMO and the Group HMO. Both deliver services through a group practice established to provide health services to HMO members. In the Staff HMO, physicians are salaried staff of the HMO; in the Group HMO, the health plan contracts with a group practice to provide services, which is reimbursed on a capitation basis. An IPA contracts with an association of physicians from various settings and reimburses practitioners on a fee-for-service basis. A Network HMO reimburses contracted practitioners on a capitation basis (Christianson et al., 1991).

## ACKNOWLEDGMENTS

Martin Ruef and Michael Lounsbury provided helpful comments and suggestions that strengthened this paper considerably. Thank you both for inviting me to contribute to this volume.

## REFERENCES

- Abbott, A. (1988). *The system of professions: An essay on the division of expert labor*. Chicago: University of Chicago Press.
- Adams, S. B. (1997). *Mr. Kaiser goes to Washington: The rise of a government entrepreneur*. Chapel Hill, NC: University of North Carolina Press.
- Albert, S., & Whetten, D. A. (1985). Organizational identity. *Research in Organizational Behavior*, 7, 263–295.
- Aldrich, H. (1999). *Organizations evolving*. London: Sage.
- American Medical Association. (1997). *AMA membership 1947–1996 (internal reports)*. Chicago: Division of Survey and Data Resources, American Medical Association.
- Baron, J. (2004). Employing identities in organizational ecology. *Industrial and Corporate Change*, 13, 3–32.
- Bergthold, L. (1990). *Purchasing power in health: Business, the state, and health care politics*. New Brunswick, NJ: Rutgers University Press.
- Brown, L. D. (1983). *Politics and health care organization: HMOs as federal policy*. Washington, DC: The Brookings Institution.
- Campion, F. D. (1984). *The AMA and U.S. health policy since 1940*. Chicago: Chicago Review Press.
- Caronna, C. A. (2000). *Organizational identity in changing contexts: The case of Kaiser Permanente and the U.S. health care field*. Unpublished dissertation. Stanford University, Stanford, CA.
- Caronna, C. A. (2004). The misalignment of institutional ‘pillars’: Consequences for the U.S. health care field. *Journal of Health and Social Behavior*, 45, 45–58.
- Carroll, G. R., & Hannan, M. T. (2000). *The demography of corporations and industries*. Princeton, NJ: Princeton University Press.
- Christianson, J. B., Sanchez, S. M., Wholey, D. R., & Shadle, M. (1991). The HMO industry: Evolution in population demographics and market structures. *Medical Care Review*, 48, 3–46.
- Clark, B. R. (1970). *The distinctive college: Antioch, Reed and Swarthmore*. Chicago: Aldine.
- Collen, M. F. (1988). *History of the Kaiser Permanente medical care program, an oral history conducted in 1986 by Sally Smith Hughes*. Berkeley: Regional Oral History Office, The Bancroft Library, University of California.
- Crosson, F. J. (1999). United decision to place medical decisions in the hands of doctors emulates 54-year practice at Kaiser Permanente. Statement from November 9. Available at: <http://www.kaiserpermanente.org/newsroom/releases/statements.html>. Website accessed on April 18, 2000.
- Cutting, C. C. (1971). Historical development and operating concepts. In: A. R. Somers (Ed.), *The Kaiser-Permanente medical care program: One valid solution to the problem of health care delivery in the United States* (pp. 17–22). New York: The Commonwealth Fund.
- Cutting, C. C. (1986). *History of the Kaiser Permanente medical care program, an oral history conducted in 1985 by Malca Chall*. Berkeley: Regional Oral History Office, The Bancroft Library, University of California.
- DiMaggio, P. J. (1988). Interest and agency in institutional theory. In: L. G. Zucker (Ed.), *Institutional patterns and organizations: Culture and environment* (pp. 3–22). Cambridge, MA: Ballinger.

- DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48, 147–160.
- Dobrev, S. D., & Barnett, W. P. (2005). Organizational roles and transitions to entrepreneurship. *Academy of Management Journal*, 48, 433–449.
- Dutton, J. E., & Dukerich, J. M. (1991). Keeping an eye on the mirror: The role of image and identity in organizational adaptation. *Academy of Management Journal*, 34, 517–554.
- Fleming, S. (1997). *History of the Kaiser Permanente medical care program, an oral history conducted in 1990 and 1991 by Sally Smith Hughes*. Berkeley: Regional Oral History Office, The Bancroft Library, University of California.
- Foster, M. S. (1989). *Henry J. Kaiser: Builder in the modern American West*. Austin, TX: University of Texas Press.
- Freidson, E. (1970). *Professional dominance: The social structure of medical care*. Chicago: Aldine.
- Gioia, D. A., & Thomas, J. B. (1996). Identity, image, and issue interpretation: Sensemaking during strategic change in academia. *Administrative Science Quarterly*, 41, 370–403.
- Golden-Biddle, K., & Rao, H. (1997). Breaches in the boardroom: Organizational identity and conflicts of commitment in a nonprofit organization. *Organization Science*, 8, 593–611.
- Greenwood, R., & Hinings, C. R. (1993). Understanding strategic change: The contribution of archetypes. *Academy of Management Journal*, 36, 1052–1081.
- Greenwood, R., & Hinings, C. R. (1996). Understanding radical organizational change: Bringing together the old and the new institutionalism. *Academy of Management Review*, 21, 1022–1054.
- Greenwood, R., Suddaby, R., & Hinings, C. R. (2002). Theorizing change: The role of professional associations in the transformation of institutionalized fields. *Academy of Management Journal*, 45, 58–80.
- Gruber, L. R., Shadle, M., & Polich, C. L. (1988). From movement to industry: The growth of HMOs. *Health Affairs*, 7, 197–208.
- Hannan, M. T., & Freeman, J. (1989). *Organizational ecology*. Cambridge, MA: Harvard University Press.
- Health Maintenance Organization Act of 1973. (1974). *Explanation of act and text of public law 93-222. Prepared for the subcommittee on health of the committee on labor and public welfare, United States Senate*. Washington, DC: U.S. Government Printing Office.
- Heiner, A. P. (1989). *Henry J. Kaiser American empire builder: An insider's view*. New York: Peter Lang.
- Hendricks, R. (1991). Medical practice embattled: Kaiser Permanente, the American medical association, and Henry J. Kaiser on the West Coast, 1945–1955. *Pacific Historical Review*, 60, 439–473.
- Hendricks, R. (1993). *A model for national health care: The history of Kaiser Permanente*. New Brunswick, NJ: Rutgers University Press.
- Hoffman, A. J. (1999). Institutional evolution and change: Environmentalism and the U.S. chemical industry. *Academy of Management Journal*, 42, 351–371.
- Kay, R. H. (1987). *History of the Kaiser Permanente medical care program, an oral history conducted in 1985 by Ora Huth*. Berkeley: Regional Oral History Office, The Bancroft Library, University of California.
- Keene, C. H. (1971). Kaiser industries and the Kaiser-Permanente health care partnership. In: A. R. Somers (Ed.), *The Kaiser-Permanente medical care program: One valid solution to*

- the problem of health care delivery in the United States* (pp. 13–16). New York: The Commonwealth Fund.
- Keene, C. H. (1986). *History of the Kaiser Permanente Medical Care Program, an oral history conducted in 1985 by Sally Smith Hughes*. Berkeley: Regional Oral History Office, The Bancroft Library, University of California.
- Kennedy, E. M. (1972). *In critical condition*. New York: Simon and Schuster.
- KFMCP (Kaiser Foundation Medical Care Program). (1960–1978). Annual Reports.
- KPMCP (Kaiser Permanente Medical Care Program). (1979–1997). Annual Reports.
- KPMCP. (2000a). Fast facts about Kaiser Permanente. Available at: <http://www.kaiserpermanente.org/fastfacts.html>. Website accessed on April 18, 2000.
- KPMCP. (2000b). In the hands of doctors. Available at: <http://www.kaiserpermanente.org/healthplans/doctorshands.html>. Website accessed on April 18, 2000.
- KPMCP. (2000c). Who we are. Available at: <http://www.kaiserpermanente.org/about/whoweare.html>. Website accessed on April 18, 2000.
- Krause, E. (1997). *Death of the guilds: Professions, states, and the advance of capitalism, 1930 to the present*. New Haven, CT: Yale University Press.
- Link, G. E. (1986). *History of the Kaiser Permanente medical care program, an oral history conducted in 1985 by Malca Chall*. Berkeley: Regional Oral History Office, The Bancroft Library, University of California.
- Lounsbury, M., & Glynn, M. A. (2001). Cultural entrepreneurship: Stories, legitimacy, and the acquisition of resources. *Strategic Management Journal*, 22, 545–564.
- Luft, H. S. (1987). *Health maintenance organizations: Dimensions of performance*. New Brunswick, NJ: Transaction Books.
- Maguire, S., Hardy, C., & Lawrence, T. B. (2004). Institutional entrepreneurship in emerging fields: HIV/AIDS treatment advocacy in Canada. *Academy of Management Journal*, 47, 657–679.
- Marmor, T. R. (1970). *The politics of medicare*. Chicago: Aldine.
- Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, 83, 340–363.
- Miller, R. H., & Luft, H. S. (1994). Managed care plans: Characteristics, growth, and premium performance. *Annual Review of Public Health*, 15, 437–459.
- Morrison, E. M., & Luft, H. S. (1990). Health maintenance organization environments in the 1980s and beyond. *Health Care Financing Review*, 12, 81–90.
- Rao, H., Morrill, C., & Zald, M. N. (2000). Power plays: How social movements and collective action create new organizational forms. In: B. M. Staw & R. I. Sutton (Eds), *Research in organizational behavior* (pp. 237–281). New York: Elsevier Science.
- Ruef, M. (2000). The emergence of organizational forms: A community ecology approach. *American Journal of Sociology*, 106, 658–714.
- Ruef, M., Aldrich, H. E., & Carter, N. M. (2003). The structure of founding teams: Homophily, strong ties, and isolation among U.S. entrepreneurs. *American Sociological Review*, 68, 195–222.
- Ruef, M., & Scott, W. R. (1998). A multidimensional model of organizational legitimacy: Hospital survival in changing institutional environments. *Administrative Science Quarterly*, 43, 877–904.
- Scott, S. G., & Lane, V. R. (2000). A stakeholder approach to organizational identity. *Academy of Management Review*, 25, 43–62.
- Scott, W. R. (2001). *Institutions and organizations* (2nd ed.). Thousand Oaks, CA: Sage.

- Scott, W. R., Ruef, M., Mendel, P. J., & Caronna, C. A. (2000). *Institutional change and healthcare organizations: From professional dominance to managed care*. Chicago: University of Chicago Press.
- Smillie, J. G. (1987). *History of the Kaiser Permanente Medical Care Program, an oral history conducted in 1986 by Ora Huth*. Berkeley: Regional Oral History Office, The Bancroft Library, University of California.
- Smillie, J. G. (1991). *Can physicians manage the quality and costs of care? The story of the Permanente medical group*. New York: McGraw-Hill.
- Somers, A. R. (1971a). *Health care in transition: Directions for the future*. Chicago: Hospital Research and Education Trust.
- Somers, A. R. (Ed.) (1971b). *The Kaiser-Permanente medical care program: One valid solution to the problem of health care delivery in the United States*. New York: The Commonwealth Fund.
- Starr, P. (1982). *The social transformation of American medicine*. New York: Basic Books.
- Statement of Expansion Policy of the Kaiser Permanente Medical Care Program. (1983). Unpublished memo.
- Stenger, A. (1994). Kaiser Permanente's twist on managed care: A talk with David M. Lawrence, M.D. *Postgraduate Medicine*, 95, 65–68.
- Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review*, 20, 571–610.
- Time Magazine. (1953). Push-Button hospital. *Time Magazine*, June 29, pp. 55–56.
- Time Magazine. (1954). What is free choice? *Time Magazine*, July 5, p. 37.
- Traska, M. R. (1988). Vohs: Kaiser has withstood the test of time. *Hospitals*, June 5, pp. 108–109.
- Trefethen, E. E., Jr. (1971). The Kaiser health plan: A business-professional partnership. In: A. R. Somers (Ed.), *The Kaiser-Permanente medical care program: One valid solution to the problem of health care delivery in the United States* (pp. 8–9). New York: The Commonwealth Fund.
- Velie, L. (1953). Supermarket medicine. *Saturday Evening Post*, 225(51), 22–23, 44, 49–50.
- Whetten, D. A., & Godfrey, P. C. (Eds) (1998). *Identity in organizations: Building theory through conversations*. Thousand Oaks, CA: Sage.
- Williams, G. (1971). *Kaiser-Permanente health plan: Why it works*. Oakland, CA: Henry J. Kaiser Foundation.
- Yedidia, A. (1987). *History of the Kaiser Permanente medical care program, an oral history conducted in 1985 by Ora Huth*. Berkeley: Regional Oral History Office, The Bancroft Library, University of California.

# ENTREPRENEURSHIP AT THE MARGINS OF SOCIETY: FOUNDING DYNAMICS IN GRAY (SEX SHOPS) AND BLACK MARKETS (MAFIA)

Luca Solari

## ABSTRACT

*Institutional theory and organizational ecology have long proposed alternative (albeit not always contradictory) processes to interpret founding and creation of a novel organizational form. Much of the debate has dealt with the issue of how legitimation processes shape such important events or acts. Empirical research on both sides is rich with interesting results, while much of the controversy regards how legitimation is empirically captured and the ways it unfolds over time.*

*Recently, within organization ecology this specific issue has received increasing attention in the search for a theory of forms and identities. A central piece of the proposed theory links identities to specific audiences or constituencies, both internal and external, which act by attributing legitimation to novel constructions. The new formulation has originated different efforts aimed at better understanding how audiences develop and how they are shaped by wider social movements. Existing research has mainly been dealing with organizations (and forms), which appear to be legitimate (albeit not legitimated) from their inception, benefiting from*

---

**The Sociology of Entrepreneurship**

**Research in the Sociology of Organizations, Volume 25, 337–368**

**Copyright © 2007 by Elsevier Ltd.**

**All rights of reproduction in any form reserved**

**ISSN: 0733-558X/doi:10.1016/S0733-558X(06)25010-9**



*the generalized acceptance of business organizations in modern societies. Limited attention has been devoted to analyzing contrasted forms, i.e. organized forms of action which act at the border or outside the border of established economic and social action. I contend that it is by analyzing these extreme cases that a clearer interpretation of legitimacy and legitimation processes can be achieved. By analyzing the evolution and the principal dynamics of three populations that are operating in gray and black market, I propose a critique to existing theories of legitimacy.*

The creation of a new organization attracts attention because it involves a complex web of social and economic processes that coalesce into an action and an actor or a group of actors (Ruef, Aldrich, & Carter, 2003; Shane & Venkataram, 2000). McMullen and Shepherd (2006, p. 134) define an entrepreneur as someone who “respond(s) and create(s) change through (his) entrepreneurial actions, where *entrepreneurial action* refers to behavior in response to a judgmental decision under uncertainty about a possible opportunity for profit”.

If entrepreneurs act on the basis of social representations of economic activity, it is rational for them to use the endowments of existing, legitimated forms of organizational activity, either imitating their attributes or establishing communication strategies in order to ensure that categorization processes allow them to be incorporated into existing belief systems and social representations. Such a process – commonly referred to as the legitimation process (Suchman, 1995) – has been at the core both of neo-institutional and organizational ecology. However, it may be hard to disentangle legitimation processes operating at the level of individual organizational forms from the general endowment stemming from the acceptance of organized forms of economic activity in modern societies (Strang & Meyer, 1994). When a new business is created in the form of an organization, is it being legitimated for its specific characteristics or as an organization “per se”?

Research on the legitimation of novel organizational forms has been generally concerned with the legitimation trajectory of populations of organizations, which overall adhere to existing belief systems and social representations, albeit differentiating in terms of the audiences and needs they address (Carroll & Hannan, 2000). Clearly, there are conceptual and empirical differences among the two major research traditions, neo-institutionalism and organizational ecology, on the concept of legitimacy. The debate over the issue of legitimacy is a long-standing one with occasional upheavals (Zucker, 1989; Carroll & Hannan, 1989; Baum & Powell, 1995;

Hannan & Carroll, 1995) that seem to center on the way legitimacy has been incorporated and treated by the two traditions.

In this paper, I propose to treat legitimation as a process of social categorization where language plays a key role, as discussed in Phillips, Lawrence, and Hardy (2004). Under this perspective, the relationship between density and legitimacy requires rethinking. After briefly reviewing the key issues in this debate, I propose to look at the differences in legitimation according to the context where it comes to operate. While most research has dealt with white markets, I contend that we can really understand the nature of legitimation only by looking at the dynamics of entrepreneurial activity in gray and black market, where we can rule out the background legitimacy conferred to business organizations in modernity.

I propose to analyze the paths to legitimacy – or at least diffusion – of three organizational populations that cover the continuum between black and white markets: mafia, sex shops, and samba schools. Based on these three accounts, I close with a discussion on where organization researchers should point their attention in defining a sound and convincing perspective on legitimation of a novel organization form and I review the first effort by other scholars along this direction.

## **LEGITIMACY AND DIFFUSION OF ORGANIZATIONAL FORMS**

When a new organizational form produces products and services that are inherently accepted by society, the diffusion process enjoys the initial endowment granted to any form of legal business activity. New products and services are diffused, they spur imitators, and if successful, originate the growth of a new industry or industry segment. Under these conditions, we empirically observe the classical dual density dependence, where density apparently absorbs the legitimacy effect, through taken-for-grantedness. This depiction of legitimation in organizational ecology is criticized by neo-institutionalists.

Strang and Meyer (1994) critique the idea that diffusion is only rooted in relational phenomena. The authors argue, instead, that research on diffusion should consider the importance of “culturally theorized understandings of the nature of social actors and of diffusing practices” (p. 100). Diffusion processes in reality are “complex exercises in the social construction of identity” (p. 102) where factors like perceived similarity, theorization of adopters, theorization of diffusing practices, and theorization as a diffusion

mechanism play an important role. Under this perspective, [Strang and Meyer \(1994\)](#) emphasize that modernity in itself satisfies these properties for diffusion, as we observe in the struggle for modernity in Brazil that brought to life the Samba schools which we will describe later on.

The controversy is hard to resolve, given the extraordinary amount of empirical data on the part of organizational ecologists. However, I think that a better ground for comparing the two perspectives is by looking at legitimation where it can hardly be backed by a general process of legitimation of business activity in modern societies. If we use data on new organizational forms that appear to operate in white markets, in fact, our results may be confounded by the fact that modernity in itself satisfies the major properties for diffusion as [Strang and Meyer \(1994\)](#) pointed out.

While business activity has enjoyed an evident diffusion from the early nineteenth century, giving rise to a never-ending ramification of industries through specialization and innovation, a striking feature of economic and social change is the blurring of boundaries between different arenas of society. Goods that once were believed to be out of the scope of business are nowadays at the core of important industries (a powerful example is the health care industry – see [Caronna \(2007, this volume\)](#); [Scott, Ruef, Mendel, & Caronna, 2000](#)), and this process is extending to more and more areas of collective production of value.

As [Lounsbury and Rao \(2004\)](#) have pointed out, the increasing specialization of products and services is not a simple consequence of innovation in technical features, but the outcome of a complex pattern of definition and redefinition of the existing structure of categories within industries. While [Lounsbury and Rao \(2004\)](#) document this process within the mutual fund industry, their perspective could be easily extended across industries and adopted to provide a different account of the creation of a new organizational form. If social categories should be interpreted as “outcomes negotiated by participants in organizational fields” ([Lounsbury & Rao, 2004, p. 970](#)), and new organizational forms become social categories ([Hannan, Polos, & Carroll, 2004](#)), legitimation should not be considered density-dependent, but a complex process emanating from interactions among actors. The fact that density empirically fits the data of a dual density-dependent model does not solve this theoretical inconsistency, because it is possible that, once acquired, legitimacy paves the way for diffusion rather than diffusion paving the way for legitimacy. In such a case, empirical results should be considered the result of a mistaken causality between density and legitimacy. In the meantime if we consider legitimacy as a discrete status (existing or not), originating from a process of interaction where

“quantities” (e.g. the count of articles on a new product, or any other quantitative measure) do not count, it is evident that it does not show up in models because (1) it operates before the birth of the first organization in a new form, and (2) it cannot be measured and weighted.

Considering legitimacy as a result of a complex process of legitimation requires us to explicitly analyze its unfolding and relate it to a process of social categorization.

## LEGITIMACY AS CATEGORIZATION

The process of legitimation is analogous to a process of categorization, by which an organization is either attributed to a cluster of previously known organizations or declared dissimilar and attributed to a novel category. Membership in a category represents the target for entrepreneurs when a prize is awarded, but it can also mean exposure to potential sanctions for violation of characteristics considered relevant by audiences. This process is ambiguous for entrepreneurs because it presents a trade-off. On one side, it is easier to gain legitimacy and its prize by adhering to a pre-existing blueprint. On the other, a greater prize might be attained by presenting the organization as different from any other incumbent. Therefore, exploring legitimation requires a closer look at social categorization processes.

The nature of categorization and attribution processes is a key issue in defining what an organizational form is. According to [Arcuri and Castelli \(2004\)](#) there are two approaches. First, an “Aristotelian” process, which posits that categories are to be defined using a limited set of necessary and sufficient criteria that allow us to categorize and assign membership to a category (rule-based approach). It is the ideal method for highly structured field of knowledge like geometry or mathematics. The second approach (similarity-based) sustains the idea that natural categories cannot be defined on the basis of necessary and sufficient attributes and conditions. Membership into a category is not clear-cut, but rather foggy. Under such premises, a major issue becomes the definition of degrees of similarity that give origins to prototypes (ideal collection of attributes that represent a category) or exemplars (a set of attributes derived by past experiences and encounters).

There are two elements of research in categorization that are relevant to the development of our field. First, the structuration of categories into hierarchical attributes ([Rosch, 1978](#)) allows us to see patterns of use of categories dependent upon the context. Research has suggested that individuals

will tend to use base categories most of the time, resorting to higher level categories when they need to face important issues that affect values and perspectives. Second, research suggests that prototypes and exemplars come into play according to situations. When we need to describe unknown realities we will resort to prototypes, while we will employ exemplars when we face an extraordinary event or experience.

Moreover, categorization processes within and between groups have been demonstrated to vary greatly. Group members perceive the group as complex and articulated, while members of external groups are considered highly consistent and virtually undifferentiated. [Park and Rothbart \(1982\)](#) and [Linville, Fischer, and Salovey \(1989\)](#) provide two alternative explanations. According to [Park and Rothbart \(1982\)](#), categories used within groups are more detailed than categories used between groups. Fischer and Salovey (1989) emphasize the availability of information and perceptions, which are more fine-grained for group members than for outsiders.

Research in social cognition emphasizes the fuzziness of categories and clearly implies heterogeneity among individuals and variation (through time/context) within single individuals. The model of exemplars, though, allows categorization to be determined by experience and encounters with individuals claiming to pertain to a pre-form. A pre-form can be defined as an entity which clearly is not yet a form, but aspires at becoming one and is actively enforced by actors interested in its legitimation into a form. A major tool in achieving the attention of an audience and developing into a full-fledged form is language in the form of definition of one's desired identity. The relationship between legitimacy and language is a key element absent in the contemporary debate in organization theory, while it is evident that even if we accepted density-dependent legitimation, we should consider that it requires new members of a population to be perceived as different from other organizations and consequently attributed a new name.

## **LEGITIMACY AND NEW ORGANIZATIONAL FORMS: THE ROLE OF LANGUAGE AND DISCOURSE**

In a recent article, [Phillips et al. \(2004\)](#) define institutions as “social constructions constituted through discourse” (p. 638) and emphasize the role of the production of texts in such a process. Texts explain, legitimate, validate, and promote the institutionalization of forms because they provide clear self-regulating, socially constructed mechanisms that enforce their application.

According to Phillips et al. (2004), texts are produced whenever either sensemaking or legitimation actions are required by events. Sensemaking (Weick, 1995) is triggered by surprises, puzzles, or problems which require the crafting of stories that attribute sense to events. Legitimation requires the creation of explanations and justifications for the existence of individuals, practices, and organizational forms.

The authors emphasize the importance of processes which allow the incorporation of texts into discourse. According to their view, this process is affected by the nature of the producer of a text, the genre of the text, and the existence of links to pre-existing texts and discourse. Once a text becomes part of discourse, its journey toward institutionalization is far from concluded. At this stage, the characteristics of the discourse in itself come to play a role. On one side, internal consistency and structuration of discourse heighten the chances of institutionalization because they provide a more precise account of events and practices to be institutionalized. On the other, a discourse is inserted into a dynamic context where it will interact with other legitimated discourses and its institutionalization will depend upon broader discourses which support it. At this stage, an ecology of discourse, much similar to the one suggested by Dawkins (1976) could be in place because institutionalization will be affected by competing discourses.

## **THE DEFINITION OF ORGANIZATIONAL FORMS AS SOCIAL CATEGORIZATION**

The difference in the two approaches to categorization (i.e. rule-based and similarity-based) is apparent in the comparison of two recent works by Hannan et al. (2004) and Polos, Hannan, and Carroll (2002).

In the 2002 article, the authors clearly address the issue of definition of a form adhering to a rule-based approach when they deal with identities as social codes and then somewhat revert to the idea of prototypes as basis for the definition of forms. More recently, the 2004 working paper takes a different stance. The authors affirm “we have shifted to the view that the audience perceives features and makes similarity judgments in terms of shared and divergent features among sets of organizations” (2004, p. 3) where they apparently resort to the similarity-based approach. Actually, the new position moves away from the idea of identity as created by internal audiences and form as enforced by external identity codes. This latter assertion in itself was sustained in the 2002 article by the authors affirming

“we elide most differences and assume that evaluations can be meaningfully aggregated within each of two sets of actors: insiders and outsiders”.

The new formulation assigns an important role to audiences, which appear to be at the core of the theoretical argumentation. Audiences are defined as “a set of agents with (1) an interest in the domain, (2) some control over material and symbolic resources that affect the valuations of some organizations in the domain and thereby affect the success and failure of those organizations in the domain, (3) relatively homogeneous domain-specific tastes”. The line of reasoning proceeds then from audiences to organizations, but it fails to address an apparent tautology in the logic when the concept of domain appears to be defined by the existence of at least one audience with a domain-wide interest, while organizations “announce their claims to competence for membership in a domain” (2004).

A more subtle problem is the explicit adoption of both approaches to social categorization as if they were easily combined. My view is that this amounts to an epistemological violation and actually appears as just an elegant way out that allows the paper to later resort once again to a rule-based approach. Simply speaking, audiences scan a domain, find a set of organizations that closely resemble one another, define a name for them (typification), assign a name to a code (a set of feature values considered appropriate for that name by an audience), and use codes to “sweep away (...) ambiguity” constructing the feature space.

Moreover, it is evident from social categorization literature that the categorization of a novel organizational form will take different shapes according to the nature of the audiences, which act in the process. Insiders will tend to define narrow categories and perceive the form as radically different from existing forms, while outsiders will probably relate the form to broader categories. For instance, entrepreneurs and employees of a sex shop will describe in detail the nature of processes and services and attempt at communicating it and legitimating it, while external audiences will possess different schemes. Potential customers will relate a sex shop to known products and past experiences in different settings (for instance in other countries) and act accordingly in the legitimacy process. Others will probably relate a sex shop to a broader category, like “sex” and react according to their personal stance on such a delicate issue: either providing “passive” legitimacy (accepting the existence of this form), active legitimacy (defending the right of such a form to exist, though not being involved neither as participants nor customers), or opposing the form (trying to restrain its diffusion or even calling for institutional action to cancel it from the social landscape) (Suchman, 1995). According to this depiction of processes of

social categorization, the idea of legitimacy being acquired through taken-for-grantedness (prompted by mere diffusion of organizations claiming to adhere to a pre-form) clearly appears as a very special case. If we follow the perspective by Phillips et al. (2004), taken-for-grantedness depends more on social construction than on availability of material elements and encounters with a novel organizational form. Hence, if density could capture legitimacy, then this should be a specific case where we can prove that the number of organizations adhering to a prototypical “pre-form” (i.e. something which clearly is not yet a form, but aspires at becoming one) interferes with social construction processes, for instance increasing the probability that powerful actors will incorporate text into their discourse or that existing discourse will take their text into account and incorporate it. The latter case could shed some light on the role of social movements in the development of a novel form. Social movements provide the background, broader discourse where a novel form can place its own discourse.

Our approach is going to follow a track which proceeds from an in-depth analysis of processes of creation of a “word” (the name of an organizational form, be it a Samba school, a sex shop or a mafia clan), representing something socially different according to the use we make of it. The word in itself, as we will see, will provide some grounding for defining how different people will experience its use in reality.

## **GRAY AND BLACK MARKETS: A BORDERLINE CASE FOR LEGITIMACY**

A “gray market” can be defined as the marketing of *unmentionables*, i.e. all those products and services that are difficult to publicize due to the delicacy of the involved subjects: pornography, funeral rites, etc. They include a variety of objects, situations, and concepts whose value and significance are embedded both in time and space. The same object can therefore acquire a different meaning in different contexts and in the same context at different times. It is for this reason that it becomes hard to classify objects, situations, and concepts as *unmentionables*.

Wilson and West (1988) distinguish between two macro-categories. The first category includes everything not accepted by the morals of a society, but tolerated (pornography, prostitution, etc.). The second category includes all those legitimate products and services that the society names with reluctance (funeral products, sanitary products, etc.). Obviously, as a consequence of their relative nature, what belongs to one or the other





*Fig. 1.* Representation of Generic Gray Market. *Note:* In relation to the classification of Wilson and West (1988), the left part of the continuum are products of the first category, i.e. the *unmentionables* defined in this way because of customer decency, but legitimate. The right part shows classified products belonging to the second category. In relation to the “*continuum*”, the Italian pornography, e.g., will be classified in the middle point with a slow, but natural, movement toward the white market.

category in one specific country may not be included in any of the two in another country.

Beyond the macro-division proposed by the two authors, further consideration defines the gray market as a *continuum* between the extremes of the “white market”, the one of pure legality and legitimacy, and the “black market”, the one that falls completely outside the scope of the law. Everything in the middle is just a dynamic shade of the two ends (Fig. 1).

“Who” and “what” determine the dynamic from one side to the other of the continuum? We can attribute this dynamic to the process of legitimation, or structuration (Powell & DiMaggio, 1991), which may derive from the provision of normative, regulative, or cultural-cognitive type. But this is not enough for our scope.

In other words, what determines this dynamic? What are the sources that legitimate the unmentionables market and the organizational forms that populate it? A journey through the evolution of three populations that move along this continuum is helpful to analyze how legitimation is a complex process.

#### *A Black Market and Legitimation: The Diffusion of Mafia in Milan*

The evolution and diffusion of mafia organizations represents an interesting setting for studying processes of legitimation of illegal social movements. When we use the term “mafia”, we refer to a formally organized criminal association where explicit roles and rules are prescribed, followed, and enforced, even through extreme brutality and violence (Falcone, 1991). Strictly speaking, what differentiates the mafia from other forms of criminal activity is the existence of an internal organization ranging in different criminal businesses (extortion, drug smuggling, and dealing, etc.) and extending to legitimate businesses and activities where illegal revenues come to be

invested. According to Masciandaro (1993), wealth acquisition is the principal goal – if not the unique – that explains the strategic and tactical decisions of mafia organizations. A finality that confirms the statement by Nichi Vendola, former member of the Antimafia Parliamentary Commission, is that mafia members might well be individuals who appear to be harmless. Indeed, it should be mentioned that violence is not the final purpose of mafia, but a means as any other to reach the real objective. A central issue of the political debate over mafia is its stronghold in the regions where it has been dominating under different names, but similar organization (‘ndrangheta in Calabria, Camorra in Campania, Sacra Corona Unita in Puglia, and Cosa Nostra or Stidda in Sicily). Mafia appears to be highly legitimated in those regions as a consequence of historical processes dating back to ancient times (there are some sources that date it back to the thirteenth century).

A striking feature of mafia in its different forms is the importance of myths and rituals that have two functions. On one side, they identify the choice of a new member and the crossing of a boundary. On the other, they define the difference between members and non-members. Gambetta (1994) provides an extraordinary account of the origins and processes of the traditional mafia in Sicily (Cosa Nostra). In his account, the role of protection, the “industry of protection” as he calls it, is a key aspect of the evolution of Cosa Nostra. In his account, the legitimation of Cosa Nostra in its original cradle (Sicily) is based upon its ability to interact with the larger society and exploit four resources: information, violence, reputation, and publicity. Service of protection are greatly enhanced by the perception of the needs of the protected (information), the availability of violence to restore order, a good reputation that decreases the need for violent actions, and therefore the costs of protection, and a network to sell protection (publicity).

The origins of mafia organizations and the way they acquired legitimacy are at the roots of the work of several important scholars (Arlacchi, 1983; Falcone, 1991; Gambetta, 1994). However, in addition to its traditional strongholds, the mafia has been extending its reach in wealthier Northern Italy regions, starting from Milan, the financial and economic center of Italy. How do criminal organizations gain legitimacy in such novel contexts?

Using the categorizations identified by Suchman (1995), mafia organizations gain pragmatic, moral, and cognitive legitimacy. They gain pragmatic legitimacy from audiences that directly benefit from their operations, like members and intermediaries. They gain moral legitimacy because their operations are characterized by strict rules and generally reduce uncertainty for citizens, albeit at the cost of freedom and bribery. Finally, they gain cognitive legitimacy because they are legitimated by their apparent

persistence through time and changes in regimes and because a collective discourse, animated by their very enemies, keeps the attention on them. As it can be perceived, legitimacy for mafia organizations can take very different flavors according to the specific audiences we consider.

Thanks to data collected by Direzione Distrettuale Antimafia (DDA) of Milan and by Omicron (Milan Observatory on Northern Italy organized criminality), integrated with those derived from articles appearing in major newspapers, my analysis focuses on the evolution of Milan mafia activities between 1980 and 1996.

Beyond the existing criminal activity of local criminal gangs (specialized in bank robbery), the beginning of the 1970s in Milan is characterized by a rapid rise in the consumption of drugs. Exploiting this opportunity, mafia clans decided to join drug trade exactly at that time. In the case of mafia organization, a clan can be considered operative when the art. 416 *bis*<sup>1</sup> can be applied and when it can be recognized as a “continuous and effective” violent activity.

Mafia organizations appeared in Milan through different paths. Some grew up from the association of mafiosi that operated in Milan as individual criminals and decided only at the end to build organized structures. Some others represent the result of merger processes between different pre-existing entities unrelated to other mafia organizations (e.g. a clan initiated to operate as agency quite independent of a Sicilian organization). Eventually, a third path derived from the intervention of police and law enforcement agencies that, impacting only on parts of these organizations, had the effect of changing the balance of existing powers and favoring the emergence of new clans. Taken together these processes lead to the development of an intricate network of mafia clans as depicted in Fig. 2. The dynamics of mafia organization follow a common path. First, a new group is established and it creates its order through the use of gunfire. Eventually, a war between clans might burst out, but this happened only once in Milan and was linked to problems between associated clans in Sicily. More often, when a new group emerges, usually composed of younger and more aggressive members, the older group adapts and limits itself to a niche of the market that in Milan is mainly linked to drug dealing. Instead of competing, clans in Milan tend to cooperate. The reason is linked to the great availability of resources deriving from the market for drugs and narcotics. Most of these processes occur without a clear perception by citizens of the emergence of organized criminal activity. In fact, the control of commercial activities (*pizzo*<sup>2</sup>) is not diffused as in Southern Italy. Most revenues for mafia organizations originate from other markets and this allows the city to perceive itself as virtually devoid of mafia.

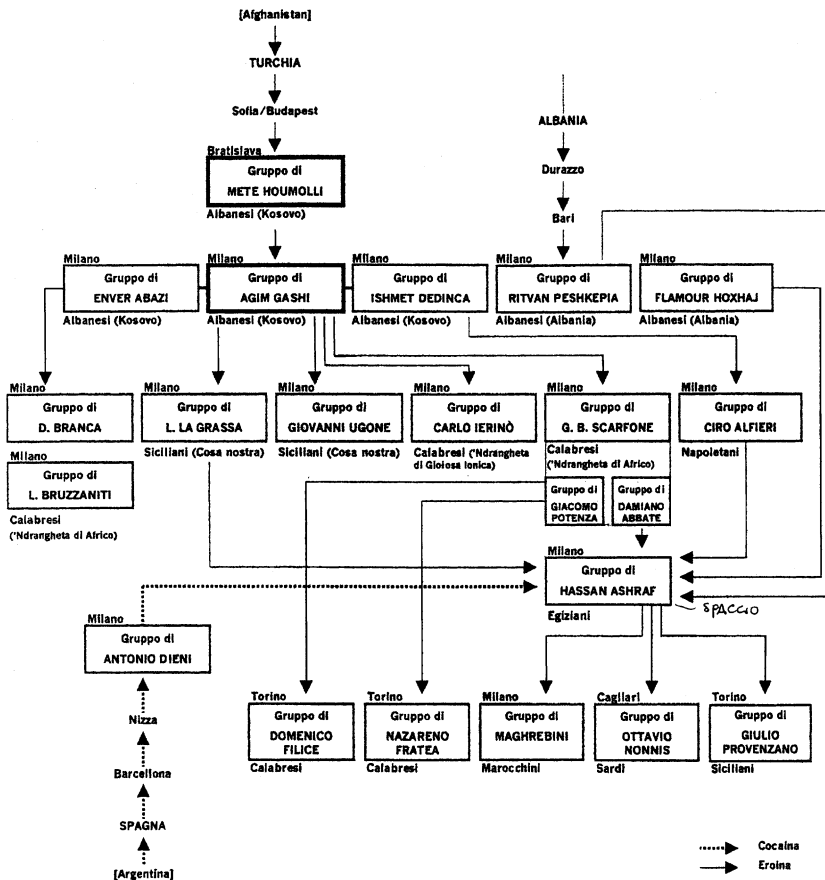


Fig. 2. A Criminal Network in Milano (Omicron, 14).

I collected data, thanks to the help of a famous prosecutor, Antonio Spadaro, who agreed to reconstruct the history of mafia clans in Milan using data from his past investigations. With his help, I was able to create a longitudinal dataset of founding events, which is the basis for the Poisson regression presented in Table 1.

The first model considers the existence of a relation between birth rates and density in the population of criminal mafia organizations and the results show a positive relation. The second model introduces the squared density and, quite surprisingly, the signs appear reverted. The first-order term shows

**Table 1.** Poisson Estimates of Mafia Organization Foundings in Milan, 1980–1996.

Independent Variable	Model			
	1	2	3	4
Density	0.156 <sup>a</sup> (0.049)	-0.573 <sup>a</sup> (0.153)	-1.210a (0.411)	-1.301 <sup>a</sup> (0.471)
Density squared		0.071 <sup>a</sup> (0.019)	0.131 <sup>a</sup> (0.0449)	0.135 <sup>b</sup> (0.048)
People arrested for drug dealing/1,000			0.725 <sup>b</sup> (0.267)	0.603 (0.393)
GDP in Lombardia/1,000				0.022 <sup>b</sup> (0.010)
$\chi^2$	61.043	120.630	147.233	156.339
Degrees of freedom	1	2	3	4

<sup>a</sup> $p < 0.01$ ;

<sup>b</sup> $p < 0.05$ .

a negative sign, while the second-order term has a positive sign. Other models add some covariates.

Apparently, the early dynamics of this population are not consistent with previous research in organizational ecology. A possible explanation considers that mafia organizations are secret and criminal, and they “continually and efficiently” produce the same output: *violence*.<sup>3</sup> Mafia organizations, at the end, get the upper hand with violence and illegality driven by the search for a monopoly in the activity (legal or illegal) of interest. Thanks to these considerations, it is possible to recognize other processes, in addition to legitimation and competition, sustain mafia activities.

Being organizations that try to impose their supremacy with violence, they will face opposition and resistance (which could be considered as competition with common criminals) to their actions and, consequently, to their development and growth. Talking about opposition and resistance, it is necessary to distinguish between situations characterized by a low and a high level of density. We expect that population density would be particularly high in areas where mafia finds a high level of cognitive legitimacy. In other words, we consider that the opposition/resistance effect would disappear when there is a high density.

When organizational density is low, it is difficult to consider the mafia as invincible and its actions as unavoidable, as it happens in the traditional areas. It is therefore reasonable to have a strong opposition from

organizations that could be impacted by mafia violence (in case of Milan, other criminal endeavors). In addition to this, low density often is not a “natural data” point, but the fruit of repressive action by law enforcement agencies.

In other words, at the beginning an illegal activity faces an increasing opposition until it gains a certain legitimacy. After mafia organizations gained legitimacy, the relationship between birth rate and density switches. In sum, at a low level of density, the processes of opposition and repression that determine a negative relationship between birth rate and density are more significant; at a high level of density, the processes of legitimation assume more importance.

These results could be explained by the characteristics of the city of Milan. Indeed, we can consider Milan as a young area in the field of mafia organizations, especially if we compare Milan with other areas of Italy – e.g. Sicilia, Campania, and Calabria – where there are such institutionalized mafia organizations as *Cosa Nostra*, *Camorra*, and *‘Ndrangheta*. So, given these characteristics, there are enough resource to avoid conflicts between different clans, i.e. the environmental carrying capacity is still to be reached. In the future, we could observe more and more conflicts if the number of illegal organizations reached the threshold of *carrying capacity*. However, we have to make clearer the point that mafia organizations are not always in the same niche. On the contrary, they operate in different ways, in different places, in different niches but always using the same instrument, violence. Moreover, as we have already said, high levels of density could determine mutualistic relations rather than competition between clans. The increase in the number of mafia organization, for example, could lead to an increase in the number of illegal goods as well as to the contamination of the “white market”.

Legitimacy for illegal activities shows a complex pattern of relations with taken-for-grantedness and, apparently, density dynamics reflect this in very peculiar ways. The interplay of audiences, which either strongly support, or strongly oppose, or passively assist the diffusion of mafia organizations, renders highly improbable that legitimacy could be determined by increases in density.

#### *Gray Market and Legitimacy: The Evolution of Sex Shops in Milan*

The Italian debate on the issue of sexual morality has become more and more lively since the end of the nineteenth and beginning of the twentieth century, roughly a century later than other countries in Europe. The path to

the legitimacy of sexual issues has been (and still is) much harder than what happened in other countries, at least for two reasons. Beside the involvement of lay moralists, there is the important role of the Catholic Church and a religious morality that are a fundamental and unavoidable part of Italian culture and that have never been neutral on these topics.

From a strictly regulative point of view, Italian law simultaneously allows and inhibits any activity linked to sexual topics (see artt. 21 and 33, Cost.; artt. 527 and 528 cod. pen.). In order to be compliant with law, activities linked with sex have to be in line with public morality, “decent and civil habits”,<sup>4</sup> and “common decency”.<sup>5</sup> In other words, *common sense* does not refer to just a pre-defined *quid*, or to single offended individuals or to the judges that have to apply the law, but to the mean sentiment of the collectivity of *that* population in *that* specific period (Cass. 07/03/1953). In these terms, the pornographic market is therefore positioned between tolerance and legitimation (surely not legality).

Within this shifting regulative context, the Italian market for pornography has been developing since the 1970s. It is in 1972 when the first so-called “temple of perversion”, the first sexy-shop opened in Milan. What are sexy-shops? Interestingly, it is not a typo-error, because it is “sexy-shop” and not sex shop as an English-speaking person might imagine. The name will remain unchanged until today, although people might not realize the semantic difference.

A sexy-shop is a place where pornographic material is sold (videos, magazines, sexual instruments, etc.). The most diffused form in Milan is the one of limited company (ltd), with regular sale license and emission of fiscal receipts. Despite compliance with certain legal, administrative, and sale requirements, it remains difficult to find with precision where these temples are located. As commercial shops they are registered at the Chamber of Commerce, but they are not codified according to the real market activity they bring on. This data gap dovetails with what has already been mentioned because codification would imply legitimation and credit to an activity that is close to illegality.<sup>6</sup> For this reason, the analysis of the sexy-shop in the Milan area meant real “field research”<sup>7</sup> into the sale and rental of videos, lingerie, and even orthopedic objects!

After defining the population (Milan sexy-shops) and the time period (1972–2000), I decided to focus on birth rates of sexy-shops. Attention has therefore been given to the detection of the initial entrepreneurial event, i.e. the date of start-up after the regularization. After clarifying these aspects of the research design, I proceeded with the observation of the population, highlighting in particular the number of births per year and the density.

The first “official” sexy-shop in Milan starts in 1972. Since then the number of sexy-shops increased to 72 establishments in 2000 that were dedicated to the sale of unmentionable products (pornographic videos, “particular” objects, lingerie, and sexy clothing).

Even if the first start-up was in the 1970s, the births had a “singular” trend. As evident from Fig. 3, only at the end of the 1980s and during the 1990s it is possible to observe sustained growth: 83% of births take stage in the 1990s. Compared with the other European countries, these birth rates are quite low, if we exclude the exceptional figure of 1999, with 22 start-ups.

In Fig. 4, it is possible to observe the trend of density per year. The graph displays a clear increasing path, but with neither a linear nor an exponential trend: this is due to the fact that births have no increasing trend, but are

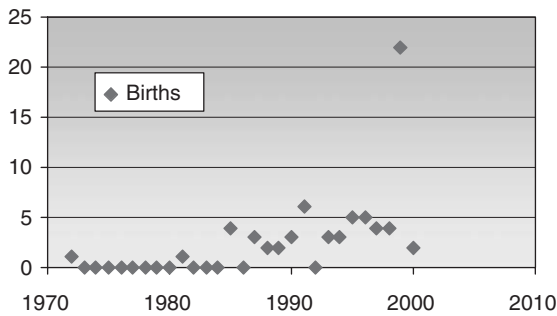


Fig. 3. Sexy-Shops Births Per Year in Milan. Note: Eighty three percent of the births occurred during the 1990s. Until 1985, Just two sexy-shops were present; since then, the growth has been more consistent, with at least a start-up per year. The record year is 1999 with a huge number (22) of births.

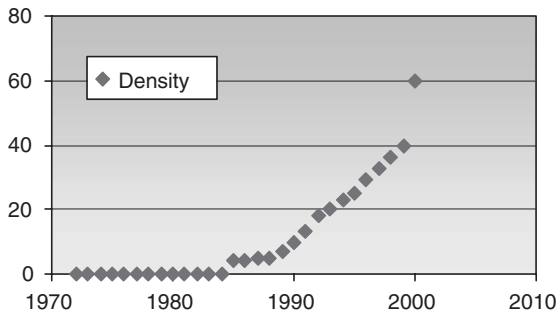


Fig. 4. Trend of Sexy-Shop Density Per Year.



quite flat (Fig. 3). Even in this case we observe that in 1999 something happened: something that makes density explode.

The total number of collected observations is 72 sexy-shop start-ups between 1972 and 2000. I now present a Poisson model estimating birth rates among these establishments. Alongside density dependence, I consider a number of other covariates. First, I target events that might emphasize the visibility of pornography and provide media coverage:

1. The election of Cicciolina (a famous porno-star) to the Parliament in 1987 (PAR).
2. The death of Moana Pozzi (the most famous and iconic Italian porno-star) in 1994 (DMP).
3. The first public offering of a German sexy-shop chain on a stock exchange in 1999 (QUOT).

Second, according to the opinion of an expert in this industry, “the exemplary customer is not the classical employee who comes home full of desires and takes the wife into fantastic erotic games” (Michele Baraldi).<sup>8</sup> The typical consumer is a professional with a level of high education, who travels around Europe (for work) and has had the opportunity to know foreign pornography markets. So, the higher the education level, the greater will be the mental openness of the population and consequently the acceptance of the existence of sexy-shop and therefore the increase of their presence.

Third, the twentieth century has been marked by widespread processes of social change in the way relationships and families are interpreted: from a greater emancipation of women to an end of arranged marriages. It might be possible that the decline in traditional morals and values (marriage, family, love, and religion) has a positive impact on sexy-shop birth rate, i.e. an increase in the decline of traditional values increases the number of new sexy-shops.

As a good proxy of this crisis of moral values, I choose the number of marriages (civil and religious) and considered separately religious marriages. The results presented in Table 2 illustrate a complex pattern of relationships. Density dependence appears to have only positive effects on birth rates – along with education levels and the initial public offering of a German sex shop chain in 1999 – while the election of a porno-star in the Italian Parliament appears to have a negative impact. Obviously, the model shown has only indicative value, but its direction appears consistent with our argument.

It may be premature to attempt to interpret these results, but interestingly it appears as if the diffusion of sex shops is following a different fate if considered from the standpoint of entrepreneurs rather than the general audience. The positive impact of density might reveal a growing availability

**Table 2.** Poisson Estimates of Sexy-Shop Foundings in Milan.

Independent Variable	
Constant	2.174 <sup>a</sup> (0.507)
Density	0.591 <sup>a</sup> (0.076)
Density squared	-0.153 (0.895)
PAR	-2.072 <sup>b</sup> (0.851)
DMP	-0.257 (0.326)
QUOT	20.753 <sup>a</sup> (0.575)
Percentage of illiterates	-1.092 <sup>c</sup> (0.479)
Percentage of graduates	1.087 <sup>c</sup> (0.478)
Number of marriages	0.250 (0.472)
Number of religious marriages	-0.001 (0.597)
Number of observation	72
R <sup>2</sup>	0.986
Log-likelihood	-141.236
Degrees of freedom	9

<sup>a</sup> $p < 0.01$ ;

<sup>b</sup> $p < 0.05$ ;

<sup>c</sup> $p < 0.1$ .

of a prototype for entrepreneurs and signal a pragmatic legitimacy (Suchman, 1995). The positive impact of education might reveal a pattern of moral legitimacy or cognitive legitimacy resting on comprehensibility. Finally, the negative effect of the election of the famous porno-star Cicciolina might signal that the general audience would react differently to the attempt at taken-for-grantedness implied by this situation.

*Moving from Black to White through Gray: Samba Schools in Rio*

No doubt one of the most striking popular forms of celebration is the Carnival that is deeply rooted in ancestral myths and has survived, albeit in different forms, the process of modernization of our societies. When it

comes to the Carnival, there are several key sites that come to mind, like Venice in Italy, New Orleans in the US and, of course, Rio de Janeiro in Brazil. As most forms of ritual, the Carnival has undergone quite a lot of investigation by anthropologists, but scant attention has been devoted to the way it became a subject for organizing. Most Carnivals operate thanks to a semi-formal organizing structure that is involved in parallel activities throughout the year, but there is a notable exception, the role of Samba schools in the Carnival in Rio.

The origins of the Carnival in Rio are deeply intertwined with the evolution of samba as a music genre, an evolution that has been the subject of a sharp confrontation between legitimated authority and popular will. There are a number of hypotheses on the origins of the word in itself, but it is rather clear that it first appeared linked to a musical genre on February 3, 1838 when Friar Miguel do Sacramento Lopes Gama used it to define a musical genre popular in rural and peripheral areas in Brazil. Actually the word “samba” referred to a set of different musical genres generally imported from Africa by slaves and rooted in different geographical areas in the Eastern part of Brazil (these genres included, for instance, tambor de mina do Maranhão, Milindo in Piauí, Bambolo in Rio Grande do Norte, etc.).

At that time, Rio de Janeiro played a great role by being the melting pot of a radically heterogeneous set of influences brought by free and forced immigrants. According to Moura (1983), who provides one of the most detailed accounts of music in Rio, the early samba was influenced in particular by other genres, like modinha, choro, and lundu.<sup>9</sup> Although samba reveals an ancient origin, the way it became a key ingredient of Carnival in Rio followed a path full of conflicts and social tensions.

Early forms of Carnival had been introduced since the sixteenth century by colonial Portugal in the form of the *entrudo*, rooted in India but then subjected to a process of anchoring and objectification (Palmonari, Cavazza, & Rubini, 2002; Moscovici, 1961) alongside Christian principles and values. In the early stages, *entrudo* was an organized rite before the start of Lent, where participants would battle with flour, water, and *limões de cheiro*.<sup>10</sup> It clearly was a form of entertainment limited to dominant classes. *Entrudo* did not entail music and dance and remained an institution in Brazil until the second half of the nineteenth century, when repression under the Emperor took its more brutal form (people suspected of participating in *entrudo* could be sentenced to death). The modern Carnival in Rio emerged on the ashes of *entrudo*, after an explicit ban between the seventeenth and twentieth centuries by legitimate authorities. One of the motivations for this ban was the goal of modernizing the country and abandoning old and

traditional institutions that were not consistent with the desired image and identity of Brazil.

The Carnival in Rio struggled to acquire a definite form as different alternatives competed for dominance. It was in 1846 that a new form of Carnival emerged. Zé Pereira was born when people started strolling around beating drums and visiting newspapers to acquire visibility. The Zé Pereira had origins in the Portuguese culture and gave birth to new songs on a drum basis, still without any link to dance. It was an entertainment mainly followed by poorer people. White, dominating elites, in fact, resolved to import the European-style Carnival with dance-floors and favored the creation of “Grandes Sociedades” (literally, large Carnival societies), like the “Sumidades carnelescas” that was the first to arrange a public catwalk in 1855 using masks, flowers, and allegoric floats or carts. The Grandes Sociedades explicitly denied the value of entrudo and Zé Pererira and advocated a modern form of entertainment where luxury and European flavor should dominate (one of them was named União Veneziana after Venice in Italy). Their themes had political and social targets both inside Brazil and internationally. In 1907, they would start their catwalk with motorized vehicles on Avenida Rio Branco, giving birth to what would become the “corso”. Among the Grandes Sociedades, three would survive until 1940: Tenenti do Diabolo (born in 1867), Democráticos (created in 1867 and devoted to the abolition of slavery), and Fenianos (1869, of Irish origin).

Other forms of collective celebration were still in place. Catholic processions had been a constant if we consider that von Martius in 1818 ([von Martius & von Spix, 1823–1832](#)) had calculated 35 religious festivities apart from Sundays where slaves were allowed to dance, sing, and play music! The African traditions were entangled with Catholic traditions and originated Congos and Congadas, ritual dances where the rei do Congo (Mani Congo, an ally to Portugueses) would struggle with a pagan king. According to [Efege \(1965\)](#), Ze Pereira evolved into cordões where alongside drummers, Indians and other roles would become institutionalized. In cordões Indians came first acting as the “Ameríndios”, preceding the pano<sup>11</sup> and a second group of the “Re do Diabolo”. Each participant had a role and a costume and the group would play a scenic, symbolic struggle. They started being performed during the Feast of Our Lady of the Rosary and assumed different forms, like the cordão de Velhos or the cucumbis. The cordões would end in newspapers’ and journals’ offices, where the banner would be left for months. This allowed the cordões to become popular and attract newspaper coverage, starting in 1886, in turn helping their diffusion among middle-upper-class citizens. The different cordões would compete for

funeral wreaths promised by Funeral House owners passing by their sites. Later they would all rejoin in Praça Onze. The cordões enjoyed immediate success, but were later banished from time to time due to the violence they aroused between participants. In 1911, journals decided to abandon cordões and started promoting a different form, Os ranchos carnavalescos, and many former participants in cordões decided to join them. The important role of newspapers in defining the success of alternative forms of Carnival thus continued.

Ranchos evolved to become the most immediate ancestor of Samba schools, though their tradition is rooted in religious and temporal (Dias do Reis – Days of the King) ceremonies. They abandoned the attitude toward confrontation of former forms and embraced irony and subtlety. The first rancho (Dois do Ouro) can be dated back to 1872 and already showed a structure similar to the *afoxê* typical of Bahia: *obà* (the king), *ibiquejiobà* (vice-king), *baizo-orum* (counsellor), and so forth. The rancho in Rio de Janeiro was moved to Carnival instead of being held on January 6th, the Epiphany, because of an explicit prohibition. The ranchos differentiated themselves from the cordões, attempting to becoming legitimated among the police through the mediation of journalists. In 1894, the *Rei do Ouro*, a classical figure of the ranchos, was received at Itamaray Palace by the President of the Brazilian Republic, Marshall Floriano. Ranchos and cordões were later to be prosecuted by the legitimate authority but were able to survive. Actually, cordões became ranchos or sometimes *bloco* and it is interesting to note that on December 31, 1918 the *Cordão de Bola Preta* was founded. It never really acted as a cordão, but it has kept in its statute the goal of preserving the memory of cordões until today.

The early twentieth century brought moral and legal opposition against these forms of collective entertainment. The Catholic Church criticized the Feast of the Penha, while the municipality and police prohibited samba, arresting those who were found playing it. It was only in 1917 that samba acquired its legitimacy with the song *Pelo Telefone* that was used by the *Democráticos* in their catwalk. Its genre, though, was going to change soon when in 1930 a group of young African musicians (*Estaciò de Sá*) proposed a new rhythm that allowed a more frenetic dance.

Ranchos rapidly acquired a widespread legitimacy and were soon invited to the parade with the *Grandes Sociedades*. They could be easily distinguished because of the strict consistency of their themes and colors and their well-defined internal structure. Their dominance continued until the crisis that took place in the 1940s when solidarity and economic resources were lacking.

In 1916, the *Jornal do Brasil* announced the presence of a “bloco di sujeitos”<sup>12</sup> – dressed up like monkeys that would amuse people with tricks and games – at the Feast of the Penha (a church where a miracle had been said to have happened in 1635). It must be noted that the Feast of the Penha had an important role because it allowed people from very different classes to interact. It was from this feast though that another form took its lights, the bloco. They were more informal than the other forms, but it was there that samba started to accompany the parades. Different blocos emerged, some of them with a unified structure, others completely informal. Again it was newspapers that established the first competitions for blocos, which would make their parade the Sunday before Carnival. Blocos continued to operate in parallel with other forms until 1965 when the few remaining created an association. In the 1970s, blocos were associated with Samba schools and this lasted until 1987 when they were finally banned from the Samba schools parade.

On August 12, 1928, *Deixa Falar*, the first Samba school, was born. The idea had been to transform a bloco into something similar to a rancho in order to obtain a permit from the police. It was therefore necessary to define a location for the school and since this location was close to a real school, it appeared natural to define *Deixa Falar* as a Samba school. Strange as it might seem, *Deixa Falar* was never really a Samba school in the true sense of the term, but rather something in between a rancho and a well-organized bloco. Its destiny was obscure, because in 1933 it merged with another bloco and disappeared, ironically in the same year when the first competition between Samba schools took place. Samba schools emerged rapidly between 1928 and 1932 following a well-defined internal structure: modern samba music, a samba dance, a cortege of people dancing samba, a set of specific drums, and a wind of *baiane* (young dancers). Other elements were imported from ranchos: *enredo* (theme), *mestre de sala* and *porta-bandeira* (carrying the banner and dancing), *alegorias* (painted banners carried on carts), and *comissão de frente* (with the same dress, usually important persons or older members). Samba schools were founded in different parts of Rio de Janeiro, mainly in the poorer districts, where people lived in favelas. Almost independently, several blocos in different parts of the city evolved, following the rules defined in a famous meeting at *Zé Espinguela*’s house (Candomblé priest and polygamous founder of Samba School *Mangureira*).

The first competition among Samba schools was organized in 1932 by *Mario Filho*, a journalist at *Mundo Sportivo*. Soon to follow was the involvement of another journal, *Globo*. Actually, Samba schools were radically different from previous forms, because they entailed a goal of social

inclusion. In 1934, the schools founded UES (União das Escolas de Samba), their association that was able to negotiate an autonomous role in the Carnival with the local authority. The evolution in time of Samba schools was characterized by a strong presence of the State that granted funding but exerted control. Today, Samba schools have acquired a well-defined structure and a clear economic stance. Revenues from the Carnival are distributed among them. There are different groups of Samba schools that compete. Every year the winners get promoted and losers are demoted to a lower level competition. The internal structure is formalized into two parts: a business-like organization and the Carnival organization. Roles are defined and structured alike in different schools. They dominate the Carnival in Rio and have progressively absorbed the previous forms, like ranchos, blocos, grandes sociedades, incorporating some of their features.

It is interesting at this stage to take a closer look at what this process of legitimation can tell us. Clearly, the whole history that traces back to *entredo* is the only way to explain why and how Samba schools came to operate a crucial role in the Carnival in Rio. The dynamics of the multicultural society of this city and the evolution toward modernity all provided essential ingredients for this history.

The legitimation of Samba schools was not an easy process, opposed as they were by state institutions and previous alternative forms. In order to capture legitimacy, early Samba schools appealed to continuity with earlier forms and benefited from the actions of very specific individuals, like journalists, *candomblé* priest, or *tia maria* (iconic figures around whom *sambistas* would gather). There was not a prototype which sparked their diffusion, because the first organization to use the name was radically different in nature and form, more closely resembling a *bloco* with ambitions to become a *rancho*. It also seems strange that a name was created before a new form was born. Once the name was in place, specific entities (organizations) arose to fill the niche and craft its boundaries to self-propel the expansion. Density did not seem to play any role, because Samba schools were founded by mutating previous existing informal or semi-formal organizations. It was not the number of these schools that prompted Mario Filho to propose a competition and journalists were rather skeptical when he published his advertisement. People outside the *morros* (the areas where poorest people lived) had no idea what a Samba school was and few among them could imagine that 19 of them would participate in 1932.

Audiences did play a role if we consider how homogeneous the early founders of Samba schools were, usually originating from the poorest components of the Rio society, but the interplay between audiences and the

novel organizational form is not straightforward. While audiences saw Samba schools as a way to exit marginality, Samba schools were a legitimated and controlled instrument for voicing other issues that were at the roots of processes of social exclusion. I think that this case closely resembles the findings by Greve and Rao (2006) on Norwegian early insurance companies. Legitimation is acquired and density increases by virtue of initial imprinting rather than the contrary.

*Three Populations, a Single History*

The stories of these three populations differ in timing and context, but appear similar in detailing the complexity of the relations involved in the process of legitimation of an organizational form (Table 3).

**Table 3.** A Comparison of the Three Cases.

	Mafia	Sex Shops	Samba Schools
Endorsing actors	Criminals Former members of original clans	Entrepreneurs Interested customers	Lower classes Brazilians of African origin
Opposing actors	Police Prosecutors	Liberals Conservatives Local communities Police Prosecutors	Police Prosecutors Elites
Pre-form	Local branches/ members of traditional mafia	Illegal commerce of materials and videos	Bloco
Density dependence	Inverted signs	Single density dependence	Rancho n/a
Legitimation	Opposed in the early stages, but once they emerge they create an action set that favors further rapid diffusion among criminals, thanks to “invisibility” to larger public	Density dependent through taken-for-grantedness by endorsing actors, but competition effect is triggered by excessive diffusion and counterbalancing pressures by social movements	Almost instantaneous after the name was created, but required a long and conflicted path to legitimacy



The three cases illustrate the changing patterns of legitimation when the creation of a new form interacts with different audiences. An important role is played by what I call “pre-form” and the nexus of meaning relations that this entity initiates. Legitimation appears to start operating well before a new form is established or even defined by a specific word/name. It is at this level that the social processes between audiences and ideologies define the path toward legitimacy. Obviously, it is important to notice that if we consider a new form that is deeply linked to an implicit exemplar (i.e. a legitimate business activity), we might not find evidence of this and empirically reveal a link with density and taken-for-grantedness through diffusion.

## DISCUSSION AND CONCLUSION

This paper has attempted to provide some hints as to how theorizing about the legitimacy of novel organizational forms should be improved before moving to empirical research. In doing so, I mainly used results from previous research projects that dealt with borderline organizations to illustrate important deviations from mainstream predictions. I believe that these cases are thought-provoking and pinpoint exactly those areas where a more careful analysis of existing theories of social categorization and social representations would greatly help craft better theories and better accounts of empirical findings.

The three cases I described all point to the necessity of rethinking existing approaches to legitimacy, considering organizational forms as social constructions which co-evolve with heterogeneous audiences and assume that meaning cannot be considered homogeneous. The fate of a novel organizational form does not proceed through a linear process by which codes emerge and are enforced as default simply because audiences see objective things through different lenses and processes. The mafia organizations enjoy diffusion dynamics where legitimacy increases only after a minimum density is obtained due to interaction effects and competition at community level. Sexy-shops legitimate themselves only in the eyes of a minority audience but fail to develop further, not only because the material resources (demand and customers) are limited, but because their diffusion prompts reaction by other audiences. Still they can benefit from legitimacy gained from sex shops elsewhere in Europe, a reason for the curious name they ended up choosing in Italy. Finally, the Samba schools provide an interesting case where legitimation does not proceed through conformity to previous audience expectations, nor selection of a deliberate audience, nor creation of new

audiences and new beliefs (Suchman, 1995), but through an intersubjective process by which audience and organizational form co-evolve. There is no reason to search for a set of causal actions because they are a product of collective sensemaking that gives birth to an organization that, while forming, creates its audience; none of these events comes first, they are a collective act of social creation. Under this perspective, both the account of legitimacy provided by Suchman (1995) and a *fortiori* those in mainstream organization ecology fail to properly explain the emergence of Samba schools. Both approaches can find substantive empirical results, but fail to analyze the real causal relation: it is because the form is already there that the processes they suggest operate, but it is not the processes that *explain why* the form is there.

Moreover, apart from suggesting a different approach to the analysis of legitimation, these results point to the fact that there can be different interpretations of legitimacy operating at the same time for the same population. It is possible not to find a perfect correspondence between what norms allow and what people take for granted. In the sexy-shop market this is particularly evident, as the rules regarding the offense against decency and sexual honor still exist and are actually the same as in the past while new forms are diffusing thanks to different interpretation of those same norms and regulations. The fact that legitimacy is a process deeply linked to audiences and individuals suggests that we might think of extending the ecology of organizations to an ecology of audiences. The fate of a specific population would then be deeply linked to the values, norms, and behaviors of audiences. Interestingly, this is a new direction toward exploring the space of resources where populations operate. We might identify interesting relations between organizations that do not seem to operate in similar niches. Fashion and identity movements favor at the same time different kinds of organizations that share some ideological characteristics. Think for example at the diffusion of healthy habits as the creation of an audience that would favor the diffusion of organizations such as organic growers, sustainable producers, new political movements, and so forth. Researchers could move to analyzing a sort of ecology of communities where different organizations operating in different niches compete as part of an ideological community. Past research on the institutional competition between state and kibbutz, or between kibbutz and moshav by Simons and Ingram (2003, 2004) provides examples of such processes. Barnett and Woywode (2004) explore the ecological dynamics of ideology through the related dynamics of newspapers in Vienna. Their findings suggest that patterns of competition and legitimacy might vary when organizations are deeply rooted in

ideologies. Instead of taking place within the same ideological niche, competition is greater between neighboring ideologies with right-wing newspapers linked by a predator–prey relation with center and bourgeois newspapers. However, the richness and complexity of political processes at the time studied by the authors seems not to be well represented by the dynamics analyzed. Does competition between ideologically different newspapers reflect competition between ideologies, and if so, is it appropriate to capture it through density of different newspapers? My perception is that this is a case where employing a causal framework is misleading because cause and effect are intersubjectively determined.

Another problem arises if we consider the contribution of theory and research in categorization processes. Social realists are moving into a field where social construction defines the realm for research, and it is hard to find a common ground unless we move from objects to words and language, following the direction declared by Wittgenstein (1967). Words are objects with social constructed meaning, and discourse is observable by social realists, still being debatable according to social construction approaches. A form, before being any specific real object, is a word in language that comes to be used; the name of a new business (bank, internet search motor, etc.) is again a word that comes to be created and diffused. In this same direction, Hannan (2005) proposed that scholars look into new directions by either exploring the idea of a multidimensional space of audience distinctions, or exploring the semantics and texts, which help audiences in defining that very space. His advice is apparently at the core of recent empirical research that sometimes departs from earlier formulation of legitimacy. McKendrick and Carroll (2001) investigate the origins of forms by addressing a “non-form”, i.e. disk arrays. Their contribution appears particularly interesting because they show how identity could act to prevent the emergence of a new form. In their research on the diffusion of nouvelle cuisine, Rao, Monin, and Durand (2003) highlight different factors, like the sociopolitical legitimacy of activists, the number of defectors, the gains for defectors, and the theorization of new roles. Rao et al. (2003, p. 836) “suggest that diffusion is not a mindless process of replication but a mindful process of identity construction”. Hsu and Hannan (2005) restate most of previous work on identities and forms, but eventually start accepting a different view of the process that leads to a form. This process “may also depend on the existence of a distinct label for the set of organizational actors” (p. 478).

I perceive these important contributions as a major advancement of theorization in organization theory. Legitimation and legitimacy are being analyzed without resorting to empirical simplifications. Language and

naming of social constructs becomes a key aspect of new forms creation as well as individual and collective actors that participate into this process. This contribution paves the way for a complete redefinition of organizational form and new empirical direction like those explored initially by Ruef (2000). In his contribution, Ruef (2000) proposed to analyze form creation within a community ecology approach, where the carrying capacity takes the form of a latent carrying capacity, one that pre-dates the emergence of a novel organizational form. In his research, he demonstrated how this latent carrying capacity could be derived from the existing discourse in the field through a complex analysis conducted on texts. More recently, Hsu and Podolny (2005) analyzed the schemas used by film critics to evaluate products and producers in that domain. Greve and Rao (2006) illustrate the importance of resonance of organizational identities whenever organizations draw support from socially segregated groups, as for restricted mutuals and village fire mutuals in Norway. They therefore propose to reverse the relation between identities and audiences, positing that in specific instances it is audiences who attach identity to organizational forms.

To conclude, the legitimacy of legitimacy in itself requires clarification because it is a concept that hypothesizes a certain implicit level where it appears in social behaviors and therefore is socially constructed as a theory. We observe legitimacy at a certain stage, because we define it as a process that operates at a certain stage. This emphasis seems misplaced. Legitimacy is either a complex, intertwined set of deliberate and casual actions and events (so akin to history) or a discrete act where social forces coalesce and produce innovation, but in either cases it is not a social condition.

## NOTES

1. This article affirms that “an association can be considered mafia when members engage in activities that use force and intimidation connected to membership duties and resolve to criminal behaviors (like murders, control of businesses and public resource, etc.)”.

2. Pizzo is the name used to describe activities devoted to obtaining the payment of money from legitimate businesses in order to be “protected”.

3. For mafia organizations, violence represents a sanctioning mechanism for those individuals, within *and* outside the organization, that do not respect rules. Moreover, as Arlacchi (1983) said, violence is also a mechanism to obtain competitive advantage.

4. The interpretation of the “decent and civil habits” concept has been deeply discussed. Commonly, it is intended as all the rules that try to consider public opinion and protect, in a given historical moment, not just decency and sexual decency, but also the common moral sentiment, the public morality and civil life.

5. It is clear that even if intended as pertinent to just the sexual topic, both “decent and civil habits” and “common decency” result conditioned in space and time. This relativism gives a reason for the law to fluctuate in its interpretation.

6. Indeed, the market for pornography and sex is still composed by a large part of revenues coming from the black market.

7. Through the detection of commercial denomination and addresses it has been possible to understand the birth dynamic of many sexy-shops in Milan from 1972 to 2000. A second source has been the reading of annual Yellow Pages, where the category of sexy-shops is not specified until 1998. A third source with sensible credit has been the use of ISPES reports on pornography. Finally, the last dataset is derived from research in the field, by which I also verified the effective presence of the declared sexy-shops. I therefore actually verified the existence with an “address-by-address” empirical method.

8. Marketing Director of Top Line Video, one of the bigger companies in the hard-core film distribution sector in Italy.

9. Modinha was one of the earliest Brazilian genres in the eighteenth century; choro was a particular arrangement of national and international songs; lundu was characterized by the presence of African rhythm and drums.

10. Limões de cheiro were lemons, filled with perfume and coated in wax that people would throw at each other.

11. The pano was the banner.

12. Group of masked individuals who would sing and dance.

## ACKNOWLEDGMENTS

I would like to thank Flavia Blumetti and Elena Perondi for their invaluable help in revisiting the three case studies presented. I am indebted to Stefano Mureddu, Tiziana Ferro, and Silvia Marchisio who did an extensive research in the field. Moreover, I would like to thank Martin Ruef because many of the ideas presented in this paper were first stimulated by his working paper that I discussed at the Second Strategic Management Conference at Stanford School of Business in 1998.

## REFERENCES

- Arcuri, L., & Castelli, L. (2004). *La cognizione sociale. Strutture e processi di rappresentazione*. Roma: Editori Laterza.
- Arlacchi, P. (1983). *La mafia imprenditrice. L'etica mafiosa e lo spirito del capitalismo*. Bologna: Il Mulino.
- Barnett, W. P., & Woywode, M. (2004). From Red Vienna to the Anschluss: ideological competition among Viennese newspapers during the rise of national socialism. *American Journal of Sociology*, 109(6), 1425–1499.

- Baum, J. A. C., & Powell, W. W. (1995). Cultivating an institutional ecology of organizations: Comment on Hannan, Carroll, Dundon, and Torres. *American Sociological Review*, 60(4), 529–538.
- Caronna, C. A. (2007). Turning identity into form: The cause and consequence for Kaiser Permanente of becoming an HMO. In: M. Ruef & M. Lounsbury (Eds), *The sociology of entrepreneurship*. Research in the Sociology of Organization.
- Carroll, G. R., & Hannan, M. T. (1989). On using institutional theory in studying organizational populations. *American Sociological Review*, 54(4), 545–548.
- Carroll, G. R., & Hannan, M. T. (2000). *The demography of corporations and industries*. Princeton, NJ: Princeton University Press.
- Dawkins, R. (1976). *The selfish gene*. Oxford: Oxford University Press.
- Efegê, J. (1965). *Ameno resedá: O rancho que foi escola*. Rio de Janeiro: Letras e Artes.
- Falcone, G. (1991). *Cose di cosa nostra*. Milano: Rizzoli.
- Gambetta, D. (1994). *La mafia siciliana. Un'industria della protezione privata*. Torino: Einaudi.
- Greve, H. R., & Rao, H. (2006). *Mutual consent: The demography of organizational identity in the insurance industry in Norway: 1820–1940*. Working Paper.
- Hannan, M. T. (2005). Ecologies of organizations: Diversity and identity. *Journal of Economic Perspectives*, 19(1), 51–70.
- Hannan, M. T., & Carroll, G. R. (1995). Theory building and cheap talking on legitimation: Reply to Baum and Powell. *American Sociological Review*, 60(4), 539–544.
- Hannan, M. T., Polos, L., & Carroll, G. R. (2004). *Toward a theory of forms: Similarity and categorization*. Working Paper.
- Hsu, G., & Hannan, M. T. (2005). Identities, genres, and organizational forms. *Organization Science*, 16(5), 474–490.
- Hsu, G., & Podolny, J. M. (2005). Critiquing the critics: A methodological approach for the comparative evaluation of critical schemas. *Sociological Science Research*, 34, 189–214.
- Linville, P. W., Fischer, G. W., & Salovey, P. (1989). Perceived distributions of the characteristics of in-group and out-group members: Empirical evidence and a computer simulation. *Journal of Personality and Social Psychology*, 57, 165–188.
- Lounsbury, M., & Rao, H. (2004). Sources of durability and change in market classifications: A study of the reconstitution of product categories in the American mutual fund industry, 1944–1985. *Social Forces*, 82(3), 969–999.
- Masciandaro, D. (1993). Analisi economica della criminalità, teoria della regolamentazione e riciclaggio finanziario. In: S. Zamagni (Ed.), *Mercati illegali e mafie: l'economia del crimine organizzato*. Bologna: Il Mulino.
- McKendrick, D. G., & Carroll, G. R. (2001). On the genesis of organizational forms: Evidence from the market for dick drive arrays. *Organization Science*, 12, 661–682.
- McMullen, J. S., & Shepherd, D. A. (2006). Entrepreneurial action and the role of uncertainty in the theory of the entrepreneur. *Academy of Management Review*, 31(1), 132–152.
- Moscovici, S. (1961). *La psychanalyse, son image et son publique*. Paris: PUF.
- Moura, R. (1983). *Tia Ciata e a Pequena Africa no Rio de Janeiro*. Rio de Janeiro: FUNARTE.
- Palmonari, A., Cavazza, N., & Rubini, M. (2002). *Psicologia Sociale*. Bologna: Il Mulino.
- Park, B., & Rothbart, M. (1982). Perception of out-group homogeneity and levels of social categorization: Memory for the subordinate attributes of in-group and out-group members. *Journal of Personality and Social Psychology*, 42, 1051–1068.

- Phillips, N., Lawrence, T. B., & Hardy, C. (2004). Discourse and institutions. *Academy of Management Review*, 29(4), 635–652.
- Polos, L., Hannan, M. T., & Carroll, G. R. (2002). Foundations of a theory of social forms. *Industrial and Corporate Change*, 11(1), 85–115.
- Powell, W., & DiMaggio, P. (1991). *The new institutionalism in organizational analysis*. Chicago: University of Chicago Press.
- Rao, H., Monin, P., & Durand, R. (2003). Institutional change in Toque Ville: Nouvelle Cuisine as an identity movement in French gastronomy. *American Journal of Sociology*, 108(4), 795–843.
- Rosch, E. (1978). Principles of categorization. In: E. Rosch & B. B. Lloyd (Eds), *Cognition and categorization*. Hillsdale, NJ: Erlbaum.
- Ruef, M. (2000). The emergence of organizational forms: A community ecology approach. *American Journal of Sociology*, 106(3), 658–714.
- Ruef, M., Aldrich, H. E., & Carter, N. M. (2003). The structure of founding teams: Homophily, strong ties and isolation among U.S. entrepreneurs. *American Sociological Review*, 68(2), 195–222.
- Scott, W. R., Ruef, M., Mendel, P. J., & Caronna, C. A. (2000). *Institutional change and healthcare organizations. From professional dominance to managed care*. Chicago: The University of Chicago Press.
- Shane, S., & Venkataram, S. (2000). The promise of entrepreneurship as a field of research. *Academy of Management Review*, 25(1), 217–226.
- Simons, T., & Ingram, P. (2003). Enemies of the state: Interdependence between institutional forms and the ecology of the kibbutz, 1910–1997. *Administrative Science Quarterly*, 44, 592–662.
- Simons, T., & Ingram, P. (2004). An ecology of ideology: Theory and evidence from four populations. *Industrial and Corporate Change*, 13, 33–59.
- Strang, D., & Meyer, J. W. (1994). Institutional conditions for diffusion. In: W. R. Scott, J. W. Meyer & Associates (Eds), *Institutional environments and organizations* (pp. 100–112). Thousand Oaks, CA: Sage.
- Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review*, 20(3), 571–610.
- von Martius, C. F. P., & von Spix, J. B. (1823–1832). *Reise in Brasilien*. Munich.
- Weick, K. E. (1995). *Sensemaking in organizations*. Thousand Oaks, CA: Sage.
- Wilson, A., & West, C. (1988). The marketing of unmentionables. In: F. Miglioli (Ed.), *Social and economic business*. Milano: Cisalpino.
- Wittgenstein, L. (1967). *Ricerche filosofiche*. Torino: Giulio Einaudi Editore.
- Zucker, L. G. (1989). Combining institutional theory and population ecology: No legitimacy, no history. *American Sociological Review*, 54(4), 542–545.