department of economics
"S. Cognetti de Martiis"

Via Po, 53 – 10124 Torino (Italy)
Tel. (+39) 011 6704917 - Fax (+39) 011 6703895
URL: http//www.de.unito.it

WORKING PAPER SERIES

INVENTORSHIP AND AUTHORSHIP AS ATTRIBUTION RIGHTS: AN ENQUIRY INTO THE ECONOMICS OF SCIENTIFIC CREDIT

Francesco Lissoni and Fabio Montobbio

Dipartimento di Economia "S. Cognetti de Martiis"

LEI & BRICK - Laboratorio di economia dell'innovazione "Franco Momigliano"
Bureau of Research in Innovation, Complexity and Knowledge, Collegio Carlo Alberto

Working paper No. 21/2012

Università di Torino

INVENTORSHIP AND AUTHORSHIP AS ATTRIBUTION RIGHTS: AN ENQUIRY INTO THE ECONOMICS OF SCIENTIFIC CREDIT⊗

Francesco Lissoni•, Fabio Montobbio★

•Università degli studi di Brescia (Italy)
★Università di Torino, Dept. of Economics "S.Cognetti de Martiis" (Italy)
• and ★KITES – Università "L. Bocconi", Milan (Italy)
francesco.lissoni@unibocconi.it, fabio.montobbio@unito.it

> "Why does your name even appear on the paper?"
>
> "I am the one who suggested the problem [...] I prepared the grant application to the NIH. [...] Without such support [my student] could do nothing. I'm not just talking about the fellowship. [...] There's both a teacher-apprentice relationship and collegiality."
>
> (Djerassi C., *Cantor's Dilemma*, Penguin Books, 1989; pp.50-51).

> "I think there's rarely more than one inventor. I mean, if you wake up and you have an idea, that's the invention. And then there's all this work around it, of course ... [The postdoctoral researchers] contributed to the work, but they didn't do any really innovative work [...] They don't have time to think as much, they have a lot of manual labor to do"
>
> (McSherry C., *Who Owns Academic Work?*, Harvard Univ. Press; 2003; p.84)

Abstract: Authorship and inventorship contribute to the reputation of individual scientists and are distributed across several individuals, due to the importance of teamwork in both science and technology. For academic teams that both publish and patent their research results, we compare the social and legal norms that affect the negotiation over the distribution of authorship and inventorship. We use text-mining techniques to identify 680 "patent-publication pairs" (related sets of patents and publications), for a sample of Italian academic scientists. On average, the number of co-authors is larger than the number of co-inventors, especially in medical-related fields. First and last authors have a lower probability of being excluded from inventorship. However, the probability of exclusion also declines with seniority, as expected from social norms. Women have a significant higher probability to be excluded, other things being equal. Long-lasting doubts on the reliability of authorship as a tool for allocating scientific credit are reinforced, and can be extended to inventorship. Results for attribution rights in science, as we obtain here, raise questions and provide insights on other settings in which attribution rights are both relevant and distributed within teams.

Keywords: economics of science; intellectual property; patent-publication pairs; scientific credit; authorship

JEL: Codes: O31, O34, L30

⊗ We thank Maurizio Tosetti and Antonio Della Malva for valuable research assistance. We thank also two anonymous referees for their comments and suggestions. Various drafts of the paper (with different titles) have been presented at Case Western Reserve University, SPRU-University of Sussex, the University of Manchester, the International Centre for Economic Research (ICER, Turin), Georgia Institute of Technology, the Copenhagen Business School, Paris XIII and the University of Piemonte Orientale (Alessandria, Italy). Mario Biagioli, Marco Giarratana, Fiona Murray, Scott Stern, John Walsh and Lorenzo Zirulia provided extended comments and encouragement. Stefano Breschi and Gabriella Pasi provided useful advice on text-mining techniques. Usual disclaimers apply.

INVENTORSHIP AND AUTHORSHIP AS ATTRIBUTION RIGHTS: AN ENQUIRY INTO THE ECONOMICS OF SCIENTIFIC CREDIT⊗

> "Why does your name even appear on the paper?"
>
> "I am the one who suggested the problem […] I prepared the grant application to the NIH. […] Without such support [my student] could do nothing. I'm not just talking about the fellowship. […] There's both a teacher-apprentice relationship and collegiality."
>
> (Djerassi C., *Cantor's Dilemma*, Penguin Books, 1989; pp.50-51).

> "I think there's rarely more than one inventor. I mean, if you wake up and you have an idea, that's the invention. And then there's all this work around it, of course … [The postdoctoral researchers] contributed to the work, but they didn't do any really innovative work […] They don't have time to think as much, they have a lot of manual labor to do"
>
> (McSherry C., *Who Owns Academic Work?*, Harvard Univ. Press; 2003; p.84)

1. Introduction

Understanding how scientific knowledge is produced and reduced to practice is a central theme of today's economic research, due to the importance of science-based innovation for market dynamics and economic growth (Nelson and Romer, 1996; Stokes, 1997). Both the sociology and the economics of science pay a great deal of attention to the system of incentives affecting academic scientists' choice of topics and dissemination tools, and to the role played by reputation and intellectual property (Stephan, 2010). This paper contributes to this major line of enquiry, by providing an empirical analysis of the distribution of reputation among scientists working in teams, and engaged in the "simultaneous disclosure" of scientific and commercial knowledge, by means of publications and patents (Gans et al., 2010).

An increasing number of scientists earn their reputation following a path of dual knowledge disclosure (Ducor, 2000; Murray, 2002; Murray and Stern 2007, 2008; Gans et al. 2010), by which they publish their discoveries in scientific journals (authorship) and, at the same time, obtain intellectual property (inventorship). Authorship plays a key role in academic careers, as scientists' reputation is built first and foremost on the number and impact of their publications (Merton, 1957; Dasgupta and David, 1994). Patents also contribute to reputation because they signal the economic value of discoveries and personal competencies to perspective employers, buyers of consultancy services, and funding agencies, especially in disciplines such as engineering and the life sciences (Murray, 2004; Link et al., 2007)[1].

Both authorship and inventorship are 'attribution rights': a form of intellectual property according both to the social norms of science (Zuckerman, 1968) and to legal regulations of "moral

[1] See also the guidelines of governmental research evaluation exercises, such as RAE (2008), and CIVR (2006). Of course, patents may also produce immediate economic benefits in the form of royalties or fees

rights" of authors and performers, as disciplined by international conventions (article 11 in UNESCO, 2001; and article 6 in WIPO, 2008). If assigned correctly, they provide information to any party interested in recruiting, promoting, rewarding or sanctioning the rights' holders, with benefits for both the parties involved and society at large (Hansmann and Santilli, 1997). They cannot (at least in principle) be sold or traded, because they also serve as liability signs in case of misconduct (such as scientific fraud, plagiarism or the violation of trade secrecy). They play an important role not only in science, but also in a wide range of economic activities where individuals are creative and build their careers on personal reputation, from engineering services to advertising to media (Fernandez-Molina and Pais, 2001; Fisk, 2006).

Assigning attribution rights is difficult when creative activities are performed by teams, rather than individuals. And this is precisely the case of science and technology, where teamwork is now a dominant feature (Katz and Martin, 1997; Jones et al., 2008; Wuchty et al., 2007; Jones, 2009), Team work bring together several individuals' contributions, whose boundaries of are often difficult to locate. Besides, ambiguities in the social and legal norms defining attribution rights exist, and negotiation within teams occur in order to resolve such ambiguities. Inefficiency in the distribution of reputation may occur, to the extent that the negotiation outcome reflects **the members' private evaluation of attribution rights or their bargaining power**, and not only their individual contribution to the team's output. Under these circumstances the negotiation may result into the destruction of information relevant for third parties, with negative welfare effects, such as ill-informed recruitment and funding decisions.

The main empirical question of this paper is to what extent the distribution of authorship and inventorship in academic teams reflect not only individual contributions, but also the hierarchical position, seniority and gender of team members. We theorize that the latter bear an influence on the private value attached by team members to being recognized as authors and/or inventors, as well as on their bargaining power. We then test our propositions by using patent publication pairs (PPPs). A patent and a paper form a pair when they disclose the same research result, and at least one author and one inventor are the same person. Using text mining techniques we build an original sample of 680 PPPs produced by 308 Italian academic inventors between 1975 and 2002, in the fields of Chemical Engineering, Electronic Engineering and Telecommunications, Pharmacology, and Biology. We complement these data with related bibliometric and biographical information on the selected academic inventors and their co-authors.

We find that inventorship is attributed more sparingly than authorship. As a result, several authors of scientific publications are excluded from the list of inventors of the related patents. Such exclusion cannot be entirely explained by heterogeneity in individual contributions to the research project (proxied by the scientist's position in the author by-line of the team's publications), as legal norms on inventorship would imply. In fact, we find that junior and women co-authors are more at risk of being excluded from inventorship, other things being equal. We interpret this evidence as suggestive of the importance of negotiation of attribution rights, and we discuss implications for the economics and policy of science, and for further research in all fields of activity where attribution rights matter.

The paper is structured as follows. In section 2 we recall the increasing importance of teams in publishing and patenting and discuss the concepts of inventorship and authorship. In section 3 we develop a conceptual model and the related hypotheses to be tested. In section 4 we describe our methodology for the identification of PPPs, the econometric model and the main variables. In section 5 we describe the data and estimate the probability for the co-author of a publication to be excluded from the related patent, as a function of her contribution to the publication, seniority, gender, and experience. We also perform robustness checks and discuss the implications and limitations of our analysis. Section 6 concludes.

2. Research teams and problems of attribution

2.1 The increasing importance of teams in publishing and patenting

The average number of authors per publication and inventors per patent has been increasing over time. By considering all scientific publications listed by the ISI Web of Science database, Wuchty et al. (2007) estimate that the average number of authors per paper has gone from 1.9 in 1955 to 3.5 in 2000. For patents at the US Patent & Trademark Office (USPTO), the same authors estimate an increase from 1.7 inventors per patent in 1975 to 2.3 in 2000[2]. According to Jones (2009), scientific work is increasingly specialized and therefore requires teams of increasing size. In addition the growing need of sharing data and facilities generate multi-team research which is conducive to multi-authorship(Katz and Martin, 1997; Jones et al., 2008).

Notably, the average number of inventors per patent remains lower than the average number of authors per publication, even for comparable technological and scientific fields (Meyer and Bhattacharya, 2004). One possible explanation is that patents originate mostly from industrial research, funded by business companies and carried out by their employees. The proprietary nature of such research limits the inventors' freedom to choose their research team partners, contrary to what happens to academic scientists, and suggests caution in putting together several teams. However, differences in the number of authors and inventors can also be found when comparing patent-publication pairs, that is patents and publications with the same contents and produced by same research team and programme (Ducor, 2000; Murray, 2002). In this case, the only possible explanation is that the qualifying criteria for

[2] Our own elaborations over data from the European Patent Office suggest an increase from 1.95 inventors per patent in 1980 to 2.46 in 1999; when considering only patents in a science-based fields such as organic chemistry, the figures are respectively 2.76 and 3.88 (data available on request).

being considered either an author or an inventor are different, or that some differences exist in the established practices of attribution. A vast literature exists, which illustrates how negotiation and social conventions play a role in authorship attribution. A similar, albeit much more limited literature can be found for inventorship. We examine both of them.

2.2 The Vexed Issue of Authorship

A vast literature exists on mis-attribution practices in scientific authorship, such as 'guest' (or 'honorary') and 'gift' authorship, which occurs when a scientist is listed in the authors' by-line of a paper to which she has not contributed (Mowatt et al., 2002). Ambiguity about individual contributions makes publications less useful as a signal of scientific credit and threatens the ethical integrity and credibility of the entire research from the rigour of the methodology to the quality of data (Biagioli, 1998). These problems are particularly severe for biomedical research, because of the great importance of responsibility attribution in that field. As a consequence, since 1985, the *International Committee of Medical Journal Editors* has published and updated the 'Uniform Requirements for Manuscripts Submitted to Biomedical Journals'. The most recent edition states that:

> "Authorship credit should be based on 1) substantial contributions to conception and design, or acquisition of data, or analysis and interpretation of data; 2) drafting the article or revising it critically for important intellectual content; and 3) final approval of the version to be published [...] Acquisition of funding, collection of data, or general supervision of the research group, alone, do not justify authorship" (ICMJE, 2007).[3]

According to the ICMJE Requirements, therefore, a heterogeneous set of authors can be listed together in the same by-line. For example, a scientist who has limited himself to an entrepreneurial role (such as chasing grants, "conceiving and designing" the paper, and revising it "critically") could be listed along with a colleague who has carried out most of the work (such as acquiring, analysing and interpreting the data, drafting the manuscript, and providing the technical expertise). Despite such latitude, the ICMJE Requirements have been largely ignored by the scientific community. For example, Bates et al. (2004) find that 60% of 72 articles surveyed in 2002 in the *Annals of Internal Medicine* and 21% of 107 articles in the *British Medical Journal* have at least one author that does not meet the first ICJME criterion. Similar results are found by Hwang et al. (2003) for the *Journal of Radiology* (see also references therein on *Lancet* and the *Dutch Medical Journal*). This suggests that authorship attribution remains a subjective decision, which is negotiated within research teams, according to customary rules that do not necessarily match editorial guidelines.

Name-ordering in the authors' by-line is often used to shed light on individual contributions. Although general authorship guidelines do not provide mandatory recommendations, two major

[3] Similar rules, albeit less detailed, can be found in the authors' guidelines of the International Electrical and Electronic Engineering association (IEEE, 2008; Section 8.2.1.A).

traditions exist: alphabetical ordering, which is typical, for example, of the social sciences, and contribution-related ordering, which is most common in the hard sciences and is explicitly recommended by some scientific societies (for a review, see Rennie and Flanagin, 1994; and Drenth, 1998). In their study on medical publications, Mowatt et al. (2002) calculate that 76% of by-lines assign the first position to the person who contributed most significantly to the study, while only 2% list authors alphabetically. Of the remaining 22%, seniority criteria were involved, such as listing the senior author last. Zuckerman's (1968) seminal work on Nobel laureates revealed that name ordering decisions are most often delegated to senior investigators, who base their judgement both on contribution and seniority.

The message conveyed by the first and last positions in a non-alphabetical by-line is relatively unambiguous: the first author is usually the scientist, possibly a junior one, who has contributed most to the paper; the last is a more senior investigator, who runs the lab, chases the grants, and sets the research strategy. The same cannot be said for the authors in between. These may be either effective contributors to the paper (although less important and/or more senior ones than the first author), but they may also be guest authors of many sorts (such as laboratory technicians rewarded for their dedication, or very senior scientists listed out of deference).

Contribution measurement is also difficult because individuals tend to overestimate their own inputs (Hoen et al., 1998; Johnson and Orback, 2002; for a discussion in economic terms, see Van den Steen, 2004). Maintaining some degree of ambiguity in the notion of "author" is then necessary to temper tensions within the team.

2.3 Inventorship

Unlike scientific authorship, inventorship is a legal concept which bears direct economic consequences. In the US, a patent may be declared invalid if the designated inventors' contribution does not match the legally defined one[4]. According to Title 35 of the US Code (as amended in 1984), two individuals can be designated as inventors on the same patent only if they have worked "jointly" and provided some kind of "inventive" contribution (Fasse, 1992, pp. 172-173). In particular, *each* person named on a patent must have contributed to the *conception* step in the invention (as defined by the claims). Conception is "the formation, in the mind of the inventor of a definite and permanent idea of the complete and operative invention, as it is to be applied in practice" (Hybritech Inc. v. Monoclonal Antibodies, Inc.).[5]

[4] See for example *Yeda Res. & Dev. v. ImClone Systems Inc.* in 2006.
[5] 802 F.2d 1367, 1376 (Fed. Cir. 1986)

In Europe, even with patents issued by the European Patent Office (EPO), inventorship is ultimately defined by the various national legislations. For example, in the United Kingdom the inventor is defined as the "actual deviser of the invention...", who in turn is the person who contributes to the novelty (inventive step) of the claims listed in the patent application (s7-3 Patents Act, 1977). In Italy, as in many other countries, no specific definition of inventor is provided by legal texts. The legal doctrine on the identification of authors and inventors coincide, with the latter being simply defined as the “author of an invention”. Mis-attribution of inventorship does not appear to threaten the validity of the patent, but it may cause re-allocation of the property rights.

Existing norms in both the US and the European legislations for defining inventorship are more restrictive than the editorial rules defining authorship. As stressed by the legal opinions of university TTO officers and IPR consultants, being involved in the conception of the invention is a requirement that several authors of scientific publications may fail (Bennett and Biswas, 1997; Hutchins, 2003; Vinarov, 2003). For example, current interpretations of the US law suggest that “merely suggesting a desired result” or “having entrepreneurial involvement” do not qualify as inventorship. A scientist who raises funds, conceives the initial experiment, and revises the draft paper could qualify as the author of a project-related paper (for example, according to the ICMJE guidelines we described above), but not as the inventor of any project-related patent. At the opposite end, “following the complete instructions” of a colleague or superior does not qualify anybody as an inventor; and even joining a research team too late, after its members have conceived the key characteristics of the desired invention, may be a cause for exclusion from inventorship[6].

However, much controversy surrounds also the concept of inventorship and its applications to team work (Fasse, 1992). It is also likely that decisions on inventorship attribution, very much like those on authorship, depend upon the discretionary judgement of the most senior members of the team, who often manage the economic details of the research and exercise authority Evidence on this point is provided by Colyvas (2007), for the case of Stanford university. An inventors’ survey by Jaffe et al. (2000) and several interviews to academic inventors conducted by MacSherry (2003) show that inventors share with authors the tendency to over-value their contribution to the team effort.

The practicalities of inventorship attribution also leave room for negotiation among team members. Very much like journal editors, patent office examiners leave the identification of inventors entirely to the applicants. At most, signed declarations are required. If not challenged in court, these initial attributions remain un-scrutinized because patent offices pay attention only to the technical contents of the patents they are called to judge, and not to the inventors’ identity.

---

[6] . The latter cases bring to mind situations in which a junior scientist or a graduate student may be rewarded with authorship for her brilliant assistantship, but not with inventorship. For a case of a student’s exclusion from a patent, see Fasse (1992; p. 282). More cases of disputes within academic teams are mentioned by McSherry (2003) and Seymore (2006).

3. Negotiation over Authorship and Inventorship

We focus on patents and publications by academic scientists that are joint products of team work. Individual contributions to the research results vary by contents (organizational, conceptual, technical…), intensity, and creativity; however, they can be acknowledged by only two forms of attribution rights, namely authorship and inventorship. Team members need to decide who among them can be defined as author and/or inventor and negotiate to provide all team members with the right incentives to contribute to the research effort, and with the least incentive to litigate *ex post*. Legal ambiguities, together with the difficulty for non-team members to detect and sanction mis-attribution, grant the team members some freedom in trading one form of attribution rights for another, or to give up one or both altogether, in exchange for some form of compensation.

In particular, we expect team members to be more generous towards authorship attribution, which entails only a reputational reward, than towards inventorship, which could also lead to tangible economic benefits and legal hiccups. In particular, no team member whose contribution is unambiguously unfit to be identified as an inventor ought to be named as such, in order to minimize the risk of litigation. Moreover by applying non-alphabetical **name ordering**, team members have the possibility to fine-tune the distribution of reputation derived from authorship, so that the addition of an author to a publication (especially if in a middle position) detract less from the reputation accruing to other team members than the addition of an inventor to the patent.

As a result, negotiation is likely to end up with team members agreeing upon distributing authorship rather widely, while restricting inventorship to a subset of team members.

The process of negotiation may also be affected **by seniority and gender**. Life-cycle models of scientists' behaviour suggest that junior scientists who pursue an academic career invest heavily in building a reputation, while their senior colleagues may choose to cash in the reputation they have already acquired, or to trade it for immediate economic returns (Stephan and Levin, 1992; Audretsch and Stephan, 1999). As a consequence, junior scientists **will value authorship more than inventorship**. When compared to patents, papers circulate more widely and contain a much more readable description of the research results and of their relevance for the academic community. In addition, although possibly submitted after the filing of the patent, they start circulating earlier[7].

[7] Under the Patent Cooperation Treaty (PCT), as well as rules followed by EPO, patent applications remain secret until the publication of the search report, a document produced by the patent examiner that assesses the novelty and non-obviousness of the patent claims. For non-PCT applications at USPTO, secrecy may last until the patent is granted, that is several years after the filing date. The refereeing and publication process at scientific journals is much shorter, and in any case does not impede the circulation of working papers and conference proceedings. Besides, no established diffusion channels and procedures exist for not-yet-published patents.

Moreover junior scientists may find that the expected financial and reputational benefits generated by being acknowledged as inventor are lower than the costs. The latter consist in the extra effort to deal with the patenting procedure (drafting a new document, discussing with lawyers and TTO staff) and, in the possibility to enter into conflict with senior team members (from whose patronage a junior scientist's career depend), in case they are reluctant over sharing or giving up inventorship. On the contrary senior scientists, whose reputation as authors is established and opportunity of further career advancements may be more limited, may attach **more importance to the potential economic benefits derived from the patent, and to increasing their reputation as inventors**. So we expect that ambiguous cases of attribution have a higher probability to be solved, other things being equal, by denying inventorship to junior scientists while granting it to senior ones.

This explanation of the relationship between seniority and the distribution of authorship and inventorship may be complemented by two others. First, a senior scientist is more likely to be a permanent team member, who may claim inventorship (besides authorship) also as **compensation for past services** rendered to the team, or for **future ones** (such as help for developing the invention into an innovation). Second, but similarly, **a junior scientist may have reached the team after the research project had started and the decision to file the patent had been taken, inclusive of a decision over inventorship distribution**.

Gender may also affect the negotiation process. Several recent studies find that academic inventors' publishing activity affects positively their propensity to patent, but also that controls for gender are statistically significant: women scientists apparently patent less than men, other things being equal (Breschi et al., 2005; Azoulay et al., 2007; Stephan et al., 2007; see also Whittington and Smith-Doerr, 2008). Explanations for this result are put forward by Ding et al. (2006) and Murray and Graham (2007), on the basis of longitudinal data on careers and interviews. Women appear to have fewer connections to operators in the market place, which diminishes their opportunities to commercialize their research results or to cash in their reputation through consultancy or participation to high-tech companies. Women also appear to be more concerned with the difficulty of reconciling their academic career with engagement in commercial ventures. This suggests that, like junior team members, women scientists may value inventorship less than authorship, and therefore be more willing to give it up when it comes to negotiating attribution rights[8].

The literature we discussed in section 2.2 suggests that the negotiation process is largely implicit, and that most often the decisions are taken by the laboratory head or project leader (therefore a senior scientist), as part of his or her managerial duties. Our own informal interviews confirm this. This

[8] Notice that this type of gender bias in patenting comes on top of the well-documented gender bias in scientific productivity and academic career opportunities. As far as our analysis is concerned, the latter may affect women scientists' type of contribution to research resulting in joint patents and papers, which we expect to be reflected by the scientist's position in the author by-line. We will confront this problem when discussing the specification of our econometric model.

assumption leaves some room for the possibility of arbitrariness, by which we mean a senior scientist's decision to exclude a junior or female team member from inventorship, irrespective of the latter's contribution and will.

Summing up, for any set of related patents and publications (joint products of the same research effort), we expect to find:

i. A higher number of authors than inventors, that is evidence of some authors' exclusion from inventorship.

ii. A relationship between an author's position in the (non-alphabetically ordered) by-line of the publication, and the probability of his/her exclusion from inventorship. In particular, we expect first and last authors, whose contribution are signalled as distinctive, to be less likely to be excluded from the patent than middle authors.

iii. A negative correlation between authors' seniority (as measured by age or publication stock) and her probability of exclusion, other things being equal.

iv. A higher probability for women authors to be excluded from inventorship, other things being equal.

Hypothesis iii and iv are especially important, as they would suggest that the negotiation outcome would reflect not only the specific individual contributions, but also the social norms deriving from the structure of incentives and career paths in science.

Finally, we have noticed how the discussion of authorship attribution has been much more lively for the life sciences than for other disciplines, which may be indicative of more controversy over attribution rights (possibly due to more important ethical implications of mis-attribution). Besides, journals from different disciplines may follow different editorial guidelines concerning authorship. As a consequence, although we do not have a clear a priori on the influence of discipline on attribution, we think it necessary to control for it.[9]

4. Data and Methodology.

Our methodology is based on the identification of *patent-publication pairs* (PPPs). Theoretically, a patent and a paper form a pair when they represent an instance of "simultaneous disclosure" of a set of

[9] We control also for ownership patterns. Patents by academic scientists may be either assigned to business companies or to the inventors' universities to the inventors themselves (Lissoni et al., 2008; and references therein). But in the absence of any appreciable result, and of any strong a priori on our part, we dropped this part of the analysis.

research results having both scientific interest and commercial value (Gans et al., 2010). Empirically, we define a patent and a paper to form a pair when the same idea is described to some extent in both documents, and at least one author and one inventor are the same person. Scientific papers and patents differ widely in contents. The former describe a set of theories and/or experimental results, and emphasize the originality and neatness of the results, whereas the latter describe the features of a new product or process, of which they emphasize the novelty and utility, by laying out a list of claims. However, in "science-based" technologies and engineering, it is often the case that a patentable advancement is also worth publishing in refereed journals. In this case, we may expect highly specific words to be present in both documents.

A set of recent papers have made use of PPP datasets built more or less manually. Ducor (2000) performed a manual search of various databases for proteins with specific genetic or aminoacid sequences, finding 40 pairs. Murray's (2002) study concerned a single patent-paper pair on tissue engineering in cartilage. Murray and Stern (2007 and 2008) compared 340 articles published in *Nature Biotechnology* between 1997 and 1999 with their authors' patents at the USPTO, ending up with 169 PPPs, all of them selected through careful reading of both types of documents. The number of patents and publications needed for our analysis is so large that we could not rely on manual search and reading. So we applied established methods of data mining and information retrieval, as follows:

(1) From the KEINS patent database, we extracted all Italian academic inventors from the four academic disciplines with the highest share of inventors over the total number of scientists, namely: Chemical Engineering, Biology, Pharmacology, and Electronic Engineering & Telecommunications. These are 218 individuals, who appear as inventors in 389 EPO patent applications from 1978 to 2001.[10]

(2) For the selected academic inventors, we collected all publication data from the ISI-*Web of Science* (ISI-WoS), from 1975 to 2003.[11]

(3) Based upon titles and abstracts, we matched the selected academic inventors' patents to their scientific articles, thus obtaining a pool of patent-publication pairs (PPPs). More details on this in the following section.

(4) Again from ISI-WoS, we collected all the publication data for the academic inventors' *co-authors*, in order to establish the latter's' first year of activity (first year in which a publication in their names appears in ISI-WoS) and their number of publications[12].

[10] The KEINS database contains information on all academic scientists designated as inventors on EPO patent applications filed either by universities, public research organizations or business companies, for a number of European countries (Lissoni et al., 2006, and 2008). It also contains information on individual characteristics of the scientists (such as age, affiliation, academic rank, discipline), as well as any information from the front page of their patents (priority dates, titles, abstracts, and applicants' names). Italian scientists listed in the KEINS database include professors from all ranks (assistant, associate and full), but no PhD students, post-docs or other non-tenured faculty.

[11] More details on these data in Breschi et al. (2007, and 2008).

(5) We established the gender of as many as possible co-authors, by manually retrieving several their publications and looking at their first names[13].

4.1 Patent-publication pairs: methodology

Given *t* the priority year of a patent and *i* the individual listed among its designated inventors, a *potential* patent-publication pair is defined as the association between the patent and a publication that has individual *i* among its authors and has been published in the period [*t-2, t+2*]. After excluding all duplications (which may occur when two or more patents or two or more publications have the same co-inventors or co-authors and title), all publications with no abstracts, and all patents which their inventors declared to be unrelated to any publication of theirs the final sample of *potential* patent-publication pairs is composed of 6810 pairs, 389 patents and 2838 publications.[14]

For all documents in this *potential* PPP set we examined the title and abstract, and transform them into comparable information sets. The first step of the transformation consisted in removing uninformative terms such as pronouns, conjunctions, and the most frequent nouns and verbs ("stop words") from both titles and abstracts.

In the second step, we applied a traditional data-mining technique, the *bag of words* method, effective (Salton and McGill, 1983; Leopold et al., 2004). For each disciplinary field we built a complete set of words from the titles and abstracts of all the patents and publications, so that each document *j* (patent or publication) could be represented by a vector. Each cell (*i,j*) in the vector has a value equal to 1 if word *i* appears in document *j*, and 0 otherwise (Bassecoulard and Zitt, 2004). This vector representation may be used to produce a large number of “similarity measures” between patents and publications. The most common one, which we adopted, is the *cosine similarity measure* (*S*).

---

[12] Due to problems of homonymy we selected, for each co-author's name, only the publications in fields "similar" to those of the related academic inventors (for a total of 99 fields). In order to do so, we applied a methodology proposed by Engelsman and Van Raan (1992) and Breschi et al, (2003). Let I be the set of academic inventors. Let $F_{ik} = 1$ if professor i signed at least one article in a journal in field k (k = 1. ...99). and $F_{ik} = 0$ otherwise. $N_k = \sum_{i\in I} F_{ik}$ will be the number of professors with at least one article in field k. We calculate the number of professors with articles in both fields $\ddot{k}$ and $\bar{k}$ as:
$C_{\ddot{k}\bar{k}} = \sum_{i\in I} F_{i\ddot{k}} F_{i\bar{k}}$.
and produce a square (99 x 99) symmetrical "matrix of co-occurrences", with generic cell $C_{\ddot{k}\bar{k}}$. We then calculate a symmetrical “similarity matrix” of the same size, whose generic cell contains a similarity score $S_{\ddot{k}\bar{k}}$ defined as:

$$S_{\ddot{k}\bar{k}} = \sum_{k=1}^{99} C_{k\ddot{k}} C_{k\bar{k}} \Bigg/ \sqrt{\sum_k C^2{}_{k\ddot{k}}} \sqrt{\sum_k C^2{}_{k\bar{k}}}$$

with continuous range (0,1). Finally, we used the similarity matrix as an input to a multidimensional scaling algorithm, which arranges the various fields on a 2-D plane by reducing the dimensionality of the data (Klavans and Boyack 2006). The mapping of the subfields over the first two dimensions suggested the existence of four meta-fields: Biomedicine. Pharmaceuticals. Materials Chemistry and Engineering. Fields falling in the same meta-fields were then defined as "similar".

[13] Since ISI-WoS list authors' surnames and name initials, our initial database did not have information on gender, whose retrieval was however requested by the referees (and rightly so), Due to the high costs of the manual retrieval, however, we limited the search to the individuals who enter the ultimate sample used in the regressions (see section 4 below). This returned information on 841 out of 899 individuals.

[14] Academic inventors’ declarations on the existence publications related to their patents were collected by means of structured phone interviews. Among other things, interviewees were asked, with reference to each of their patents, whether or not they had published any related research results. Responses were obtained from 154 out of 308 inventors, for a total of 372 patents out of 552. Overall, interviewees confirmed the existence of a patent-related publication for 86% of the patents.

If x_{ij} is the value of the binary variable for document *j* and word *i*, S measures the similarity between a document *k* and *s* as follows:

$$S(k,s) = \sum_i x_{ki} x_{si} \Big/ \sqrt{\sum_i x^2{}_{ki}} \sqrt{\sum_i x^2{}_{si}}$$

Theoretical values of *S* are in the continuous [0,1] range. In our application, S takes values comprised between 0 and 0.75. For our analysis, we selected those PPPs whose S value falls in the top 10% of the distribution, which is comprised between 0.145 and the maximum, for a total of 680 PPPs, resulting from 213 patents, 1138 different authors and 450 publications.[15]

It is important to note that, differently from manual methodologies, our bibliometric approach does not presume a one-to-one match between patents and publications (one patent corresponding to just one publication, and *vice versa*). On the contrary, we produce a large number of one-patent-to-many-publications matches, and several many-to-many ones. This is not unexpected: a good research project will certainly produce more than one result worth of publication, and possibly more than one patent. [16]

The large number of PPPs derived from one-to-many and many-to-many matches suggests that the appropriate unit of analysis may be the overall team of authors (inventors) listed in a set of related publications (patents). This is because, within a research team, the negotiation of authorship and inventorship may refer not to the single item (publication or patent) but to the overall set: for example, an author who has been excluded from one patent can be included in a related one.

Aware of this possibility, in the empirical analysis we mainly use our selected 680 PPPs as distinct units of analysis, but, we also run a set of additional regressions in which the unit is the set of all patents linked to one publication, either from one-to-one, one-to-many, or many-to-many matches.

4.2 Model and the main variables

We estimate the probability of an author's exclusion from inventorship, and we arrange the database accordingly. In particular, each PPP *j* is repeated as many times as the number of authors *i* in the publication related to PPP *j*. So for each author *i* and PPP *j* we know whether she is excluded or not

[15] Table A1 gives an example of a PPP. In order to check the robustness of the matching method we also used three other selection methods to find the actual patent-publication pairs: (1) For each potential PPP, we compared the patent and publication abstracts, and calculated the number of words that are the same in the two documents. Then we calculated the share of words that are the same in the total number of words in the patent abstract. (2) We selected actual PPPs simply on the basis of the answers to the phone interviews mentioned in the previous footnote, for a total of 3380 pairs. (3) We applied again the bag-of-words method, but with cells in the vectors containing frequencies (number of occurrences for each word in the documents) instead of dummies, and we selected once again the patent-publication pairs in the top 10 percentile (S ranging from 0.206 to 0.81). The descriptive results we found were always similar and are available on request.

[16] One-to-one matches produce 44 PPPs out of 680. As for one-to-many matches, they involve 76 patents matched to 271 publications, and originate 271 PPPs. Many-to-many matches account for a total of 346 PPPs. The many-patents-to-one-publication case is much rarer, with 6 publications associated with 20 patents, for a total of 20 PPPs. It is likely that scientists facing patentable research results will tend to publish them separately (in order to keep the length of articles under control, or to follow a "salami slicing" strategy), but to patent them jointly. In fact, the patent fee structure provides many incentives to pool several claims into a single application.

from the patent related to PPP *j*. We model the probability of exclusion as a function of both the author's contribution to the research effort and her personal (biographical, professional) characteristics. Our dependent variable *y* is the exclusion event, with y_{ij}=1 if author *i* of a publication in PPP *j* is excluded from the inventorship of a patent in the same PPP, and y_{ij}=0 otherwise. $Pr(y_{ij}=1|x)$ is the probability that author *i* is excluded from a patent in PPP *j*, conditional on a set of variables *x* that describe the characteristics either of the author or of the PPP.

The author's characteristics we consider are:

- The contribution to the publication, as measured by the author's position in the by-line, transformed into three dummy variables: FIRST, LAST, and MIDDLE (reference case). Following the discussion in section 3, we expect both FIRST and LAST to bear a negative sign[17].

- Seniority, measured either in absolute terms or relative to the other authors of the publication. We measure author *i*'s absolute SENIORITY, as the difference between the priority year of patent in PPP *j* (time of the invention t_{patj}) and the year of her first publication (t_{fpi}). As for relative seniority, we measure it with a continuous variable, ranging from 0 to 1, defined as:

  $$RELATIVE\ SENIORITY_{ij}=(t_{fpi}-t_{0j})/(t_{1j}-t_{0j})$$

  where t_{0j} and t_{1j} are the years of the first publication of, respectively, the most and the least experienced among all the authors of the publication in PPP *j*. Alternatively, we measure relative seniority with two dummy variables, MOST_SENIOR and MOST_JUNIOR, which take value one, respectively, for RELATIVE SENIORITY=1 and RELATIVE SENIORITY=0. We expect SENIORITY, RELATIVE_SENIORITY and MOST_SENIOR to bear a negative sign, and MOST_JUNIOR to bear a positive one.

- Professional experience, measured in either absolute or relative terms. In absolute terms, we use the stock of individual *i*'s publications (PUB_STOCK_i) one year before the patent's priority date (t_{patj}-1). In relative terms, we build a continuous variable, ranging from one to zero:

  $$RELATIVE\_PUB\_STOCK_{ijtpat} =$$

[17] The information provided by the name order of authors may vary between papers co-authored by several members of one research team only, and papers co-authored by authors from many teams. In the latter case, authors may be listed first according to the team they belong to (with teams ordered either according to substantive criteria or not) and then either alphabetically or according to their contribution within the team. In the case of team+alphabetical order, our dummies cannot be interpreted any more as proxies of the individual's contribution. In the case of team+contribution order, the FIRST and LAST dummies still bear an unequivocal meaning (they indicate respectively the author with the highest contribution in the first team listed, and the author in supervisory position in the last team listed), but less explanatory power, because authors in middle position comprise also many authors being first and last in other teams, alongside genuine "middle" authors, that is authors who have provided more limited contributions. Our data do not allow us to control directly for the number of teams behind each paper, but only for *the number of affiliations* listed on each paper. This is because for most publications in ISI WoS, authors and affiliations are listed in separate fields, with no keys to connect them; in addition, it is often the case that one author has multiple affiliations. We control our results running a set of regressions also on a restricted sample that includes only the publications with multiple affiliations. Our results do not change in any meaningful way. They are available on request.

$= (PUB\_STOCK_{tpati}-PUB\_STOCK_{tpat0j})/(PUB\_STOCK_{tpat1j}-PUB\_STOCK_{tpat0j})$

where PUB_STOCK_{1j} and PUB_STOCK_{0j} are respectively the highest and lowest PUB_STOCK values among all the authors in PPP *j*. Alternatively, we employ two dummies for the scientists with the highest and lowest scientific experience (TOP_SCHOLAR, BOTTOM_SCHOLAR).

Professional experience is informative of a scientist's skills and rank, and as such it should decrease the probability of exclusion from inventorship. In fact we expect technicians in the research team to have a smaller publication stock than other team members; accordingly if included in the authors' by-line (a potential instance of gift authorship), they have a higher probability to be excluded from the patent. At the same time, however, the authors' by-line may include scientists of great reputation, but who have not contributed much to the research (guest authorship, as discussed in section 2.2). Guest authors may be included to increase the publication's visibility, or out of deference towards important members of a department; but they can hardly claim any stake in the patent. In this case, we expect professional experience to increase the probability of exclusion. It follows that we cannot put forward strong a priori on the sign of PUB_STOCK, RELATIVE_PUB_STOCK and TOP/BOTTOM_.

- Authors' gender, as represented by the dummy variable WOMAN. Following the discussion in Section 3 we expect a positive effect of WOMAN on the probability of exclusion. However, gender and contribution may be correlated, to the extent that women authors who appear in the MIDDLE position of a publication by-line may be more peripheral team members, and have contributed less to the research results than other authors in the same position. The same does not apply to women in FIRST and LAST positions, since such positions can be assigned to one author only and provide non-ambiguous information. Thus, we will also interact gender and contribution dummies. We expect the coefficients for FIRST*WOMAN and LAST*WOMAN to be greater than the coefficients for, respectively, FIRST_MAN and LAST_MAN, and we can safely interpret the difference as entirely due to gender. We also expect the coefficients for MIDDLE*WOMAN to be positive, but this may be due to either contribution or gender.

As for the characteristics of each PPP, we control for:

- The number of authors of the publication in PPP *j* (N_AUT_j): the larger the team of scientists, the higher the probability that some authors will be excluded, due to dilution of contributions.
- The academic inventor's discipline (dummies for ELECTRONICS, PHARMACOLOGY, BIOLOGY and CHEMISTRY), which we presume to be the same as that of co-authors.[18]

[18] As an alternative, we experimented with journal dummies, also because journals may differ in their tolerance of authorship inflation. The results did not change at all, so we do not report them, but they will are available on request.

- The difference between the publication year and the priority year of the patent ($DELTA\_YEAR_j = t_{pubj} - t_{patj}$), which controls for the accuracy of our matching exercise, and reflect the scientists' patenting strategies (see discussion in section 5.3, below)
- Time dummies for the priority years of patents, which capture any change over time in the practice of listing inventors in patents or authors in publications.

5. Results

The database that results from the different steps described in the previous section is composed of 680 PPs and 3333 observations. Clearly the same publications and patents may belong to different PPPs and each scientist may enter the sample more than once if she has more than one publication, and/or these are related to more than one patent.[19]

5.1 Descriptive Statistics

Table 1 reports the number of patents in the selected (*actual*) PPP sample, by priority year and technological field. Table 2 reports the number of publications by year and technological fields, Figure 1 shows the observed frequencies of the number of authors and inventors in each of the 450 individual publications and 213 individual patents in the PPP sample. The distribution of the number of authors has a fatter tail to the right.

[Table 1 and 2 about here]

Table 3 shows that the average number of inventors per PPP is 3.35, while the average number of authors is equal to 4.9, for a resulting difference of 1.54. Table 3 also reports similar information for the initial set of original (*potential*) PPPs: notice that, due to a much less precise matching between patents and publications, the average values of the number of authors and inventors are higher than in the *actual* PPP sample, as it is the average difference between the two (4.89 instead of 1.54).

[TABLE 3 here]

[FIGURE 1 here]

[19] If scientist *i* is the author of two publications, both related to the same patent B, he/she will enter our database twice; if scientist *i* is the author of two publications, both of them related to patents A and B, he/she will enter our database 4 times; if scientist *i* is the author of one publication related to just one patent, he/she will enter our database just once; the latter is the most common case that covers 32.3% of the number of observations.

These results confirm the existence of an exclusion process[20]. Significant differences, however, exist across disciplines. Table 3 shows that the average author-inventor difference is significantly greater than zero only in *Biology* and *Pharmacology*. In Chemical Engineering & Material Technology and in Electronics & Telecommunications we find that the average number of authors and inventors are roughly the same, and the median value of the difference across PPPs is equal to 0.

[TABLE 4 here]

[TABLE 5 here]

In order to investigate whether a specific pattern of exclusion emerges, in Table 4 we report the number of publications by number of authors, and calculate the number of authors in each position of the by-line (FIRST, LAST and MEDIUM).

Table 5 reports similar information, but it distinguishes between authors who have been included and excluded from the PPP-related patents. It shows that authors in the LAST position have the lowest frequency of exclusion, followed by those in FIRST. Authors in MIDDLE positions are more often excluded. When considering the four disciplinary fields separately, we do not detect any significant difference across fields.[21]

5.2 Estimation results

The sample we use for the estimation is built as follows. Starting from the selected 680 PPPs we exclude: (1) all publications with only one author; (2) all the publications whose author by-line is in alphabetical order and with a number of authors greater or equal to the number of inventors; (3) all the academic inventors from the KEINS database for which the probability of being excluded is zero by construction;[22] and (4) two publications whose number of authors made the data collection effort daunting (36 and 42 authors, respectively). This leaves us with 476 patent-publication pairs, 186 patents, 326 publications and 929 authors (540 men, 330 women, and 59 for whom gender is unknown). The resulting sample contains 1997 observations (1897 of which have non-missing gender information).

[20] For all many-to-many PPPs, we also counted the total number of authors and inventors and checked whether an exclusion pattern at the group level could be detected. The results we obtained are very close to those of Figure 1 and Table 3: this means that even when the same publication is related to more than one patent, it often happens that one or more co-authors are excluded from all patents.

[21] Results available on request.

[22] The academic inventors from KEINS are excluded from the regression sample only when they serve as a starting point for the PPP's construction. Conversely they are kept in the sample if they appear as co-authors in other publications and are not excluded from the related patent.

Table 6 provides the summary statistics[23]. The overall percentage of exclusions in our sample is 83%. Notice that women account for 37% of the observations and most of them have a middle position in the by-line (no position or exclusion pattern has been detected for observations with missing gender value).

[TABLE 6 here]

The correlation matrix between the main variables is displayed in the Appendix (Table A2). The dependent variable exhibits all the expected correlations with the covariates. The correlation between measures of seniority and experience is high, as it is the correlation between the absolute and relative measures of each variable. Finally, SENIORITY and PUB_STOCK are correlated with FIRST and LAST, respectively with a negative and a positive sign: this suggests that first authors are likely to be junior team partners, and last authors more senior ones.

Table 7 displays the results of a set of Logit regressions where the dependent variable is the probability of an author's exclusion from a related patent. We assume that observations are independent across individuals, but not necessarily across publications and patents by the same individual scientists. We include dummies for the calendar year and for the disciplinary field.

Column (1) reports the basic regression. In columns (2) and (3) we substitute controls for the authors' seniority and scientific experience with similar controls, but *relative* to the other co-authors, either as continuous variables or dummies. Column (4)-(5) and (6) replicate column (1), (2) and (3), but with the addition of a control for gender (WOMAN). Column (7) also controls for gender, interacted with the information on the author's position in the by-line (MIDDLE*MAN is the reference case).

[TABLE 7 and 8 here]

Our results show that both first and last authors have a significantly lower probability of being excluded from inventorship than middle authors. This result holds across all specifications in Table 7. Also first authors are less likely to be excluded than last ones. In Table 8 we calculate the changes in the predicted probability of exclusion for a discrete change in FIRST and LAST (with all other variables

[23] There are 17 observations related to 13 publications with only two authors. We kept these observations in the sample. Their exclusion does not change the econometric results in any respect.

held at their mean value), based upon regression (1) in Table 7: we obtain values equal to -0.16 and -0.12, respectively[24].

Assuming that first authors have contributed the most, and most creatively, to the publication, these results suggest that their lower probability of exclusion reflects the rule of law. The same explanation is consistent with the assumption that last authors contribute to the research effort more and more creatively than middle authors, but less than the first ones, so that their probability of exclusion is lower than the former and higher than the latter.

Table 7 also shows that the probability of exclusion decreases significantly with the scientist's years of activity, as expected from our discussion of negotiation. In specifications (1) and (4), the estimated coefficient of SENIORITY is negative and significantly different from zero; the same applies to RELATIVE_SENIORITY, in columns (3) and (6). These results are confirmed when we use relative measures of seniority and experience with variables MOST_JUNIOR and RELATIVE_SENIORITY significantly negative.

In addition, PUB_STOCK is slightly positive and RELATIVE_PUB_STOCK is positive and significant given the same level of seniority, which means that the scientists with the larger stock of publication in the team have a relatively higher probability of exclusion from inventorship. According to our discussion in section 3, we interpret it as evidence of guest authorship practices involving well-reputed scientists (with guest authors more likely to be excluded from the patent).

Table 9 reports the predicted probabilities of exclusion based upon regression (1) in Table 7, for different levels of SENIORITY. The analysis of the marginal effect of SENIORITY *for individuals who are first in the by-line* shows that the first ten years of activity decrease the probability of exclusion by approximately 0.13. The same analysis *for individuals who are last in the by-line* suggests that the same increase in seniority decreases the probability of exclusion by approximately 0.14. The following ten years of activity (that is, from the 10^{th} to the 20^{th}) reduce the probability of exclusion of first and last authors respectively by 0.20 and 0.23.

[TABLE 9 here]

[FIGURES 2 and 3 here]

These results indicate that, given the same contribution to the publication (position in the by-line), a junior scientist is significantly more at risk of being excluded from the patent than a senior one.

[24] These values are similar to the marginal effects derived from estimating the same specification with a linear probability model, that fully confirms results shown in Table 7.

Among authors who are first in the by-line, a 10-year increase in seniority gives a substantial premium in terms of reduced probability of exclusion. Last authors also benefit greatly from seniority: a 10-year increase in publication activity provides them with a substantial premium in terms of reduced probability of exclusion.

Regressions (4)-(6) in Table 7 show that women are significantly more at risk of being excluded from the patent than men. Depending on the specification, the estimated coefficient for WOMAN range between 0.72 and 0.84.

When we interact gender and the position in the by-line (regression (7) in table 7) we find that women in MIDDLE position are more at risk of exclusion than men in the same position. In this case, it may be that MIDDLE-placed women are excluded because they contributed less than men in the same position. However, we also find that women in FIRST and LAST positions have a higher probability of exclusion than men in the same positions, which suggests that gender effects is independent from the individual contribution.

Following a referee's request, we have also considered a specification with interaction effects between the author's position and both seniority and the disciplinary dummies. We estimated the effects according to Ai and Norton (2003) and Norton et al (2004) and found them never significant for position interacted with seniority, and significantly positive for the interaction between FIRST and PHARMA and between LAST and ELECTRONICS. This means that in these two cases the results displayed in Table 7 are somewhat weaker.[25]

5.3 Robustness

In section 3, we remarked that the high number of one-to-many and many-to-many patent-publication matches suggest that negotiations within a team may refer to an entire set of related publications and patents, and not just over one item at a time. Therefore, we have performed a subsidiary exercise in which the exclusion event concerns the whole set of patents matched to one single publication in the PPP.

Table 10 reports the results of a set of regressions, identical to the ones in Table 7, in which $\Pr(y_{ij}=1|x)$ is the conditional probability that author i is excluded, not just from one patent, but from all the patents related to his/her publication (that is, j does not represent one of the patents related to i's publication, but the entire set of patents related to it; the set of explanatory variables x does not change).

[25] Full results are available on request. Notice that Ai and Norton (2003) and Norton et al. (2004) show that in logit models interaction effects are present also when the coefficient of the interaction terms is assumed to be zero.

[TABLE 10 here]

The sign and significance of the estimated parameters for FIRST does not change, although their magnitude decreases. Also, the estimated parameter for LAST maintains its sign. All estimated coefficients for seniority and experience, both absolute and relative, maintain their sign and significance, and increase slightly in absolute terms. Only the control variable DELTA_YEARS becomes not significant. Also the gender effect is very strong, its the estimated coefficient being larger than in the previous case. When we interact gender and position in the by-line, the coefficients maintain their sign and significance (with the exclusion of LAST*MAN). We conclude that the core of our results remains unchanged when we alter our definition of "exclusion from inventorship". A second possible cause of concern is the potential mix of both false positives (unrelated patents and publications in same PPP) and false negatives (true PPPs we failed to identify). In particular, false positives could produce a positive bias of the estimated coefficients of LAST, as well as variables related to seniority and professional experience. This is because typically senior and more productive authors (who, as we have seen, are more likely to sit in LAST position in by-lines) sign more papers than junior scientists.

In order to control for this potential problem, we restrict our sample to the PPPs with an S similarity score in the top 5% of the distribution. This reduces the risk of false positives and leaves us with only 341 PPPs, with a minimal value of S equal to 0.174. We then run a set of regressions identical to those of Table 7. Our results (Table 11) confirm the negative sign of SENIORITY (or, alternatively, of RELATIVE_SENIORITY), with a slight increase in the absolute value of the estimated coefficient. In addition, the estimated coefficients FIRST and LAST maintain their sign and significance, with the estimated effect of LAST being larger. We also find a stronger positive effect of PUB_STOCK on the probability of exclusion. Finally, the estimated gender effects are coherent to what found in Table 7.[26]

[TABLE 11 here]

We also consider a different way to restrict our sample of PPPs, which consists in selecting only the publications appearing *after* the priority date of the related patents, for which the variable DELTA_YEARS takes a null or positive value. The rationale behind this restriction is that research teams, especially if well advised by their TTO, are more likely to publish their papers after filing the

[26] If we raise the bar further, and select only the PPPs whose similarity scores fall within the top 1% of the distribution (minimal level of S at 0.25), we still obtain similar results. In this case we are left with 68 PPPs and 156 observations in the regression sample. In particular, estimated parameters from the Logit regression for FIRST, LAST and SENIORITY become, respectively, -2.25***, -2.35*** and -0.19***. The complete results are not displayed but are available on request.

patent, in order to avoid endangering its novelty. So, we suspect that PPPs where DELTA_YEARS<0 include more false positive than those for which DELTA_YEARS≥0.

Table 12 replicates the Logit regression of column (1) of Table 7, for two different PPP samples, one for observations with DELTA_YEARS≥0, the other one for the complementary set of observations (DELTA_YEARS<0).

[TABLES 12 here]

The results for DELTA_YEARS≥0 are similar, in terms of sign and significance of the estimated parameters, to those of Table 7. The main difference consists only in the magnitude of FIRST and LAST parameters, which is respectively lower and higher than in Table 7 (the SENIORITY parameters also appear smaller). By contrast, the regression for DELTA_YEARS<0 returns a very high coefficient for FIRST and a non-significant one for LAST. which is consistent with the possibility that part of our results in Table 7 are affected by a bias due to the methodology followed for the creation of our PPP sample.

Alternatively, we can explain the results of Table 12, by recalling the possibility that, within a team of scientists, the decision to file a patent may follow two different routes, which affect differently the distribution of inventorship credits. Patents in PPPs with DELTA_YEARS≥0 may be the result of a route which included searching for IPR protection from its very beginning of the research, so that precautions were taken, including not publishing any research result before filing the patent application. Conversely, patents in PPPs with DELTA_YEARS<0 may be the result of a decision taken after finding promising scientific results. In this case, the patent generates specific additional activity of the author who has contributed most to the research activity, namely the first author. This interpretation is consistent with the very high absolute value of the coefficient for FIRST (as opposed to the lack of significance for LAST) in the case of DELTA_YEARS<0. Note, however, that the observations with DELTA_YEARS<0 account for just one third of the sample.[27]

6. Conclusions

In this paper, we have investigated the weight of social norms and legal rules on the determinants of authorship and inventorship attribution within scientific teams, whose research results are diffused through patent-publication pairs (PPPs). We have proposed that ambiguities of legal norms

[27] This interpretation is coherent with findings by Breschi et al. (2008) and Azoulay et al. (2007) on the time sequencing of patents and publications by academic inventors.

and editorial guidelines put team members in the condition of negotiating over attribution. While the negotiation does necessarily take into account each team member's contribution to the research project, attribution rights may be traded on the basis of their economic value and relative bargaining power. Our empirical analysis has confirmed most of the conclusions derived by our theory.

First, we found that the number of authors in a PPP is usually larger than the number of inventors, and that this is due to the exclusion of several authors of a publication, or set of related publications, from the related patent or set of patents.

Second we found that a scientist's contribution affects the probability of her exclusion from the patent, as suggested by the legal norms, albeit scant and confused, on inventorship attribution. Our results are also compatible with the existence of "guest" or "gift" authorship, as suggested by the vast literature on authorship attribution.

Third, we have provided evidence that the allocation of attribution rights depends also seniority and gender. Ceteris paribus, junior and women authors of a scientific publication are more likely to be excluded from a related patent. Our main rationale for this result is that junior and women scientists may be more willing to trade inventorship for authorship, if the latter is at stake, due to its higher economic value in terms of career prospects. But we cannot exclude arbitrariness in decisions taken by senior team members.

Although within-the-team negotiations may be conducted in the best interest of all members (that is, with no significant welfare loss for any of them), welfare losses for society at large may yet arise. This is because attribution rights provide information to several third parties, which ought to be revealing of the individuals' capabilities as well as of personal responsibilities in cases of misconduct (Lacetera and Zirulia, 2011). Future research ought to be directed at quantifying these losses.

Our results contribute to current criticism of the economic value of the concept of scientific authorship, and extend it to that of inventorship. Despite the dramatic rise of teamwork, scientific attribution rights are still modelled upon a view of discovery and invention as resulting from an individual's spark of genius. In contrast, other fields of human creativity have abandoned such individualistic bias (Fisk, 2006). Some steps in a similar direction have been undertaken by several scientific journals, especially in the medical sciences, which now require scientists not merely to identify themselves as "authors", but also to specify the exact contents of their contribution according to pre-determined categories. "Contributorship" is suggested as an alternative to authorship (Rennie, 1998; Biagioli et al., 1999; Hwang, 2003).

The legal figure of the inventor is also an obsolete one that dates back to a time – the XIX century – when the existence of patents had been put into question, and was defended by the creation of a public image of inventors as "heroes of the industrial revolution" (MacLeod, 2008; see also Machlup and Penrose, 1950; Long, 1991; and Bracha, 2005). In that respect, our work can be considered not only as an exploration in the field of academic patenting, but also as a first step in the direction of investigating the overall adequacy of present norms of inventorship attribution.

Besides these substantive results, our paper contribute to the emerging technical literature on PPPs by proving the usefulness of text-mining techniques for matching patents and publications. Our application suggests that complex combinations of patents and publications are likely: one-to-one matches between individual patents and publications are rarer than matches of several publications connected to a single patent or several patents.

As for immediate extensions of our work, it would be of great interest to explore differences in attribution practices across academic institutions and countries. The existence of cross-countries differences in authorship attribution is suggested by Hwang et al. (2002), who find that US scientists are more likely to comply with the ICMJE authorship guidelines than non-US ones. Such differences may extend to inventorship. It would be also interesting to study the behaviour of scientists that publish end patent employed in private companies (Lacetera and Zirulia, 2011).

More broadly, it would be also of interest to address other fields of human creativity, in which – like in science – results are achieved by teams, but careers are built upon personal reputation. In some of these fields, various forms of contributorship have emerged to fine-tune the information signals resulting from attribution. In movie-making, for example, the various professional figures are awarded specialized credits (for directing, screenwriting, shooting etc.); this does not prevent the existence of some prestige ranking (as with directors vs. more technical figures), but it allows due credit to be distributed to all participants in the creative act. On the contrary, in fields such as R&D, design, architecture, or advertising, authorship or inventorship still are the key form of attribution, so we may be interested to investigate whether negotiation among team members occurs, as in science; and what characteristics of the individuals affect its outcome.

References

Ai, C., Norton, E.C., (2003), "Interaction terms in logit and probit models," *Economics Letters* 80(1), pp. 123-129

Audretsch, D.B., Stephan, P.E. (1999), "Knowledge spillovers in biotechnology: Sources and incentives", *Journal of Evolutionary Economics* 9, 97–107.

Azoulay P., Ding W., Stuart T. (2007), "The determinants of faculty patenting behavior: Demographics or opportunities?", *Journal of Economic Behavior & Organization* 63/4, pp.573-576

Bassecoulard E., Zitt, M. (2004), "Patents and Publications. The Lexical Connection", in: Moed H.F., Glänzel W., Schmoch U., *Handbook of Quantitative Science and Technology Research*, Kluwer. Dordrecht, Ch. 30.

Bates T., Anić A., Marušić, M., Marušić A. (2004), "Authorship Criteria and Disclosure of Contributions" *Journal of American Medical Association* 292(1), pp.86-88

Bennett V.C., Biswas S.J. (1997), "Protecting the patentability of your collaborative research", *Nature Biotechnology* 15, pp. 472-473

Biagioli M. (1998), "The Instability of Authorship: Credit and Responsibility in Contemporary Biomedicine", *FASEB Journal* 12, pp.3-16

Biagioli M., Crane J., Derish P., Gruber M., Rennie D., Horton R. (1999), *Authorship Task Force White Paper*, Council of Science Editors (http://www.councilscienceeditors.org/services/atf_whitepaper.cfm, last accessed: May 2008)

Biagioli M., Galison P. (2002), *Scientific Authorship: Credit and Intellectual Property in Science*, Routledge

Bracha O. (2005), *Owning Ideas: A History of Anglo-American Intellectual Property (Ch.4: United States Patents)*, S.J.D. Dissertation, Harvard Law School

Breschi S., Lissoni F., Malerba F. (2003) "Knowledge Relatedness in Firm Technological Diversification", *Research Policy* 32/1, pp.69-87

Breschi S., Lissoni F., Montobbio F. (2007), The scientific productivity of academic inventors: new evidence from Italian data, *Economics of Innovation and New Technology* 16/ 2, pp.101-118

Breschi S., Lissoni F., Montobbio F. (2008). University patenting and scientific productivity. A quantitative study of Italian academic inventors. *European Management Review* 5(2): 91-109

Breschi, S., Lissoni F., Montobbio F. (2005). "From publishing to patenting: Do productive scientists turn into academic inventors?" *Revue d'Economie Industrielle* 110 (2), 75–102.

CIVR (2006), *Valutazione triennale della ricerca 2001-2003. Relazione Finale*, Comitato d'Indirizzo per la Valutazione della Ricerca, Roma (http://vtr2006.cineca.it/index_EN.html; last access: May 2008)

Colyvas J.A. (2007), "From divergent meanings to common practices: The early institutionalization of technology transfer in the life sciences at Stanford University", *Research Policy* 36, pp. 456-76

Dasgupta P., David P.A. (1994), "Toward a new economics of science", *Research Policy* 23, pp.487-521

Ding, W., Murray F., Stuart T. E. (2006), "Gender difference in patenting in the academic life science" *Science* 313, pp. 665–667

Djerassi C. (1989), *Cantor's Dilemma*, Penguin Books, London

Drenth JP. (1998), "Multiple authorship: the contribution of senior authors", *Journal of the American Medical Association* 280, pp. 219-21

Ducor P. (2000), "Coauthorship and Coinventorship", *Science* 289, pp.873-875

Engelsman E. C., van Raan A.F.J. (1992), "A patent-based cartography of technology", *Research Policy* 23, pp. 1-26.

Fasse W.F. (1992), "The Muddy Metaphysics of Joint Inventorship: Cleaning Up after the 1984 Amendments to 35 U.S.C. § 116", *Harvard Journal of Law and Technology* 5, pp.73-74

Fernandez-Molina J.C., Pais E. (2001), "The Moral Rights of Authors in the Age of Digital Information", *Journal of the American Society for Information Science and Technology* 52/2, pp. 109-117

Fisk C.L. (2006), "Credit Where It's Due: The Law and Norms of Attribution", *Georgetown Law Journal* 95/1, pp.49-118

Gans J. S., Murray F E., Stern S. (2010), "Contracting Over the Disclosure of Scientific Knowledge: Intellectual Property Protection and Academic Publication", *Working Paper* (http://ssrn.com/abstract=1559871; February 26)

Garfield E. (1983), "Carl Djerassi: Chemist and Entrepreneur", *Chemtech* 13, pp. 534-538

Hansmann H., Santilli M. (1997), "Authors' and Artists' Moral Rights: A Comparative Legal and Economic Analysis", *Journal of Legal Studies* 26(1), pp.95-143

Hoen W.P., Henk C.W., Overbeke A.J.P.M. (1998), "What Are the Factors Determining Authorship and the Order of the Authors' Names?", *Journal of American Medical Association* 280, pp. 217-218

Hutchins M. (2003), "Common mistakes that undermine patent protection and how to avoid them", *International Journal of Medical Marketing* 3, pp. 204-211

Hwang S.S. et al. (2003), "Researcher Contributions and Fulfillment of ICMJE Authorship Criteria: Analysis of Author Contribution Lists in Research Articles with Multiple Authors", *Radiology* 22, pp.16-23

ICMJE (2007), *Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Writing and Editing for Biomedical Publication*, International Committee of Medical Journal Editors (http://www.icmje.org)

IEEE (2008), *IEEE Publication Services and Products Board Operations Manual (revised version)*, (http://www.ieee.org/portal/cms_docs_iportals/iportals/publications/PSPB/opsmanual.pdf)

Jaffe A.B., Trajtenberg M., Fogarty M.S. (2000) "Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors", *American Economic Review* 90/ 2, pp. 215-218

Johnson J.C., Orback M.K. (2002), "Perceiving the political landscape: ego biases in cognitive political networks", *Social Networks* 24, pp.291-310

Jones B.F., Wuchty S., Uzzi B. (2008), "Multi-University Research Teams: Shifting Impact, Geography, and Stratification in Science", *Science* 322, pp. 1259-1262

Jones B.M. (2009), "The Burden of Knowledge and the 'Death of the Renaissance Man': Is Innovation Getting Harder?" *Review of Economic Studies* 76(1), pp. 283-317

Katz J.S., Martin B.R. (1997), "What is research collaboration?", *Research Policy* 26, pp. 1-18

Klavans R., Boyack K.W. (2006), "Identifying a Better Measure of Relatedness for Mapping Science", *Journal of the American Society for Information Science and Technology* 57/2, pp.251-263.

Lacetera N., Zirulia L. (2010) Individual Preferences, Organization, and Competition in a Model of R&D Incentive Provision, DSE Working Paper 624, University of Bologna.

Lacetera N., Zirulia L. (2011); The Economics of Scientific Misconduct. *Journal of Law, Economics and Organization,* forthcoming.

Leopold E., May M., Paaß (2004), "Data Mining and Text Mining for Science & Technology Research", in: Moed H.F., Glänzel W., Schmoch U. (eds.), *Handbook of Quantitative Science and Technology Research.* Kluwer, Dordrecht

Link A.N., Siegel D.S., Bozeman, B. (2007), "An empirical analysis of the propensity of academics to engage in informal university technology transfer", *Industrial and Corporate Change* 16, pp. 641–655

Lissoni F., Mairesse J., Montobbio F., Pezzoni M. (2011) "Scientific Productivity and Academic Promotion: A Study on French and Italian Physicists", *Industrial and Corporate Change* (forthcoming)

Lissoni F., Llerena P., McKelvey M., Sanditov B. (2008), "Academic Patenting in Europe: New Evidence from the KEINS Database", *Research Evaluation* (forthcoming)

Lissoni F., Sanditov B., Tarasconi G. (2006), "The Keins Database on Academic Inventors: Methodology and Contents", *CESPRI Working Paper* 181, Bocconi University

Long P.O. (1991), "Invention, Authorship, 'Intellectual Property', and the Origin of Patents: Notes toward a Conceptual History", *Technology and Culture* 32/4, pp.846-884

Machlup F., Penrose E. (1950), "The Patent Controversy in the Nineteenth Century", *Journal of Economic History* 10/1, pp. 1-29

MacLeod C. (2008), *Heroes of Invention: Technology, Liberalism and British Identity, 1750-1914*, Cambridge University Press

McSherry C. (2003), *Who Owns Academic Work*, Harvard University Press, Cambridge MA

Merton R.K. (1957), *Social Theory and Social Structure*, Free Press, Glencoe, Ill.

Meyer M., Bhattacharya S. (2004), "Commonalities and differences between scholarly and technical collaboration. An exploration of co-invention and co-authorship analyses", *Scientometrics* 61, pp. 443-456

Mowatt, G., Shirran, L., Grimshaw J.M., Rennie D., Flanagin A., Yank V., MacLennan G., Gotzsche P.C., Bero L.A. (2002), "Prevalence of Honorary and Ghost Authorship in Cochrane Reviews", *Journal of the American Medical Association* 287, pp.2769-2771

Murray F. (2002) "Innovation as co-evolution of scientific and technological networks: exploring tissue engineering", *Research Policy* 31, pp. 1389–1403

Murray F. (2004), "The role of academic inventors in entrepreneurial firms: sharing the laboratory life", *Research Policy* 33, pp. 643-659

Murray F., Graham L. (2007), "Buying science and selling science: gender differences in the market for commercial science", Industrial and Corporate Change 16, pp. 657 - 689

Murray F., Stern S. (2007) "Do formal intellectual property rights hinder the free flow of scientific knowledge? An empirical test of the anti-commons hypothesis", *Journal of Economic Behavior & Organization* 63/4, pp.648-687

Murray F., Stern S. (2008) "Learning to Live with Patents. A Dynamic Model of Knowledge Community's Response to Legal Institutional Change",MIT Sloan School Working Paper

Murray F.E. (2004) "The role of academic inventors in entrepreneurial firms: sharing the laboratory life", *Research Policy* 33(4), pp. 643-659

Nelson R.R., Romer P.M. (1996), "Science, Economic Growth, and Public Policy", *Challenge* March-April, pp. 9-21

Norton E.C., Wang H., Ai C., (2004). "Computing interaction effects and standard errors in logit and probit models" *Stata Journal* 4(2), pp. 154-167

OECD (2003), *Turning Science into Business: Patenting and Licensing at Public Research Organisations*, Organization for Economic Co-operation and Development, Paris

RAE (2008), *Research Assessment Exercise*, Higher Education Funding Council for England (http://www.rae.ac.uk/; last accessed; May 2008)

Rennie D. (1998), "Freedom and Responsibility in Medical Publication: Setting the Balance Right", *Journal of American Medical Association* 280, pp.300-302

Rennie D., Flanagin A. (1994), "Authorship! Authorship! Guests, ghosts, grafters, and the two-sided coin", *Journal of American Medical Association* 271, pp. 469-471.

Salton G., McGill M.J. (1983), Introduction to Modern Information Retrieval, McGrawHill, New York

Seymore S.B. (2006), "My Patent, Your Patent, or Our Patent? Inventorship Disputes within Academic Research Groups", *Albany Law Journal of Science and Technology* 16, pp.125-167

Stephan P.E. (2010), "The Economics of Science", in: Hall B.H., Rosenberg N. (eds.), *Handbook of Economics of Innovation,* North Holland/Elsevier

Stephan P.E., Levin, S. (1992). *Striking the Mother Lode in Science: The Importance of Age, Place, and Time.* Oxford University Press, New York

Stephan P.E., Gurmu, S., Sumell, A., Black, G. (2007), "Who's patenting in the university? Evidence from the survey of doctorate recipients". *Economics of Innovation and New Technology* 16(2), pp. 71–99.

UNESCO (2001), *A Guide to Human Rights. Institutions, Standards, Procedures*, United Nations Educational, Scientific and Cultural Organization, Paris

Van den Steen E. (2004), "Rational Overoptimism (and Other Biases)", *American Economic Review* 94/4, pp. 1141-1151

Vinarov S.D. (2003), "Patent protection for structural genomics-related inventions", *Journal of Structural and Functional Genomics* 4, pp. 191-209

Whittington K.B., Smith-Doerr L. (2008), "Women Inventors in Context: Disparities in Patenting across Academia and Industry", *Gender & Society* 22/2, pp. 194-218

WIPO (2008), *Berne Convention for the Protection of Literary and Artistic Works*, World Intellectual Property Organization, Geneva (http://www.wipo.int/treaties/en/ip/berne/; last accessed: May 2008)

Wuchty S., Jones B.F., Uzzi B. (2007), "The Increasing Dominance of Teams in Production of Knowledge", *Science* 316, pp. 1036-1039

Zuckerman H.A. (1968), "Patterns of Name Ordering among Authors of Scientific Papers: A Study of Social Symbolism and Its Ambiguity", *American Journal of Sociology* 74/3, pp.276-291

TABLES

Table 1. Number of patents by field and priority years in the selected PPPs

	Chemistry	*Electronics*	*Pharma*	*Biology*	*Total*
1988	0	1	0	0	1
1989	2	6	0	1	9
1990	1	6	3	4	14
1991	2	2	6	4	14
1992	2	10	6	4	22
1993	0	4	3	2	9
1994	2	10	1	9	22
1995	1	13	3	7	24
1996	2	9	4	7	22
1997	3	16	12	7	38
1998	1	16	7	3	27
1999	1	6	0	1	8
2000	0	3	0	0	3
Total	17	102	45	49	213

Table 2. Number of publications by publication year in the selected PPPs

	Chemistry	*Electronics*	*Pharma*	*Biology*	*Total*
1990	0	2	0	1	3
1991	4	9	8	18	39
1992	3	10	10	10	33
1993	1	8	8	16	33
1994	3	11	6	22	42
1995	4	22	3	22	51
1996	4	25	19	23	71
1997	1	27	7	27	62
1998	4	26	10	20	60
1999	1	24	6	6	37
2000	0	12	1	3	16
2001	0	3	0	0	3
Total	25	179	78	168	450

Table 3. Summary statistics on the number of authors and number of inventors for each potential and selected patent-publication pairs, total samples and by scientists' fields

	No. of author (a)	No. of Inventor (b)	(a)-(b)
Selected PPPs			
Obs. (No. of PPPs)	680	680	
Mean	4.90	3.35	1.54
Median	4	3	1
St. dev.	2.67	2.50	
Min	1	1	
Max	19	21	
Potential PPPs			
Obs. (No. of PPPs)	6810	6810	
Mean	8.51	3.62	4.89
Median	5	3	2
St. dev.	1.41	3.53	
Min	1	1	
Max	517	21	
Pharmacology (selected PPP)			
Obs. (No. of PPPs)	104	104	
Mean	6.46	3.60	2.86
Median	6	3	3
St. dev.	2.71	2.01	
Min	2	1	
Max	14	10	
Biology (selected PPP)			
Obs. (No. of PPPs)	222	222	
Mean	5.94	3.55	2.39
Median	6	3	3
St. dev.	2.51	3.91	
Min	2	1	
Max	13	21	
Chemical Eng. & Materials Tech. (selected PPP)			
Obs. (No. of PPPs)	27	27	
Mean	4.48	4.78	-0.30
Median	4	4	0
St. dev.	1.60	2.10	
Min	2	2	
Max	8	11	
Electronics and Telecom (selected PPP)			
Obs. (No. of PPPs)	327	327	
Mean	3.73	3.03	0.70
Median	3	3	0
St. dev.	2.27	1.12	
Min	1	1	
Max	19	6	

Tab. 4. Number of authors, by number of authors in each publication and position in the by-line

Nr of authors in the publication	Number of publications	Nr of authors by position: FIRST	MIDDLE	LAST
1	4	4		
2	78	78		78
3	167	167	167	167
4	138	138	276	138
5	80	80	240	80
6	66	66	264	66
7	52	52	260	52
8	25	25	150	25
9	26	26	182	26
10	9	9	72	9
11	14	14	126	14
>11	21	21	240	21
Total:	680	680	1977	676

Table 5. Count of exclusions (and non exclusions) from inventorship. by position of the author in the by-line of the publication in the PPP

	FIRST		MIDDLE		LAST		
Position in the by-line	Non excluded	Excluded	Non excluded	Excluded	Non excluded	Excluded	Total
1	336	344					680
2			215	383	60	18	676
3			122	309	103	64	598
4			73	220	78	60	431
5			49	164	42	38	293
6			22	125	36	30	213
7			18	77	24	28	147
8			10	60	8	17	95
9			8	36	17	9	70
10			5	30	4	5	44
11			3	18	9	5	35
>11			2	28	6	15	51
Total	336	344	527	1450	387	289	3333

Table 6. Summary statistics for the regression sample.

Variable	Obs.	Mean	Std. Dev.	Min	Max
Y	1997	.83	.37	0	1
FIRST	1997	.17	.37	0	1
MIDDLE	1997	.70	.46	0	1
LAST	1997	.14	.35	0	1
SENIORITY	1997	7.70	7.78	-2	26
PUB_STOCK(T-1)	1997	18.76	34.81	0	299
N_AUT	1997	7.21	3.36	2	19
DELTA_YEARS	1997	.48	1.30	-2	2
MOST_JUNIOR	1997	.36	.48	0	1
MOST_SENIOR	1997	.15	.36	0	1
TOP_SCHOLAR	1997	.10	.30	0	1
BOTTOM_SCHOLAR	1997	.32	.47	0	1
RELATIVE SENIORITY	1997	.39	.39	0	1
RELATIVE PUB_STOCK	1997	.24	.33	0	1
CHEMISTRY	1997	.04	.19	0	1
ELECTRONICS	1997	.24	.43	0	1
PHARMA	1997	.20	.40	0	1
BIOLOGY	1997	.52	.50	0	1
WOMAN	1897	.36	.48	0	1
FIRST*WOMAN	1897	.07	.25	0	1
MEDIUM*WOMAN	1897	.27	.45	0	1
LAST*WOMAN	1897	.02	.15	0	1

Table 7. Probability of exclusion from inventorship: Logit regressions.

Dep. Var. $=y_{ij}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)
FIRST	-1.05***	-0.95***	-1.02***	-1.04***	-0.96***	-1.03***	
	(0.28)	(0.26)	(0.27)	(0.29)	(0.27)	(0.28)	
LAST	-0.86***	-0.90***	-0.87***	-0.72***	-0.75***	-0.76***	
	(0.24)	(0.27)	(0.25)	(0.25)	(0.28)	(0.25)	
DELTA_YEARS	-0.18***	-0.18***	-0.15***	-0.18***	-0.19***	-0.16***	-0.18***
	(0.056)	(0.058)	(0.058)	(0.055)	(0.058)	(0.057)	(0.056)
N. OF AUTHORS	0.034	0.056	0.046	0.046	0.069*	0.070**	0.045
	(0.034)	(0.035)	(0.036)	(0.033)	(0.035)	(0.032)	(0.033)
SENIORITY	-0.079***			-0.074***			-0.073***
	(0.017)			(0.017)			(0.017)
PUB_STOCK(T-1)	0.0060*			0.0067*			0.0068*
	(0.0036)			(0.0036)			(0.0035)
MOST_JUNIOR		0.90***			0.92***		
		(0.34)			(0.34)		
MOST_SENIOR		-0.079			0.077		
		(0.35)			(0.34)		
TOP_SCHOLAR		0.43			0.50		
		(0.36)			(0.36)		
BOTTOM_SCHOLAR		0.54			0.47		
		(0.36)			(0.37)		
RELATIVE_SENIORITY			-1.62***			-1.59***	
			(0.39)			(0.38)	
RELATIVE_PUB_STOCK			0.75*			1.00***	
			(0.43)			(0.38)	
WOMAN				0.75**	0.84**	0.72**	
				(0.37)	(0.37)	(0.31)	
FIRST*WOMAN							-0.43
							(0.63)
FIRST*MAN							-0.90***
							(0.30)
MIDDLE*WOMAN							0.93***
							(0.30)
LAST*WOMAN							-0.13
							(0.65)
LAST*MAN							-0.65**
							(0.27)
Constant	0.30	-1.07	0.32	0.13	-1.25	0.36	0.051
	(1.61)	(1.71)	(1.69)	(1.57)	(1.60)	(1.62)	(1.53)
Observations	1997	1997	1997	1897	1897	1897	1897
Time dummies	Y	Y	Y	Y	Y	Y	Y
Field dummies	Y	Y	Y	Y	Y	Y	Y
Pseudo R-squared	0.13	0.14	0.13	0.14	0.15	0.14	0.14

Note: Robust standard errors in parentheses (clustered for individuals)
*** p<0.01. ** p<0.05. * p<0.1

Table 8. Change in predicted probabilities of exclusion from inventorship. for changes in the author's position in the by-line. as from regression (1) and (4)

	Excluded	Non Excluded
FIRST	0.73	0.27
	[0.62. 0.85]	
NOT FIRST	0.89	0.11
	[0.86. 0.91]	
Difference	-0.16	
LAST	0.76	0.24
	[0.66. 0.85]	
NOT LAST	0.88	0.12
	[0.85. 0.91]	
Difference	-0.12	
WOMAN	0.91	0.09
	[0.87. 0.96]	
MAN	0.83	0.17
	[0.78. 0.88]	
Difference	0.08	

Note. 95% confidence intervals in parenthesis

Table 9. Change in predicted probability of exclusion from inventorship. for changes in author's position in the by-line and different levels of SENIORITY, as from regression (1) in Table 7

	Last author		First author	
SENIORITY	Non Excluded	Excluded	Non Excluded	Excluded
0	0.14	0.86	0.17	0.83
5	0.21	0.79	0.23	0.77
10	0.28	0.72	0.30	0.70
15	0.36	0.64	0.40	0.60
20	0.46	0.54	0.50	0.50

Table 10. Probability of exclusion from inventorship (entire set of publication-related patents): Logit regressions.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
FIRST	-0.65**	-0.58**	-0.62**	-0.65**	-0.59**	-0.66**	
	(0.26)	(0.27)	(0.27)	(0.28)	(0.28)	(0.30)	
LAST	-0.39	-0.45	-0.39	-0.30	-0.36	-0.36	
	(0.27)	(0.29)	(0.27)	(0.27)	(0.30)	(0.27)	
DELTA_YEARS	-0.060	-0.067	-0.015	-0.051	-0.063	-0.014	-0.049
	(0.058)	(0.061)	(0.059)	(0.059)	(0.063)	(0.059)	(0.059)
N. OF AUTHORS	0.066*	0.090**	0.076*	0.077**	0.10**	0.11***	0.075**
	(0.038)	(0.040)	(0.039)	(0.038)	(0.041)	(0.038)	(0.038)
SENIORITY	-0.10***			-0.094***			-0.094***
	(0.022)			(0.022)			(0.022)
PUB_STOCK(T-1)	0.0058			0.0070*			0.0071*
	(0.0040)			(0.0040)			(0.0040)
MOST_JUNIOR		0.68*			0.64*		
		(0.35)			(0.35)		
MOST_SENIOR		-0.31			-0.10		
		(0.47)			(0.45)		
TOP_SCHOLAR		0.28			0.40		
		(0.43)			(0.43)		
BOTTOM_SCHOLAR		0.90***			0.90***		
		(0.28)			(0.29)		
RELATIVE_SENIORITY			-1.73***			-1.60***	
			(0.55)			(0.53)	
RELATIVE_PUB_STOCK			0.41			0.55	
			(0.56)			(0.53)	
WOMAN				1.01***	1.06***	0.90***	
				(0.38)	(0.37)	(0.34)	
FIRST*WOMAN							0.30
							(0.63)
FIRST*MAN							-0.57*
							(0.33)
MIDDLE*WOMAN							1.15***
							(0.35)
LAST*WOMAN							0.19
							(0.51)
LAST*MAN							-0.18
							(0.31)
Constant	0.81	-0.65	0.80	0.56	-0.92	0.20	0.52
	(1.48)	(1.77)	(1.64)	(1.43)	(1.65)	(1.58)	(1.40)
Observations	1997	1997	1997	1897	1897	1897	1897
Time dummies	Y	Y	Y	Y	Y	Y	Y
Field dummies	Y	Y	Y	Y	Y	Y	Y
Pseudo R-squared	0.11	0.11	0.096	0.14	0.14	0.14	0.14

Robust standard errors in parentheses (clustered for individuals)
*** p<0.01. ** p<0.05. * p<0.1

Table 11. Probability of exclusion from inventorship: Logit regressions. (restricted PPP sample: top 5% of the S score)

Dep. Var. $=y_{ij}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)
FIRST	-0.92***	-0.81**	-0.92***	-0.93***	-0.83**	-0.92**	
	(0.35)	(0.33)	(0.34)	(0.35)	(0.33)	(0.37)	
LAST	-1.01***	-1.04***	-1.06***	-0.90***	-0.90***	-0.96***	
	(0.29)	(0.31)	(0.29)	(0.29)	(0.31)	(0.29)	
DELTA_YEARS	-0.21***	-0.18**	-0.17**	-0.22***	-0.18**	-0.18**	-0.21***
	(0.074)	(0.075)	(0.075)	(0.074)	(0.075)	(0.076)	(0.075)
N. OF AUTHORS	-0.030	0.0013	-0.018	-0.016	0.014	0.025	-0.018
	(0.059)	(0.061)	(0.062)	(0.058)	(0.062)	(0.055)	(0.059)
SENIORITY	-0.089***			-0.081***			-0.081***
	(0.018)			(0.018)			(0.018)
PUB_STOCK(T-1)	0.010**			0.010**			0.010**
	(0.0044)			(0.0043)			(0.0043)
MOST_JUNIOR		1.32**			1.30**		
		(0.52)			(0.51)		
MOST_SENIOR		0.027			0.16		
		(0.39)			(0.38)		
TOP_SCHOLAR		0.42			0.52		
		(0.44)			(0.44)		
BOTTOM_SCHOLAR		-0.057			-0.095		
		(0.51)			(0.50)		
RELATIVE_SENIORITY			-2.11***			-2.00***	
			(0.41)			(0.40)	
RELATIVE_PUB_STOCK			1.60***			1.72***	
			(0.53)			(0.48)	
WOMAN				0.55	0.69*	0.52	
				(0.36)	(0.36)	(0.32)	
FIRST*WOMAN							-0.38
							(0.73)
FIRST*MAN							-0.91**
							(0.40)
MIDDLE*WOMAN							0.59*
							(0.32)
LAST*WOMAN							-0.55
							(0.64)
LAST*MAN							-0.85***
							(0.33)
Constant	0.31	-1.11	0.22	0.16	-1.27	0.23	0.15
	(0.85)	(0.81)	(0.86)	(0.85)	(0.82)	(1.06)	(0.87)
Observations	960	960	960	900	900	900	900
Time dummies	Y	Y	Y	Y	Y	Y	Y
Field dummies	Y	Y	Y	Y	Y	Y	Y
Pseudo R-squared	0.14	0.13	0.14	0.14	0.14	0.14	0.14

Robust standard errors in parentheses (clustered for individuals)
*** p<0.01. ** p<0.05. * p<0.1

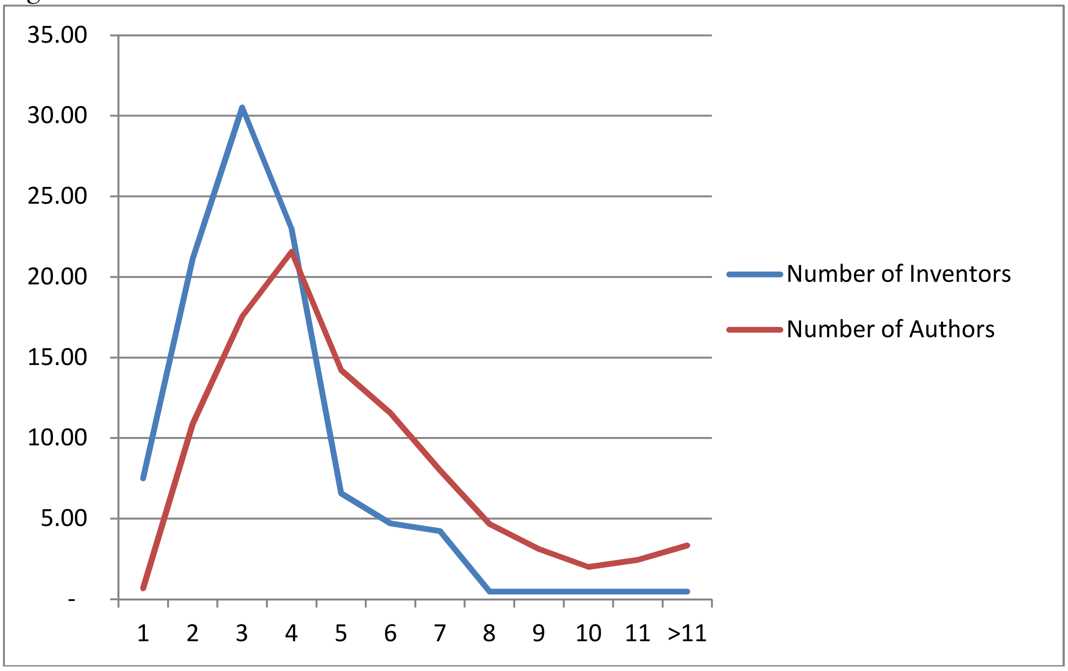
Table 12. Probability of exclusion from inventorship: Logit regressions. by values of DELTA_YEARS

Dep. Var. =y_i	DELTA_YEARS≥0 (1)	DELTA_YEARS<0 (2)	DELTA_YEARS≥0 (3)	DELTA_YEARS<0 (4)
FIRST	-0.75***	-2.43***	-0.74***	-2.47***
	(0.26)	(0.48)	(0.28)	(0.49)
LAST	-0.97***	-0.67	-0.82***	-0.53
	(0.24)	(0.57)	(0.24)	(0.57)
DELTA_YEARS	0.14*	-0.29	0.14	-0.35
	(0.086)	(0.31)	(0.089)	(0.30)
N. OF AUTHORS	0.039	0.024	0.054	0.039
	(0.034)	(0.098)	(0.034)	(0.10)
SENIORITY	-0.068***	-0.16***	-0.062***	-0.16***
	(0.017)	(0.034)	(0.017)	(0.038)
STOCK_PUB	0.0069*	0.0030	0.0073*	0.0058
	(0.0039)	(0.0044)	(0.0039)	(0.0051)
WOMAN			0.72**	1.11*
			(0.36)	(0.57)
Constant	-0.77	1.94	-0.94	0.65
	(1.48)	(1.44)	(1.47)	(2.86)
Observations	1470	527	1397	500
Time dummies	Y	Y	Y	Y
Field dummies	Y	Y	Y	Y
Pseudo R-squared	0.12	0.27	0.12	0.29

Robust standard errors in parentheses (clustered for individuals)
*** p<0.01. ** p<0.05. * p<0.1

FIGURES

Figure 1. Distribution of the number of inventors and authors



Note: This figure refers to the total sample of 450 publications and 217 patents. The maximum number of co-inventors is 19. There are 11 publications with a number of authors greater than 11.

Appendix.

Table A1 – Example of a patent-publication pair

PATENT EP1012301	*PUBLICATION* ISI:000074208600018
Title Total synthesis and functional overexpression of a $i(Candida rugosa) lip1 gene coding for a major industrial lipase 5 inventors	**Title** Design. total synthesis. and functional overexpression of the Candida rugosa lip1 gene coding for a major industrial lipase 5 authors
Abstract (extract from) The dimorphic yeast Candida rugosa has an unusual codon usage which hampers the functional expression of genes derived from this yeast in a conventional heterologous host. Lipases produced by this yeast are extensively used in industrial bioconversions. but commercial lipase samples contain several different isoforms encoded by the lip gene family. In a first laborious attempt the lip1 gene. encoding the major isoform of the C. rugosa lipases (crls). was systematically modified by site-directed mutagenesis to gain functional expression in S. cerevisiae. As an alternative approach. the gene (1688 bp) was completely synthesised with an optimised nucleotide sequence in terms of heterologous expression in yeast and simplified genetic manipulation. [...]	**Abstract (extract from)** The dimorphic yeast Candida rugosa has an unusual codon usage that hampers the functional expression of genes derived from this yeast in a conventional heterologous host. Commercial samples of C. rugosa lipase (crl) are widely used in industry. but contain several different isoforms encoded by the lip gene family. among which the isoform encoded by the gene lip1 is the most prominent. In a first laborious attempt. the lip1 gene was systematically modified by site-directed mutagenesis to gain functional expression in Saccharomyces cerevisiae. As an alternative approach. the gene (1647 bp) was completely synthesized with an optimized nucleotide sequence in terms of heterologous expression in yeast and simplified genetic manipulation. [...]

Table A2 – Correlation matrix for variables in regression exercise (significance level and n. of obs.)

	Dep. Var. $=y_{ij}$	N. OF AUTHORS	DELTA YEARS	FIRST	LAST	SENIORITY	WOMAN	PUB_STOCK (T-1)	MOST JUNIOR	MOST SENIOR	TOP SCHOLAR
N. OF AUTHORS	0.0923*	1.0000									
	0.0000	-									
	1997	1997									
DELTA YEARS	-0.0734*	0.1327*	1.0000								
	0.0010	0.0000	-								
	1997	1997	1997								
FIRST	-0.1228*	-0.2048*	0.0066	1.0000							
	0.0000	0.0000	0.7681	-							
	1997	1997	1997	1997							
LAST	-0.1118*	-0.1519*	-0.0370	-0.1777*	1.0000						
	0.0000	0.0000	0.0981	0.0000	-						
	1997	1997	1997	1997	1997						
SENIORITY	-0.1523*	-0.0290	-0.1228*	-0.0667*	0.1794*	1.0000					
	0.0000	0.1954	0.0000	0.0029	0.0000	-					
	1997	1997	1997	1997	1997	1997					
WOMAN	0.1359*	0.0767*	0.0219	0.0258	-0.1679*	-0.1430*	1.0000				
	0.0000	0.0008	0.3409	0.2622	0.0000	0.0000	-				
	1897	1897	1897	1897	1897	1897	1897				
PUB_STOCK (T-1)	-0.0543*	-0.0215	-0.0424	-0.0653*	0.2031*	0.6190*	-0.1465*	1.0000			
	0.0153	0.3377	0.0579	0.0035	0.0000	0.0000	0.0000	-			
	1997	1997	1997	1997	1997	1997	1897	1997			
MOST JUNIOR	0.1685*	-0.0632*	-0.0082	0.0271	-0.0933*	-0.6896*	0.0473*	-0.3722*	1.0000		
	0.0000	0.0048	0.7152	0.2262	0.0000	0.0000	0.0393	0.0000	-		
	1997	1997	1997	1997	1997	1997	1897	1997	1997		
MOST SENIOR	-0.0554*	-0.0638*	-0.0055	-0.0797*	0.1640*	0.5764*	-0.2210*	0.3399*	-0.3176*	1.0000	
	0.0133	0.0043	0.8066	0.0004	0.0000	0.0000	0.0000	0.0000	0.0000	-	
	1997	1997	1997	1997	1997	1997	1897	1997	1997	1997	
TOP SCHOLAR	-0.0328	-0.0898*	0.0266	-0.0644*	0.1886*	0.4069*	-0.1798*	0.6133*	-0.2428*	0.4645*	1.0000
	0.1431	0.0001	0.2356	0.0040	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-
	1997	1997	1997	1997	1997	1997	1897	1997	1997	1997	1997
BOTTOM SCHOLAR	0.1503*	-0.0130	0.1119*	0.0185	-0.0786*	-0.6275*	0.0872*	-0.3440*	0.8039*	-0.2825*	-0.2260*
	0.0000	0.5602	0.0000	0.4083	0.0004	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000
	1997	1997	1997	1997	1997	1997	1897	1997	1997	1997	1997

* $p<0.01$