Inventorship and authorship as attribution rights: An enquiry into the economics of scientific credit

Published version:

DOI:10.1016/j.jebo.2013.08.016

Journal of Economic Behavior & Organization 95 (2013) 49– 69
http://dx.doi.org/10.1016/j.jebo.2013.08.016

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 labour to do"
>
> (McSherry C., *Who Owns Academic Work?*, Harvard Univ. Press; 2003; p.84)

Francesco Lissoni[1, 4,*], Fabio Montobbio[2,4], Lorenzo Zirulia[3,4]

[1] GREThA - Université Montesquieu, Bordeaux IV (France)
[2] Dipartimento di Economia e Statistics "Cognetti de Martiis" – Università di Torino (Italy)
[3] Dipartimento di Scienze Economiche – Università di Bologna (Italy)
[4] CRIOS – Università "L. Bocconi", Milano (Italy)

* Corresponding author: francesco.lissoni@u-bordeaux4.fr

Abstract: Authorship and inventorship are "attribution rights" upon which individual scientists build their reputation and career. Social and legal norms concerning their distribution within research teams are currently criticized for failing to inform third parties on individual contributions. We examine the case of teams engaged in the "double disclosure" of their research results through both publications and patents, and model the negotiation process taking place between junior or female team members and the senior (male) one. We suggest that the former may give up inventorship in order to secure authorship, even when entitled to the both. Based on a sample of 680 "patent-publication pairs" (related sets of patents and publications) we show that, very frequently, one or more authors of a publication do not appear as inventors of a related patent. This is less likely to happen for first and last authors, which is in accordance both with our model and the prevailing legal norms on inventorship. However, the probability of exclusion from inventorship also declines with seniority, and increases for women, which is compatible with our model only.

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

JEL: Codes: O31, O34, L30

1. Introduction

Understanding how scientific knowledge is produced and reduced to practice is a central theme of today's economic research. Both the sociology and the economics of science pay a great deal of attention to the system of incentives affecting academics' choice of research topics and transfer tools, with special emphasis on the role played by personal reputation and intellectual property (Stephan, 2010). We contribute to this line of enquiry by studying the distribution of reputation among scientists working in teams and engaged in the "simultaneous disclosure" of scientific and commercial knowledge, by means, respectively, of publications and patents (Gans et al., 2011). In particular, we show that the distribution of authorship (of publications) and inventorship (of patents) among members of a research team reflects not only the individual contributions to the research effort, but also the relative bargaining power and incentives of team members.

We describe both authorship and inventorship as 'attribution rights', a form of intellectual property recognized both by the social norms of science (Merton, 1957) and by international conventions on "moral rights" of authors and performers (art. 11 in UNESCO, 2001; and art. 6 in WIPO, 2008). Such rights provide signals to participants to knowledge markets, where problems of asymmetric information are particularly acute. Indeed, the scientist's record as author and/or inventor is used by funding agencies or business companies to find the best researcher to sponsor, or the most-suited collaborator or consultant.

Assigning attribution rights is however difficult when the relevant activities are performed by teams, rather than individuals, as it is increasingly the case with science and technology (Katz and Martin, 1997; Jones et al., 2008; Wuchty et al., 2007; Jones, 2009). This is because the existing social and legal norms defining attribution rights leave room to contrasts and negotiations among team members (Fernandez-Molina and Pais, 2001; Fisk, 2006). We argue that such negotiations, while possibly resolving in an optimal way internal disputes, may misinform third parties on each team member's actual contribution to the research and inventive efforts, thus possibly generating negative information externalities. As already discussed in other contexts (e.g. Aghion and Bolton, 1987, on exclusive-dealing contracts; Hansmann and Santilli, 1997, on visual artists' rights), such externalities may affect negatively the efficiency of private agreements, as stated by the Coase theorem (Coase, 1960; Hermalin et al., 2007).

With the help of a stylized theoretical model we identify a number of conditions under which inventorship may be attributed more sparingly than authorship, so that not all the co-authors of a scientific publication end up being included in the list of inventors of the related patents. In particular, we argue that junior and female co-authors can convinced to give up inventorship, other things being equal, due to lower incentives to reclaim this type of attribution right, as opposed to authorship.

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 gender information on the selected academic inventors and their co-authors. We estimate that the risk of an author's exclusion from a related patent is higher for junior and female scientists.

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 formal model and the related proposition (full analysis in Additional Material). 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 and discusses the relevance of our findings for the domain of the economics of science, and beyond.

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 moved 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.[1] According to Jones (2009), the 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).

[1] 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).

Notably, the average number of inventors per patent remains lower than that 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 the resulting knowledge output limits the inventors' freedom to choose their research partners, contrary to what happens to academic scientists. However, differences in the number of authors and inventors can also be found when comparing 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 being considered authors or inventors are different, or that some differences exist in the established practices of attribution. A vast sociological literature exists, which illustrates how negotiation plays a role in authorship attribution. A sparse legal literature on inventorship suggests the same. We examine both of them.

2.2 The Vexed Issue of Authorship

Attribution practices in scientific authorship have been largely discussed with references to malpractices, such as 'guest' (or 'honorary') and 'gift' authorship, which occur when a scientist is listed in the authors' by-line of a paper to which she has not contributed (Mowatt et al., 2002). These problems are particularly felt in biomedical research, because of the great importance attached to ethical integrity and responsibility attribution in that field (Biagioli, 1998). 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).[2]

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 herself 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 research 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. Bates et al. (2004) find that 60% of 72 articles surveyed in 2002 in the *Annals of Internal Medicine* and 21% of 107

[2] 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). More recently, several commentators have suggested that the notion of authorship in science is out-of-date, linked as it is to the idea of an integral responsibility of all contents of a paper, which is at odds with dominant practices of teamwork and division of labour. We come back to this point, as well as proposals to replace authorship with "contributorship", in the Conclusions.

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 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; Drenth, 1998; Mowett et al. 2002). 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, often 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 occasionally rewarded for their dedication, or very senior scientists included out of deference). Still, some evidence exists on senior authors' latitude in retaining first authorship or granting it to junior co-authors, depending on matters of convenience, such as the wish to support a disciple's career or the necessity to boost their own (Zuckerman, 1968)[3].

2.3 Inventorship

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]

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

[3] 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).
[4] See for example *Yeda Res. & Dev. v. ImClone Systems Inc.* in 2006.
[5] 802 F.2d 1367, 1376 (Fed. Cir. 1986)

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; www.legislation.gov.uk/ukpga/1977/37). In Italy, as in many other countries, no specific definition of inventor is provided by legal texts. The legal doctrine on 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 property rights.

Existing inventorship norms in both the US and in Europe are more restrictive than the editorial rules defining authorship. As stressed by the legal opinions of university TTO officers and IP 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. Therefore, a scientist who raises funds, conceives the initial experiment, and revises the draft paper can qualify as the author of a project-related paper (at least according to the ICMJE guidelines), but not as the inventor of any project-related patent. The same applies to whom, at the opposite end, follows "the complete instructions" of a colleague or superior.[6] Notice that, as far as inventorship is concerned, name ordering affects neither the economic rights to which inventors are entitled (all inventors are equally entitled to any compensation) nor the reputation they get (alphabetical name ordering is the norm).

Outside these extreme cases, however, the application of legal definitions of inventorship to members of a team is as controversial and open to arbitrary decisions as that of authorship (Fasse, 1992). Colyvas (2007) shows that, for the case of Stanford university, decisions on inventorship attribution, very much like those on authorship, often depend upon the discretionary judgement of the most senior members of the team, who manage the economic details of the research and exercise authority. Finally, 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.

2.4 Seniority and gender in negotiation over attribution rights

The previous discussion suggests that the attribution of both authorship and inventorship may be subject to negotiation within the team. Third parties observe only the final outcome. Such outcome may be affected by seniority, gender, and their relationship to individual team members' incentives. Life-cycle models of scientists' behaviour suggest that junior scientists who pursue an academic career invest heavily in building a reputation within the academy, while their senior colleagues may choose to

---

[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).

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, we expect junior scientists to value authorship more than inventorship. When compared to patents, papers circulate earlier, more widely and contain a much more readable explanation of the research results.[7] On the contrary senior scientists may attach more importance to inventorship, both for the expected returns from patents and as a mean to increase their reputation (as technologists) beyond the boundaries of the academic community.

Reclaiming attribution rights also entails costs. In particular, junior scientists may attach a negative value to the possibility to enter into conflict with their team's seniors, from whose mentorship their careers largely depend (Pezzoni et al., 2012; and references therein).

It is also important to take into account gender. First, women may be involved in fewer research projects than men, due to family-career trade-offs or a disadvantaged academic position, so that authorship will have for them a higher marginal value. Second, female scientists may also assign a lower value to inventorship than men. In this respect, several authors find that female scientists patent less than men with the same publication records (Breschi et al., 2005; Azoulay et al., 2007; Stephan et al., 2007; see also Whittington and Smith-Doerr, 2008). This may be explained by Ding et al.'s (2006) and Murray and Graham's (2007) analysis of longitudinal data on careers and field interviews, which show that women have fewer connections to operators in the marketplace than men. This diminishes their opportunities to commercialize their research results or to cash in their reputation through consultancy or participation in high-tech companies.[8]

3. Negotiation over Authorship and Inventorship: A Stylized Model

We propose a formal model of negotiation over attribution rights, which both summarizes the stylized facts derived from the literature and develop the proposition for the empirical analysis. We proceed as follows. We first describe the model and discuss its assumptions (3.1); then we derive the conditions for exclusion from inventorship (3.2), to be tested in sections 4 and 5. Welfare implications are discussed in section 3.3.

---

[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.

[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.

3.1 Assumptions

The model formalizes a bargaining process over attribution rights between two scientists that assign different values to authorship and inventorship, and face litigation costs. In particular we consider a team composed of two focal scientists, senior *S* (the team leader) and junior *J*, plus an indefinite number of other scientists or technicians, for whom notation is not necessary as they do not enter the negotiations. We initially ignore gender issues, which justifies our use of male pronouns and adjectives for both scientists.

The team produces research findings that originate both a scientific publication and a patent. *S* and *J* are the team members who contribute more than anybody else to the research effort, and yet one of the two (either *S* or *J*) contributes more than the other[9]. They engage in negotiations to decide who will be first author of the publication and, jointly, on who will appear as inventor on the patent.. As far as authorship is concerned, the two scientists face different alternative ("non first") positions. For *S*, this is the last position in the paper (which would signal at least his role as project leader), for *J* it is any intermediate position between first and last[10]. As for inventorship, name ordering does not matter, so *S* and *J* must simply decide whether to share it by listing both their names on the patent, or to exclude one of them from the list. *S* and *J* play a non-cooperative game with complete information. For simplicity, we assume that *S*, being the team leader, has all the bargaining power, and makes a take-it-or-leave-it offer to *J*. *S*' strategy (that is, his proposal to *J*) is composed of two parts: one part concerns who should be first on the paper (respectively, *1S* or *1J*), while the other consists of the list of inventor (*S*, *J* or *JS*). Then, for instance, (*1S*, *JS*) is a proposal according to which *S* is the first author, and both *J* and *S* appear as inventors on the patent.

We indicate the economic value of authorship with *R*, and that of inventorship with *v*. Following our discussion in section 2.4, we assume that the value of authorship differs for *S* and *J*. Formally, R_1^S and R_{N1}^S (with $R_1^S \geq R_{N1}^S$) represent the benefits from authorship for *S* as "first" and "non first" author. Similarly R_1^J and R_{N1}^J (with $R_1^J \geq R_{N1}^J$) represent the benefits for *J*. As for the value of inventorship, we assume it, for sake of simplicity, to be the same for *S* and *J*, and to be either *v* (in case one of the two is the only one listed as inventor on the patent), $\frac{v}{2}$ (shared inventorship), or zero (exclusion from inventorship). On one hand, this assumption is not

[9] Apart from *S*' leadership role, we do not distinguish between types of contribution (conception, execution etc), but simply assume that team members who contribute less are less qualified for getting both first authorship and inventorship (more details below). We also ignore all issues of team formation and research strategy (but see our discussion of welfare implications) and we do not consider explicitly the decision on the type of research to undertake and whether to patent or not.

[10] We assume that if *S* becomes the first author, he may dispose of the last position in favour of any non-focal team member more senior than *J*. Otherwise, non-focal team members will always appear in intermediate positions in the authors' by-line of the publication, or not appear at all. For simplicity, we also assume they will never be granted inventorship. As for *J*, he will get last authorship only if member of a two-member team along with *S*. In this case, however, last authorship can be hardly interpreted as signal of research leadership, so it is much less valuable than first authorship.

restrictive, because for most results what matters is the comparison between benefits of publications (which are individual-specific) and patent value. On the other hand, the model can be easily extended to patent values which differ across scientists.

The key assumption of our model is that first authorship's value (relative to inventorship) is "high" for *J* and "low" for *S*. In particular we assume that:

(i) $R_1^S - R_{N1}^S < \frac{v}{2}$; and

(ii) $R_1^J - R_{N1}^J > \frac{v}{2}$.

Finally, L^J represents the "litigation" cost faced by *J* when refusing *S*' proposal. For simplicity, we assume this to be the same, whether the proposal violates the social norms on authorship, the legal ones on inventorship, or both. The assumption is reasonable to the extent that these costs depend upon the risk of compromising his relationship with a potential mentor and the ensuing damage to his career, as discussed in section 2.4. As for L^S(the litigation costs incurred into by *S* when his proposal goes against the norm and it is refused by *J*) we assume it to consist first and foremost in a reputational loss, linked to the fact that *J*'s refusal may give publicity to *S*' deviation from norms.[11]

When evaluating the proposal by *S* and the opportunity to refuse it, *J* compares it to the prescriptions of the legal norms on inventorship and the social norms on authorships, based upon the two scientists' relative contributions to the research effort These prescriptions identify the "outside option" that *J* can implement (at a cost), by rejecting the proposal (this action is denoted with *NA*). In case the proposal is accepted (action *A*) the corresponding payoff depends solely on the authorship and inventorship value implied by the proposal. Four possible cases (a. to d.) must be considered**:**

- If *J* has contributed more than *S* to the project, but *S* has contributed enough to deserve inventorship;
    a. *J* should appear as first author on the paper and both *J* and *S* should appear on the patent; or
    b. *J* should appear as first author and as the only inventor.
- If *S* has contributed more than *J* to the project, but *J* has contributed enough to deserve inventorship;
    c. *S* should appear as first author and both should appear on the patent; or
    d. *S* should appear as the first author and as the only inventor.

Coherently with the discussion on legal and social norms on attribution rights (see section 2.3) these are the only four plausible cases to be considered. According to the norms it is not conceivable that inventorship is granted and authorship is not.

[11] As a matter of fact, the exact value and interpretation of L^S is irrelevant, as long as it is positive, since its role is to make *S* strictly preferring a norm-abiding proposal to a non-abiding one, if rejected.

Following standard practices in two-stage games with complete information, the solution concept is subgame perfection. Figures from A1 to A4 in the Additional Material report the game in extensive form for cases a. to d., respectively

3.2 Results and empirical implications

We can derive from our model a number of propositions on the distribution of attribution rights as a function of team members' contribution to research and personal characteristics (seniority and gender). We focus here on just one of them, which summarizes the equilibria of the game in terms of exclusion from inventorship of either *S* or *J* and lends itself to being tested with bibliometric data (for a complete list of propositions and their proofs, see the Additional Material).

Proposition. In equilibrium:

1. *S* is *never* excluded from inventorship if listed as first author. He can be excluded only if he has contributed to the project less than *J* *and* his exclusion is in accordance with legal norms (case b) *and* *J*'s litigation costs are low ($L^J < \frac{v}{2}$); in which case, he will appear as last author.
2. *J* *can* be excluded when listed as first author, under the condition: $\frac{v}{2} < L^J < \frac{v}{2} + R_1^J - R_{N1}^J$. *J* is *always* excluded from the patent when he is not listed as first author.

The proposition implies that exclusion from inventorship is not merely driven by legal norms. This would be the case if authors listed as first authors were never excluded. However, a junior scientist may end up being excluded from inventorship even if listed as first author, for intermediate values of litigations costs. This is never the case for his senior. In other words, when it comes to attribution rights, seniority matters. This result is intuitively explained by *S'* and *J's* different preferences for publications and patents, as captured by the assumptions (i) and (ii) introduced above. By granting *J* the first authorship while at the same time excluding him from the patent, *S* can maximize his economic return from the invention, while at the same time letting *J* obtain what is most valuable to him. The deal is sustainable for a relatively wide range of *J*'s litigation costs.[12]

Notice that seniority matters not only because junior scientists are more at risk of exclusion from inventorship, irrespective of their contribution, but also because senior scientists end up being

[12] Another way of looking at this result is that *J* can obtain the first position in the paper, although *S* has all the bargaining power and litigation cost are large (albeit not too large, as they have an upper bound). This point is discussed by the literature on multiple issue bargaining (e.g. Fershtman, 1990), where it has been shown that simultaneous bargaining on several items for which agents have different preferences can yield Pareto improvements. In the context of our model, it can be shown that there are values of L^J for which, given the scientific contribution, *J* would not get first authorship if he and *S* had the possibility to bargain over scientific authorship only, and would otherwise get it in case the bargaining included inventorship, too.

excluded only when juniors face low litigation costs L^J. This implies that senior scientists have a lower probability of being excluded from inventorship, irrespective of their contribution.

Still, our results imply that a correlation exists between an author's position in the paper's by-line line and the probability of exclusion. S is never excluded when he is first author, and, when last, he is is excluded only for low values of L^J; while J is always excluded when he is not the first author.

Gender is a factor that we do not explicitly include in our model. However, following the analogy analogy between junior and female scientists we drew in section 2.4, gender issues can be accommodated by assuming that female scientists assign either lower value to inventorship or higher value to first authorship than their male colleagues. The former hypothesis can be accommodated in the model by specifying scientist-specific values of v. The latter corresponds to a higher value of $R_1^J - R_{N1}^J$. This makes it more likely to meet the condition (ii) , under which the scientist accepts exclusion from the patent, while getting first authorship.

3.3. Welfare implications

Our model implies the existence of two potential sources of welfare losses, the first one affecting third parties, the second concerning the research team members. Further losses may appear, albeit indirectly, as a result of *S'* and *J*'s choice of the research projects to undertake, on the basis of their expectations concerning the outcome of negotiations over attribution rights.

First, third parties (e.g. recruiting companies and institutions, funding agencies) suffer a welfare loss under all equilibria in which legal and social norms are violated. This is because they receive a wrong signal and may end up recruiting or supporting an under-performing scientist.

Second, J will suffer a welfare loss whenever he has contributed the most to the research effort, so that he should get first authorship according to social norms, but gives it up when facing high values of L^J. In this case, the violation of norms leads to a decrease of overall utility of team members (see the Additional Material for proof). The intuitive explanation of this result is straightforward: while the allocation of inventorship is irrelevant (the total value for the team is v in any case), the marginal value of first authorship is higher for J than for S, so that any change of the name ordering in the publication decreases the total utility of the team. [13]

Third, more losses may materialize whenever *S'* and *J's ex ante* choice of research projects is affected by their expectations of negotiation's outcome. Suppose that, whenever they undertake a research project, S and J sustain a sunk cost (in terms of time, resources, etc). Suppose also that this occurs *before* attribution rights are negotiated, but with rational expectations on the equilibrium

[13] The opposite holds when it is S who contributes the most to the project. In this case the utility of the team weakly *increase*s when the attribution rights are not those put forth by norms. In this case, the third parties' welfare loss may be (at least partially) compensated by the team's gain.

outcome of negotiations. Under these circumstances, the scientists will be willing to participate only to projects whose benefits, in terms of attribution rights, are higher than the sunk cost[14].

Consider first the case in which *J* provides the largest contribution to the project. According to our model, any departure from the norms (*S* as first author or *J* excluded from the patent) makes *S* better off, and *J* (weakly) worse off. As a consequence, there will be a number of projects, which require relatively high sunk costs by *J*, that will *not* be initiated, due to *J*'s refusal to join in, while they would have been under a norm-abiding distribution of attribution rights. However, when the largest contribution comes from *S*, both *S* and *J* are better off under the norm-deviating equilibrium (*J* as first author) and end up undertaking more projects than otherwise.

4. Data and Methodology

Our empirical methodology makes use of bibliometric data and 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 research results having both scientific interest and commercial value (Gans et al., 2011). 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.

Over the last ten years, several papers have been published, which make 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:

[14] Let us assume also that the investment required to *S* is low enough, and therefore *S* will always be willing to participate to the project.

(1) From the KEINS patent database, we extracted all Italian academic inventors from the four academic disciplines with the highest propensity to patent, 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.[15]

(2) For the selected academic inventors, we collected all publication data from the ISI-*Web of Science* (ISI-WoS), from 1975 to 2003.[16]

(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.[17]

(5) We established the gender of as many as possible co-authors (841 out of 899), by manually retrieving their publications and looking at first names.

4.1 Patent-publication pairs: methodology

Given *t* the priority year of a patent and *i* one of the its inventors, a *potential* patent-publication pair is defined as the association between the patent and a publication that has *i* among its co-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.[18]

For all documents in this *potential* PPP set we examined the title and abstract, and transformed them into comparable information sets. The first step of the transformation consisted in removing

[15] 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.

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

[17] 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). We report it in Box 1 in the Additional Material.

[18] Academic inventors' declarations on the existence of 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.

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 (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* (*Cos*).

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

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

Theoretical values of *Cos* are in the continuous [0,1] range. In our application, *Cos* takes values comprised between 0 and 0.75. For our analysis, we selected those PPPs whose *Cos* 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.[19]

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. [20]

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.

---

[19] Table A1 in the Additional Material 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, which we describe in Box A2 in the Additional Material. The descriptive results we obtained did not change much and are available on request.

[20] 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.

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 main variables

As we estimate the probability of an author's exclusion from inventorship, we arrange the database accordingly. In particular, each PPP *j* is repeated as many times as the total number of authors appearing on the publication(s) included in *j*. So for each author *i* and PPP *j* we know whether he/she is excluded or not 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 author's position in the by-line, transformed into three dummy variables: FIRST, LAST, and MIDDLE (reference case). Following the discussion in section 2 and the model in section 3, we expect both FIRST and LAST to bear a negative sign.[21]
- 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 the author's first publication (t_{fpi}). As for relative seniority, we measure it with a continuous variable, ranging from 0 to 1, defined as:

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

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

[21] The information provided by the name order of authors may vary between papers co-authored by several members of one research team, and papers co-authored by authors from several teams. In the latter case, authors may be listed first according to the team they belong to (with teams ordered either according to criteria we ignore) and then either alphabetically or according to the within-the-team negotiation outcome (substantive order). 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+substantive order, the FIRST and LAST dummies still bear an unequivocal meaning (they indicate respectively the first author in the first team listed, and the last author in the last team listed), but have less explanatory power, because authors in middle position comprise also many authors listed as first and last by other teams, alongside with 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, until recently, authors and affiliations were listed in separate fields, with no keys to connect them (in addition, it is often the case that one author has multiple affiliations). We checked the robustness of our results by running a set of regressions also on a restricted sample that includes only the publications with multiple affiliations. The results do not change in any meaningful way and are available on request.

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:

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

  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_SCHOLAR.

- Authors' gender, as represented by the dummy variable FEMALE. Following the discussion in Section 3 we expect a positive effect of FEMALE on the probability of exclusion. However, gender and contribution may be correlated, to the extent that female 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*FEMALE and LAST*FEMALE to be greater than the coefficients for, respectively, FIRST*MALE and LAST*MALE, and we can safely interpret the difference as entirely due to gender. We also expect the coefficients for MIDDLE*FEMALE 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.[22]

- 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 he/she has more than one publication, and/or these are related to more than one patent.[23]

5.1 Descriptive Statistics

Table 1 reports the number of patents and publications in the selected (*actual*) PPP sample, by priority year and technological field. It reports also the number of authors by field. 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 Figure 1 about here]

Table 2 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 2 also reports similar information for the initial set of original (*potential*) PPPs: notice that, due to a much less precise matching between

[22] 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 are available on request.

[23] 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.

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 2 here]

These results are consistent with the existence of an exclusion process.[24] However, we observe significant differences across disciplines. Table 2 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 3 here]

[TABLE 4 here]

In order to investigate whether a specific pattern of exclusion emerges, in Table 3 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 4 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.[25]

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 or with a number of inventors greater than or equal to the number of authors; (3) all the academic inventors from the KEINS database, for which the probability of being excluded is zero by

[24] 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 2: 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.
[25] Results available on request.

construction;[26] 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). Table 5 provides the summary statistics.[27] 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 5 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, as expected.

Table 6 displays the results of a set of Logit regressions where the dependent variable is the probability of an author's exclusion from the inventorship of a related patent. We assume that observations are independent across individuals, but not necessarily across publications and patents by the same individual scientists; therefore we cluster errors by individual. 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 (FEMALE). Column (7) also controls for gender, interacted with the information on the author's position in the by-line (MIDDLE*MALE is the reference case).

[TABLE 6 and 7 here]

[26] 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.

[27] 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.

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 6. Also first authors are less likely to be excluded than last ones. In Table 7 we calculate the changes in the predicted probability of exclusion for a discrete change in FIRST and LAST (with all other variables held at their mean value), based upon regression (1) in Table 6: we obtain values equal to -0.16 and -0.12, respectively.[28]

These results are coherent both with the legal rules on inventorship we examined in section 2 and the equilibria of our negotiation model. The former state that inventorship must be awarded to team members who contribute most to the research project, while the latter indicate that, after negotiation, a correlation exists between the authors' position in the by-line and their probability to get inventorship.

Table 6 also shows that the probability of exclusion decreases significantly with the scientist's years of activity, as expected from our model's treatment of seniority. 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, the coefficients for MOST_JUNIOR and RELATIVE_SENIORITY being significantly negative.

Table 8 reports the predicted probabilities of exclusion based upon regression (1) in Table 6, 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 8 here]

These results indicate that, given the position in the by-line, a junior scientist is significantly more at risk of being excluded from inventorship than a senior one. 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. This is coherent with our negotiation model, where it showed that when senior scientist *S* gets first authorship he will also get inventorship, while the same does not apply to junior scientist *J*.

[28] 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 6.

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 from the patent. Again, this is coherent with our model, where it suggests that last authorship may not go with inventorship in two cases: when it goes to *S* and *S* cannot illegitimately reclaim inventorship, due to *J*'s low litigation costs; or when *S* gets first inventorship and concedes last authorship to either *J* or a non focal team member. In the first case, *J*'s litigation cost may be low because *S* is not so senior (*J* cannot count so much upon him for mentorship). In the second case, it is likely that the team member who gets last authorship is more junior than *S* (he certainly is if he is J).

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

When we interact gender and the position in the by-line (regression (7) in Table 6) 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, as suggested by the extension of our model to gender.[29]

Finally, PUB_STOCK is slightly positive and RELATIVE_PUB_STOCK is positive and significant, conditional on seniority, which means that the scientists with the larger stock of publication in the team are more likely to be excluded from inventorship. In line with our discussion in Sections 2 and 4, we interpret it as evidence of guest authorship practices involving well-reputed scientists (with guest authors more likely to be excluded from the patent).

5.3 Robustness checks

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

Table 9 reports the results of a set of regressions, identical to those in Table 6 but for the definition of $\Pr(y_{ij}=1 \mid x)$, which reads now as the conditional probability that author *i* is excluded, not

---

[29] 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 6 are somewhat weaker Full results are available on request. In any case, in logit models, interaction effects are present also when the coefficient of the interaction terms is assumed to be zero, due to non-linearity.

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).

[TABLE 9 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, with the only exception of DELTA_YEARS. The gender effect remains very strong, its estimated coefficient being larger than in Table 6. When we interact gender and position in the by-line, the coefficients maintain their sign and significance (with the exclusion of LAST*MALE). We conclude that, when altering our definition of "exclusion from inventorship", the core of our results remains unchanged.

A second possible cause of concern is the potential mix, in our PPP sample, of both false positives (unrelated patents and publications in same PPP) and false negatives (unrelated patents and publications we failed to identify as such). 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 be LAST) sign more papers than junior scientists.

In order to control for this potential problem, we restrict our sample to the PPPs with a *Cos* 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 *Cos* equal to 0.174. We then run a set of regressions identical to those of Table 6. Our results (Table 10) 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 with what found in Table 6.[30]

[TABLE 10 here]

[30] 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 *Cos* 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.

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 from the legal viewpoint, are more likely to publish their papers after filing the 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 11 replicates the Logit regression of column (1) of Table 6, 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 11 here]

The results for DELTA_YEARS≥0 are similar, in terms of sign and significance of the estimated parameters, to those of Table 6. The main difference consists only in the magnitude of FIRST and LAST parameters, which are respectively lower and higher than in Table 6 (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. This is consistent with the possibility that part of our results in Table 6 were affected by a bias due to the methodology followed for the creation of our PPP sample.

Alternatively, we can explain the results of Table 11 with the possibility that, within a team of scientists, the decision to file a patent may follow two different routes, which affects differently the distribution of inventorship credits. Patents in PPPs with DELTA_YEARS≥0 may be the result of a route based on searching for IP protection from the very beginning of the research project, 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 some promising results. In this case, the patent may generate specific additional activity by the author who has contributed most to the research activity, who will then have higher chances both to be retained as first author and to get inventorship. 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.[31]

[31] 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.

6. Conclusions

In this paper, we have investigated the determinants of attribution rights distribution within scientific teams whose research results are diffused jointly through patents and publications. We first argued that social and legal norms concerning authorship and inventorship are of difficult interpretation and application to teams, which leaves room for negotiations among team members. We have then formalized the negotiation process with a stylized model involving two scientists (a junior and a senior one), who assign different values to first authorship and inventorship, have asymmetric bargaining power, and face litigation costs. Under these assumption, the negotiation process is described as a two-stage game with perfect information, whose equilibrium outcomes suggest that the junior author may agree to give up his inventorship rights in order to secure first authorship, while the senior author, whenever gets first authorship, also retains inventorship. Results for the junior scientist can be intuitively extended to female ones. This amounts to say that seniority and gender matter, so that junior and female scientists have a higher probability to be excluded fro inventorship when facing high enough litigation costs.

We then test this proposition on a new and original database composed of Patent-Publications Pairs by a set of Italian academics active in the 1990s-early 2000s. We find that first and last co-authors of a focal publication are less likely to be excluded from the related patent or patents, as suggested by legal norms on inventorship. But we also find that, *ceteris paribus*, junior and female co-authors are more likely to be excluded from a related patent, as predicted by our model.

This implies that within-the-team negotiations may lead to allocations of attribution rights that do not reflect entirely the individual contributions to research advancements and inventions. As a consequence, society may incur into net welfare losses, which are not always compensated by within-the-team efficiency gains.

More generally, our results contribute to existing criticism directed at the obsolescence 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 views of discovery and invention as resulting from an individual spark of genius (Fisk, 2006). Some steps in the direction of abandoning authorship have been undertaken by several scientific journals, especially in the medical sciences, which now require authors not merely to identify themselves as such, 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 also dates back to a time – the XIX century – when the existence of patents had been put into question, and was defended by portraying intellectual property as an individual right, as well as by creating 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 applies not only to academic patenting, as it solicits an investigation on the overall adequacy of present norms on inventorship, also when applied to industrial R&D settings.

Our paper contributes to the literature also from the technical viewpoint. First, it provides a formal economic treatment of negotiation over attribution rights, a topic that so far has been quite neglected by economists. One notable exception is Engers et al. (1999), who develop a model of bargaining over authors' order in a paper, and show that alphabetic ordering (which is widespread in economics) can be a sustainable equilibrium (as opposed to the relative contribution ordering, more common in other disciplines). However, Engers et al. do not consider bargaining over inventorship, as we do, or the possibility of exclusions (from papers or patents).

Second, we contribute to the emerging bibliometric 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 less frequent 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 (our results refer only to Italy, whose academic system assigns large discretionary power to seniors; see Pezzoni et al., 2012). 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. Similar differences may be found for inventorship. For example, Häussler and Sauermann (2013) replicate in part our exercise for a sample of British and German life scientists, but do not find any evidence of a relationship between the distribution of attribution rights and gender (while they find some for seniority). They also compare the behaviour of industry scientists to that of academics, and find no substantial differences. Further theorizing on this issue should take into account the different incentives to publishing and patenting of the two categories of researchers (Lacetera and Zirulia, 2012).

Finally, our research questions can be extended to other fields of human creativity, in which – as in science – activities are increasingly performed by teams, but careers are built upon personal reputation, fuelled by attribution rights. 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 contributing to the production of a film 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 design, architecture, or advertising, individual attribution rights are still the key form of attribution. In these cases, we may be interested to investigate whether negotiations among team members occur, as in science; and what characteristics of the individuals affect their outcomes.

References

Aghion P., Bolton P. (1987), "Contracts as a Barrier to Entry", *American Economic Review* 77 (3), pp. 388-401.

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 behaviour: 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)

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

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), pp. 91-109

Coase R.H. (1960), The Problem of Social Cost, *Journal of Law & Economics* 3, pp. 1-44

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

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

Engers M., Gans J.S., Grant S., King S.P. (1999), First-Author Conditions, *Journal of Political Economy* 107(4), pp. 859-883

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

Fernandenz-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

Fershtman C., (1990). "The Importance of the Agenda in Bargaining", *Games and Economic Behavior* 2(3), pp. 224-238,

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. (2011), "Contracting Over the Disclosure of Scientific Knowledge: Intellectual Property Protection and Academic Publication", mimeo (http://ssrn.com/abstract=1559871)

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

Haeussler C., Sauermann H. (2013), Credit where credit is due? The impact of project contributions and social factors on authorship and inventorship.*Research Policy* 42(3), pp. 688-703.

Hermalin B.E., Katz A.W., Craswell R. (2007), "Contract Law" in: Polinsky A.M., Shavell S. (eds.), *Handbook of Law and Economics*. Elsevier, Amsterdam, pp. 3-138.

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)

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

Lacetera N., Zirulia L. (2012) "Individual preferences, organization, and competition in a model of R&D incentive provision", Journal of Economic Behavior & Organization 84(2), pp. 550-570.

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

Lissoni F., Llerena P., McKelvey M., Sanditov B. (2008), "Academic Patenting in Europe: New Evidence from the KEINS Database", *Research Evaluation* 17(2), pp. 87-102

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., 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", mimeo (http://fmurray.scripts.mit.edu/docs/Murray.Stern_LearningtoLivewithPatents.pdf)

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

Pezzoni M., Sterzi V., Lissoni F. (2012) "Career progress in centralized academic systems : an analysis of French and Italian physicists", *Research Policy* 41(4), pp. 704-719

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 & FIGURES

Table 1. Patents, publication, authors and inventors by field and priority years in the selected PPPs

	Chemistry	*Electronics*	*Pharma*	*Biology*	*Total*
Patents					
88-94	9	39	19	24	91
95-00	8	63	26	25	122
Total	*17*	*102*	*45*	*49*	*213*
Publications					
90-95	15	62	35	89	201
96-01	10	117	43	79	249
Total	*25*	*179*	*78*	*168*	*450*
N. of authors[+]	*72*	*311*	*253*	*527*	*1138*

[+] *The sum of the number of authors across fields is 1163, the total number of authors is 1138 because 25 authors publish in two different fields, in particular 20 authors in pharma and biology*

Table 2. 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	

Table 3. 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 4. 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 5. 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
FEMALE	1897	.36	.48	0	1
FIRST*FEMALE	1897	.07	.25	0	1
MEDIUM*FEMALE	1897	.27	.45	0	1
LAST*FEMALE	1897	.02	.15	0	1

Table 6. Probability of exclusion from inventorship: Logit regressions.

	(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)	
FEMALE				0.75**	0.84**	0.72**	
				(0.37)	(0.37)	(0.31)	
FIRST*FEMALE							-0.43
							(0.63)
FIRST*MALE							-0.90***
							(0.30)
MIDDLE*FEMALE							0.93***
							(0.30)
LAST*FEMALE							-0.13
							(0.65)
LAST*MALE							-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 7. 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	
FEMALE	0.91	0.09
	[0.87. 0.96]	
MALE	0.83	0.17
	[0.78. 0.88]	
Difference	0.08	

Note. 95% confidence intervals in parenthesis

Table 8. 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 6

	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 9. 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)	
FEMALE				1.01***	1.06***	0.90***	
				(0.38)	(0.37)	(0.34)	
FIRST*FEMALE							0.30
							(0.63)
FIRST*MALE							-0.57*
							(0.33)
MIDDLE*FEMALE							1.15***
							(0.35)
LAST*FEMALE							0.19
							(0.51)
LAST*MALE							-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 10. Probability of exclusion from inventorship: Logit regressions. (restricted PPP sample: top 5% of the similarity score)

	(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)	
FEMALE				0.55	0.69*	0.52	
				(0.36)	(0.36)	(0.32)	
FIRST*FEMALE							-0.38
							(0.73)
FIRST*MALE							-0.91**
							(0.40)
MIDDLE*FEMALE							0.59*
							(0.32)
LAST*FEMALE							-0.55
							(0.64)
LAST*MALE							-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

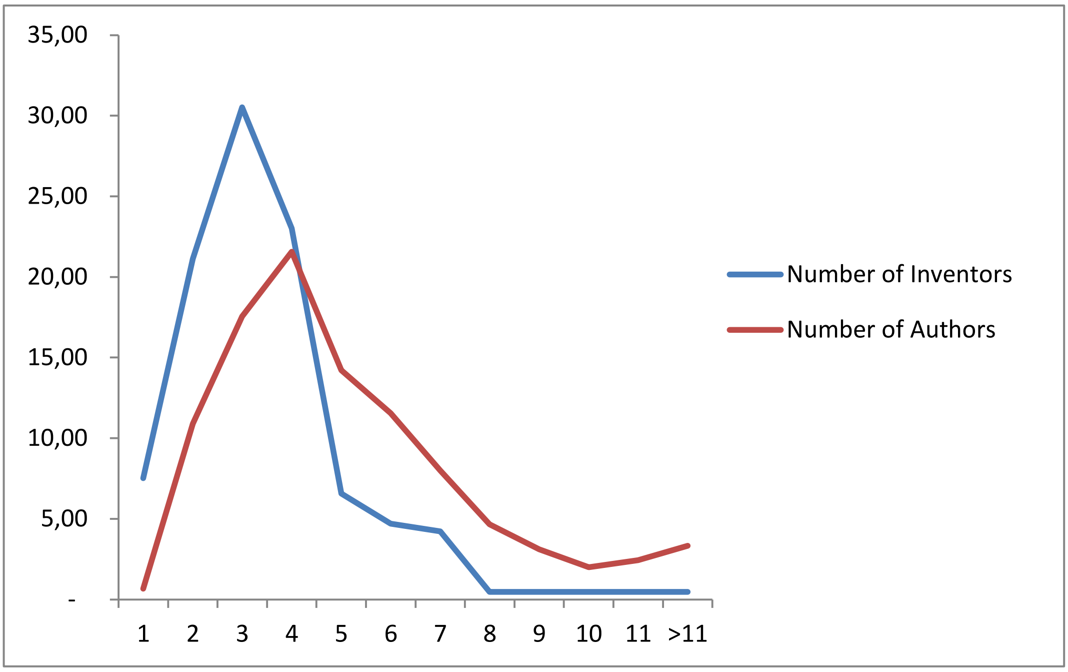
Robust standard errors in parentheses (clustered for individuals)
*** p<0.01. ** p<0.05. * p<0.1

Table 11. Probability of exclusion from inventorship: Logit regressions. by values of DELTA_YEARS

	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)
FEMALE			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

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