Paper to be presented at
the DRUID16 20th Anniversary Conference
Copenhagen, June 13-15, 2016

Sleeping beauties in technology - delayed recognition of breakthrough inventions

Elena M. Tur
TU/e
Technology, Innovation & Society
e.m.mas.tur@tue.nl

Joaquín M. Azagra-caro
INGENIO (CSIC-UPV)
INGENIO (CSIC-UPV)
jazagra@ingenio.upv.es

Koen Frenken
Utrecht University
Innovation Studies
K.Frenken@uu.nl

Abstract

This paper analyzes the phenomenon of delayed recognition of breakthrough knowledge, that is, the phenomenon that some inventions initially do not diffuse, but only later become accepted. The failure in the recognition of these radical developments depends both on the social position of their authors, but also on their technological characteristics. Following the literature on delayed recognition and sleeping beauties in scientific papers, it identifies sleeping beauties in technologies as highly cited patents that did not receive citations for a long period, and it compares them to the control group of other highly cited patents from the same period. On the one hand, breakthroughs in the most codified technological areas are more susceptible to earlier appreciation. On the other hand, a collaboration between many of authors, as well as their previous experience, also facilitates preventing delayed recognition.

Jelcodes:O33,-

Sleeping beauties in technology - delayed recognition of breakthrough inventions

Abstract

This paper analyzes the phenomenon of delayed recognition of breakthrough knowledge, that is, the phenomenon that some inventions initially do not diffuse, but only later become accepted. The failure in the recognition of these radical developments depends both on the social position of their authors, but also on their technological characteristics. Following the literature on delayed recognition and sleeping beauties in scientific papers, it identifies sleeping beauties in technologies as highly cited patents that did not receive citations for a long period, and it compares them to the control group of other highly cited patents from the same period. On the one hand, breakthroughs in the most codified technological areas are more susceptible to earlier appreciation. On the other hand, a collaboration between many of authors, as well as their previous experience, also facilitates preventing delayed recognition.

1 Introduction

Ever since the first studies on innovation as the motor of economic development, scholars have distinguished two types of inventions: incremental and radical. Incremental innovations are based on small improvements along an established technological path, while radical innovations, also known as breakthroughs, are disruptive innovations that involve major changes in the dominant technological trajectory. The established technological community, thus, might not be able to recognize at first their implications and follow for a while the conventional, well-known technological path, even after it has been rendered obsolete by a breakthrough development. Possible explanations for this tendency include path dependence of the technological process (David, 1985), the lack of technologically complementary technologies (Valentin and Jensen, 2002), or a weak social position of the inventor (Singh and Fleming, 2010).

Understanding why some breakthrough inventions experience delayed recognition would help unveiling possible failures in the process of technological diffusion. A new advance at the frontier of technological progress may be ahead of its time and remain latent until complementary technologies have been developed. An alternative hypothesis would be that the social network of inventors is determinant for the diffusion of inventions and isolated actors with a weak social position lack the means to make their inventions noticed. This second explanation would reveal a shortcoming of the technology system, since important developments would be ignored, delaying further technological development. In such a case, there would be scope for policy action to correct such frictions in the diffusion of technologies.

This study explores the case of sleeping beauties (SBs), breakthrough inventions that experienced delayed recognition, by means of patent data. Patents are legal documents that grant exclusive rights of exploitation of an invention for a limited period of time, in exchange for public disclosure. They include one or more claims, which define the protection conferred by the patent, and usually contain references to earlier patents or scientific documents. References in a patent signal the state of the art on which the patent is based, and they can limit the property rights established by its claims. A patent that is cited by many others, thus, certainly includes some technology central to further developments. Hence, patent citations can be used to study patented breakthrough inventions, identifying them as highly cited patents (Singh and Fleming, 2010; Castaldi et al, 2015). In this chapter we will add to this strand of literature by analyzing the determinants of a delay in the recognition of highly cited patents.

The structure of the chapter is as follows. Section 2 summarizes the previous work on breakthrough inventions and delayed recognition. Section 3 introduces the methodology of the study, and Section 4 presents the main characteristics of SBs. Finally, Section 5 provides some conclusive remarks.

2 Literature

2.1 Breakthrough inventions

Literature on innovation defines two sizes of inventive steps: incremental and radical. While the difference between them can be difficult to measure in practice, it is widely accepted that radical innovations are the core of technological progress and wealth creation (Schumpeter, 1934; Ahuja and Lampert, 2001). Many studies try to identify the determinants of breakthroughs in order to understand the underlying mechanisms of technological progress.

Breakthrough inventions have traditionally been linked to lone inventors, since individuals prefer to create new systems rather than to improve the systems of others (Schumpeter, 1934; Hughes, 2004). Several problems arise in creative teams, such as idea blocking, communication problems, and personal tensions (Mullen et al, 1991; Paulus and Nijstad, 2003). These problems are also linked to work in big firms, often tied by bureaucracy and the constraint to look for immediate benefits (Ahuja and Lampert, 2001). Nonetheless, the best environment for breakthroughs is unclear. Jewkes et al (1969) signalled that breakthrough inventions originated in large corporations tend to come from the work of a single outstanding individual. Nonetheless, due to the increasing costs of R&D, large corporations provide the most fruitful source of inventions (Jewkes et al, 1969). On the other hand, other studies on teamwork have also found evidence that collaborative research is less likely to produce very unsuccessful outcomes, and at the same time more likely to produce very successful ones (Singh and Fleming, 2010).

Another source of variability lays in the scientific origin of inventions. Several studies point out that basic science is a requisite for breakthrough invention in many technological areas (Valentin and Jensen, 2002; Melese et al, 2009; Malva et al, 2015). The reason is usually that breakthroughs are likely inventions that reside outside the boundaries of the current technological paradigm, yet often inspired by recent scientific advances (Malva et al, 2015). Thus, basic scientific capabilities are more likely to generate the unexpected outcomes that lead to a breakthrough invention (Sobrero and Roberts, 2001).

2.2 Delayed recognition and sleeping beauties in science

Some technological breakthroughs occur much later than the time at which the original patent was granted. In this context, we will speak of sleeping beauties, The question why sleeping beuaties occur in technological development is a new one. Yet, in studies about the dynamics of scientific research, sleeping beauties have already been analyzed. In science, sleeping beauties in science are considered to be articles that go unnoticed for a long time and then, almost suddenly, attract a lot of attention. They were first named by van Raan (2004), although the phenomenon of delayed recognition had been known to science for a long time (Barber, 1961; Cole, 1970; Stent, 1972). There can be several reasons for a paper to become a sleeping beauty (Li et al, 2014): for instance, publishing in the wrong journals (targeting an audience that is not interested), the reputation of the author, or presenting a groundbreaking theory. This last is the case of researchers that were ahead of their time.

In science, a researcher can be ahead of her time if she publishes a theory inconsistent to the established theory (the case of Mendel) or if she researches in an early field (the case of string theory presented by van Raan, 2004). This effect is reinforced when the research cannot be continued due to a delay in technological discovery. Indeed, Lachance and Lariviere (2014) found that sleeping beauties behave differently according to their science field. In arts, humanities and social sciences, sleepers had a stronger start (received many more citations in the early period after awakening) and a weaker finish than in technical sciences (biology, physics, etc).

Sleeping beauties are a particular case of delayed recognition. A paper suffers from delayed recognition if its peak of citations starts later than usual. Sleeping beauties are not only sleepers (the equivalent to delayed recognition in our metaphor), they also need to be beauties. That is to say, they have to be highly cited in their awakening period.

3 Methods

3.1 Patent data

Patents have been extensively used to measure innovation. Albeit all their flaws, they contain very detailed information, and patent applications need to fulfill two key requirements to be granted (novelty and applicability), that make them useful as innovation indicators. Moreover, patent citations have been linked to patent value since Trajtenberg (1990) pointed them out in their seminal paper in other empirical studies (i.e. Albert et al, 1991; Harhoff et al, 1999; Sampat and Ziedonis, 2005). Moreover, citations to patents have also been used to identify breakthrough inventions (Castaldi et al, 2015).

A breakthrough invention is usually defined as one that provides a framework for many subsequent inventions, contributing disproportionally to further technological development (Fleming, 2001). A highly cited patent is likely to be a breakthrough invention. References in a patent have legal binding, and they summarize the state of the art on which the patent is based (Hall et al, 2005; Harhoff et al, 2003). Hence, a patent that is cited in the reference list of many subsequent patents can be considered as a patent that has paved the way for a new trajectory of technological development.

Our data on patents is extracted from the Patstat (October 2012) database. We consider patent families,[1] together with the priority date (first application date in the family), as units of study. Considering patent families allows to group those patents that have been applied for in several patent offices, reducing the number of duplicates in the citations lists. In order to propose a conservative identification strategy of highly cited patents, we only consider those patent families that have received at least one citation over their lifetime.

The initial database is formed of 14,835,792 patent families, with priority years ranging from 1782 to 2011. In order to limit the variation in the data due to changes in the structure of society, we consider only families with recent priority dates. We use as a cutting point the year of creation of the World Intellectual Property Office (WIPO), 1967. Nonetheless, other relevant dates could have been chosen.[2] As the number of patents published per year increases over time and earlier years are poorly covered, the final number of patents is 11,913,774 families, 80% of the database, despite using only 45 years out of several centuries of

[1] From now on, we use "patent" and "patent families" interchangeably.

[2] Results remain qualitatively identical for different starting dates, including 1782 (the first year in the database), 1972 (the first quartile of the distribution of priority years), and 1978 (the year of the first PCT).

data.

All citations from the patents in one family to the patents in another family are only counted once, as one citation from the citing family to the cited family. Due to legal issues in the lifetime of patents, sometimes a patent family can cite a family with a later priority date. These cases are not very common (they only account for 0.77% of the total citations), so we can safely dismiss them.

3.2 Defining sleeping beauties

The distribution of the number of citations that patents receive is depicted in Table 1. The distribution of citations is, as expected, close to exponential: out of the patents that have received at least one citation, more than one third have received exactly 1 citation, and more than half of them have received 2 or less citations. Following the tradition in breakthrough research (Ahuja and Lampert, 2001; Singh and Fleming, 2010), we define highly cited patents as those in a top quantile of the distribution. For this analysis, we chose the top 10% of the citation distribution, because it is quite a robust number across different initial years of the database.[3] That is to say, we consider that a patent is highly cited if it has received more than 12 citations over its lifetime.

Table 1: Quantiles of the citations distribution

Sample fraction	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Quantile	1	1	1	1	2	2	3	4	7	12	2,363

We define that a patent is sleeping for a year if it does not receive any citations for that year. Thus, a patent is sleeping until it receives a citation. We could relax this definition by including years in which it received one (or few) citations, as van Raan (2004) did with scientific papers. Nonetheless, given that most patents receive very few citations, allowing for citations during the sleeping period would transform most of the patents in sleepers. The distribution of sleep lengths (the number of years before the first citation) is shown in Table 2. Some patents receive their first citation during the same year of their application. For half of the patents, the first citation arrived 3 or less years after their priority date.

As before, we define a sleeper as a patent family in the top 10% of the sleep length distribution. That is

[3] The 90% quantile of the citation distribution is 12 citations, regardless whether the initial years of the database is 1782, 1967, 1972 or 1978.

to say, a patent is a sleeper if it received its first citation more than 13 years after its priority date.[4]

Table 2: Quantiles of the length of sleep distribution

Sample fraction	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Quantile	0	1	1	2	3	3	4	6	8	13	45

A sleeping beauty is defined as a patent family that is both a sleeper (is not cited for at least 13 years after its priority date) and a highly cited patent (receives at least 13 citations). According to this definition, the database contains 3,196 sleeping beauties. Their priority years range from 1967 to 1996, almost the whole range of 14-year-old patent families in our database.[5] Their sleep periods range from 14 to 38 years, and their total number of citations range from 13 to 164. Figure 1 shows the distribution of SB families across sleep lengths and number of citations.

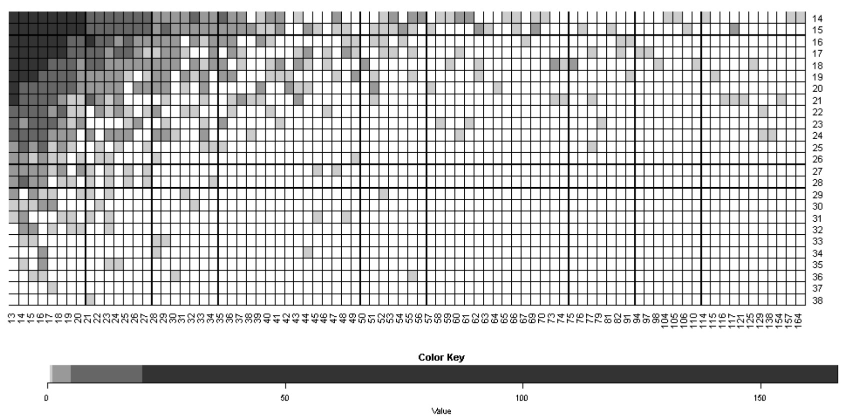


Figure 1: Distribution of sleeping beauties over sleep length (y axis) and number of citations (x axis). Darker colors indicate higher frequencies.

Figure 1 shows that SBs are not uniformly distributed. Those in the right side of the table are more cited ("more beautiful"), while those in the lower rows have experienced longer delay in their recognition (were "more sleepy"). Moreover, there are many more SBs in the left top corner than in the rest of the table. Some studies (van Raan, 2004, for instance) have defined a sleeping beauty not depending on the lifetime citations but depending on the number of citations during a fixed period after awakening. That would require the definition of an additional parameter, the length of the awakening period. Theoretically, this would allow a more accurate comparison between the older and newer patents. Nonetheless, 12 citations comes as a very

[4] The 90% quantile of the distribution of sleep length varies from 20 years of sleep if the database starts in 1782, to 13 (1967), 12 (1972) and 11 (1978).

[5] Since the database includes 2011, sleeping beauties could have had priority year 1997 in theory. Nonetheless, there was no patent from 1997 that received no citations until 2011 and more than 12 citations in 2011.

consistent 90% quantile of the number of citations overall, independently of the starting year of our database. Thus, we feel pretty confident that those patents with more than 12 citations are the highly cited patents (HCPs). Moreover, all SBs are grouped together, without consideration for their intensity: we consider that a sleeping beauty is a patent that slept for 14 years and received 13 citations as well as a patent that slept for 24 years and received 154 citations. Thus, defining different awakening periods will not change their identification in our study.

4 Results

4.1 Descriptive analysis

This section compares the population of SBs against the population of HCPs, across several typical determinants of patent citation intensity (Harhoff et al, 2003): technological classes, number of references, and number, country, and experience of of the authors. These univariate comparisons, are a first, exploratory approach to the empirical analysis of the determinants of SBs. Results of this exploratory analysis have to be taken with caution since they are unidimensional comparisons. In the next section we will introduce all dimensions in a multivariate analysis. Results shown in this section are all for full counts, but they do not change qualitatively if we consider fractional counts.

The first dimension through which we compare SBs and HCPs are technological classes, measured through IPC codes. IPC codes in patents are assigned by the examiners, based on potential applicability. They are grouped in eight main classes, coded from A to H: human necessities (A), performing operations and transporting (B), chemistry and metallurgy (C), textiles and paper (D), fixed constructions (E), mechanical engineering, lighting, heating, weapons and blasting (F), physics (G) and electricity (H). Here, we expect that patents from technology classes that are more science-based are less likely to be SBs, because knowledge in these fields is more codified, allowing it to diffuse more freely and rapidly (Foray, 2004).

Figure 2 shows the distribution of SBs (black bars) and HCPs (grey bars) across the eight technological classes, in percentages of the full count (every patent can be assigned to one or more IPC, so that patents with many technological classes are counted once in every class), although the results are not qualitatively different for the fractional counting case. To compare both distributions, we run a χ^2 goodness-of-fit test. It confirms ($X^2 = 488.6459, df = 7, p - value < 0.001$) that both distributions are statistically different. SBs are more probable than HCPs in most technological classes, except in classes G and H, that is to say physics and electricity. Physics (G) and Electricity (H), together with chemistry and metallurgy (C), are the most science-based technological classes (Azagra-Caro et al, 2006). This comparison suggests that breakthroughs in science-based technologies, in particular physics and electricity, are indeed more likely to be recognized as such earlier in their lifetime.

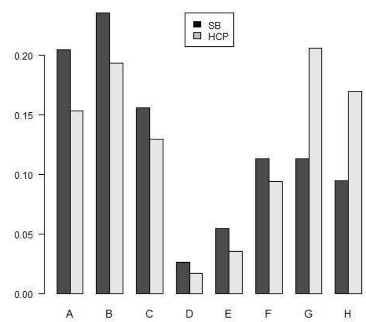


Figure 2: Technological classes

A second dimension of comparison is the number of references cited in the patents. Since the list of references in a patent can limit the protection of its claims, one would expect inventors and applicants to cite as little prior patents as possible. In order to correct this incentive not to disclose information, in some patent offices (like the USPTO) an opposition to a patent will be more likely to be successful in court if the application of the opposed patent did not include a citation to the opposing patent. Thus, the incentives are balanced for an applicant to disclose the list of prior knowledge of the patent. We expect that SBs have fewer references than other HCPs, since SBs deviate more from the current paradigm than other HCPs. Figure 3 shows the cumulative distribution of the number of references per patent for the SBs (black line) and the HCPs (grey line) groups. A Mann-Whitney U test ($W = 451183786, p - value < 0.001$) shows that both distributions are significantly different. SBs tend to have less citations than HCPs: half the SBs have 4 references or less, while half the HCPs have less than $7 - 8$ references, almost the double. This could indicate that SBs are breakthrough inventions outside the existing technological paradigm: less related to existing technologies, and therefore, with less references.

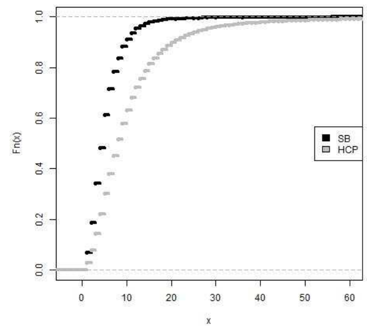
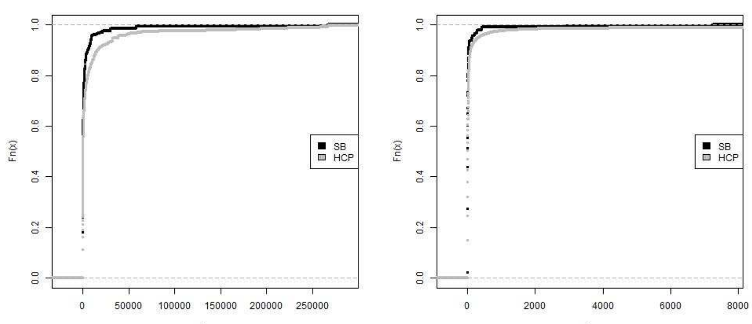


Figure 3: Number of references

Another dimension in which we expect differences between SBs and HCPs is the experience of inventors and applicants.[6] Experienced inventors and applicants that have filed many applications (or made many inventions) are more known in the technological community. Hence, one can expect that their inventions are noticed earlier. A group of authors will likely profit of the experience of the most experienced one. Thus, we consider the experience of the authors of a patent as the highest in the team. That is to say, the experience of the inventors of a patent is equal to the maximum number of patents invented by any of its inventors. Likewise, the experience of the applicants of a patent is the maximum number of patent applications that any of its applicants has filed.



(a) Experience of the applicants *(b) Experience of the inventors*

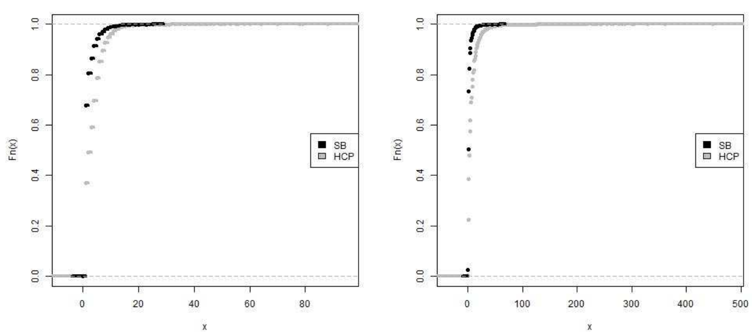
Figure 4: Experience of the authors

Figure 4 shows the distribution of the experience of the authors, both for applicants (Figure 4a) and for inventors (Figure 4b). The experience of an applicant (inventor) is the number of patents she has applied for (invented), independently of the number of citations it received. The distribution of experience of the

[6]Information about the authors can be found in Patstat.

authors is extremely skewed. The vast majority of the authors have a limited experience: around 20% of the applicants have only filed one patent application, and around 25% of the inventors have invented one or two other patents. On the other hand, there is one applicant that has filed over 300,000 patent applications, Matsushita Electric Inc Co Ltd. Four of those are in the SB group, and 2,145 of them are in the HCP group.

A Mann-Whitney test shows that the distribution of experience of the authors for SBs and HCPs are different, both for applicants ($W = 21606.5, p - value < 0.001$) and for inventors ($W = 10927.5, p - value < 0.001$). The authors of SBs tend to be less experienced than the authors of HCPs (the SB black line is above the HCP grey line). This could indicate that experienced authors achieve earlier recognition of breakthrough inventions. Nonetheless, very big players also produce some SBs, which is to be expected due to their sheer size.

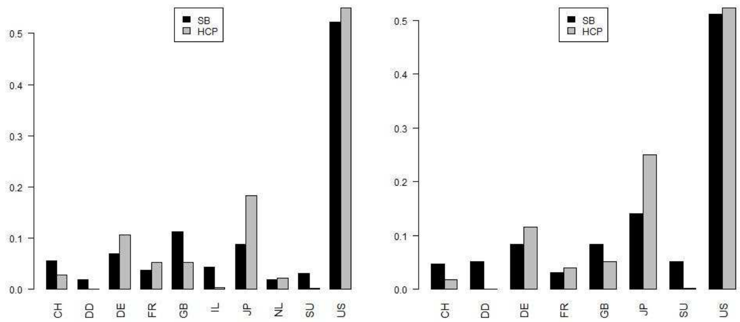


(a) Number of applicants *(b) Number of inventors*

Figure 5: Number of authors

Patents that are the result of team work are expected to achieve earlier recognition, as each team member is a source of diffusion. Figure 5 shows the distribution of the number of applicants (Figure 5a) and number of inventors (Figure 5b) per patent. A Mann-Whitney U test shows that the distribution of number of authors per patent differ from SBs to HCPs, both for applicants ($W = 801964016, p - value < 0.001$) and for inventors ($W = 707462148, p - value < 0.001$). The number of authors per patent is lower for SBs (the black line is higher). This could mean that lone inventors are more likely to create breakthroughs outside the technological paradigm, but it could also indicate that lone inventors lack the social network of authors through which an invention gets diffused.

Every applicant or inventor has a country code that can be associated to the patent country of origin. In



(a) Country of the applicants *(b) Country of the inventors*

Figure 6: Country of the authors

case of multiple applicants or inventors, every additional country of origin is added to the countries of the patent family, and every patent family is duplicated for every country it has.

Figure 6 shows the distribution of countries of the SBs and HCPs (only countries with a share higher than 0.005% are shown). A χ^2 goodness-of-fit test shows that the distribution of countries is different for SBs and HCPs, both for applicants ($X^2 = 836.8716, df = 174, p - value < 0.001$) and for inventors ($X^2 = 852.7427, df = 186, p - value < 0.001$). The result shows that there are less SBs originating in Japan than their share of HCPs would suggest. The Japanese patent system has some peculiarities, compared to the rest of the systems, that could account for this difference. For example, patents in Japan follow a "one patent, one claim" policy. This could indicate that highly cited patents in other patent systems are made of many important claims, that become highly cited patents in the Japanese system.

4.2 Regression analysis

In this section, the previous dimensions of comparison between SBs and HCPs are introduced in a multivariate analysis through a regression of the probability that a breakthrough is recognized with delay. The dependent variable takes value 1 if the patent is a SB and 0 if it is a HCP in the control group. The population consists of $735,653$ patents for which we have all information, of which $2,136$ are SBs and $733,517$ are HCPs. The suitable regression model for this kind of data is a logistic regression.[7]

[7] SBs account for 0.29% of the whole population in our database, what statistics call a rare event. Methods of logistic regressions for rare events data have been recently developed by King and Zeng (2001), but they are only adequate when the data is a sample of the entire population. Since our database includes the whole population of patents with more than 12 citations, a regular logistic regression is more suitable for our study.

Two new control variables are added to some models: application year, to account for any time trend; and number of technological classes, to separate the transversal technologies that can be relevant to many domains. The effect of the country of the authors, although interesting, is left for further analyses. Independent variables are separated in two groups: the characteristics of the patents and the IPC. We run a regression for each group and another with all dependents included (models 1-3). IPC classes are non-exclusive, so we run a second group of regressions (models 4 and 5) with a fractional count of the IPC classes that patents belong to (every patent is weighted down as many times as different IPC it is assigned to), with reference technological class Electricity (H). Results of the logistic regressions are summarized in Table 3.

	(1)	(2)	(3)	(4)	(5)
Priority year	-0.175^{***}	-0.199^{***}	-0.172^{***}	-0.201^{***}	-0.173^{***}
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Nr IPC	-0.250^{***}				
	(0.049)				
Nr references	-0.059^{***}		-0.062^{***}		-0.063^{***}
	(0.006)		(0.006)		(0.006)
Nr applicants	-0.176^{***}		-0.174^{***}		-0.178^{***}
	(0.016)		(0.016)		(0.016)
Nr inventors	-0.023^{**}		-0.024^{***}		-0.025^{***}
	(0.009)		(0.009)		(0.009)
Applicant exp	-0.000005^{***}		-0.00000		-0.00000
	(0.00000)		(0.00000)		(0.00000)
Inventor exp	0.00000		-0.00000		-0.00000
	(0.00003)		(0.00003)		(0.00003)
IPC A		0.024	0.104	0.898^{***}	0.822^{***}
		(0.067)	(0.067)	(0.089)	(0.089)
IPC B		-0.374^{***}	-0.237^{***}	0.411^{***}	0.431^{***}
		(0.061)	(0.061)	(0.090)	(0.090)
IPC C		-0.280^{***}	0.003	0.520^{***}	0.706^{***}
		(0.069)	(0.071)	(0.103)	(0.102)
IPC D		-0.210	0.086	0.561^{***}	0.821^{***}
		(0.129)	(0.129)	(0.178)	(0.176)
IPC E		-0.211^{**}	-0.145	0.609^{***}	0.538^{***}
		(0.093)	(0.093)	(0.118)	(0.118)
IPC F		-0.396^{***}	-0.238^{***}	0.347^{***}	0.389^{***}
		(0.073)	(0.073)	(0.101)	(0.101)
IPC G		-0.849^{***}	-0.733^{***}	-0.219^{**}	-0.216^{*}
		(0.077)	(0.077)	(0.112)	(0.110)
IPC H		-0.702^{***}	-0.580^{***}		
		(0.081)	(0.081)		
IPC count	*Full*	*Full*	*Full*	*Fractional*	*Fractional*
Observations	735,653	735,653	735,653	735,653	735,653
Log Likelihood	-11,962.810	-12,142.570	-11,865.620	-12,168.460	-11,872.540

Note: *p<0.1; **p<0.05; ***p<0.01

Table 3: Logistic regressions of the probability that a breakthrough is a SB

The effect of the priority year is very consistent across specifications: later patents are less likely to be SBs than HCPs. One important point is that this effect is not due to HCPs being younger patents in our database, since all patents, both in the SB population and in the control group, have priority dates in 1967-1996.

The exploratory analysis showed that SBs tend to cite less references than HCPs. Indeed, this result holds in the multivariate analysis. As already suggested, patents that cite few earlier references are less connected to already existing technologies. Thus, patents that are outside the existing technological paradigm are more likely to be recognized as breakthroughs with delay.

As suggested by the exploratory analysis, both the number of applicants and the number inventors have a negative effect on the probability to produce a SB. HCPs produced by big teams are less likely to experience delayed recognition. This can indicate either that the SBs are developments at the technological frontier, that need big teams, or that the social network of big teams is wider and facilitates the diffusion of a breakthrough more easily than the smaller network of lone inventors.

Patents assigned to several technological classes are less likely to be unrecognized as breakthroughs. Since they have more potential applications, it is logical that some class or another will recognize them immediately. Moreover, IPCs are assigned by professional examiners. That SBs have less IPC classes might also indicate that examiners ignore their future impact and applications.

Finally, patents in the technological classes G (physics) and H (electricity) are much less likely to remain unrecognized as breakthrough for a long time, compared to patents in other classes. This is evidenced by the fact that their coefficients are bigger than those of the other IPCs in models 2 and 3, and that the effect of all IPCs excepted G are positive with respect to the reference class H in models 4 and 5.

5 Conclusions

This study carries out an exploratory analysis of technological sleeping beauties (SBs). To do so, it compares SBs to other breakthroughs in the same period across several dimensions, first in univariate analyses, and then through a multivariate regression analysis. It points out some distinctive features of SBs. Some are in line with the hypothesis that SBs are ahead of their time and remain latent until complementary technologies have been developed, while others features support the alternative hypothesis that the social network of inventors affects diffusion. Indeed, the fact that SBs have less references and are often (mis)classified in only few technology classed points to patents being ahead of their time. At the same time, SBs more often have few inventors or applicants, supporting the alternative hypothesis.

This study is not exempt of limitations. Mainly, this remains a purely exploratory analysis of the characteristics of breakthroughs with delayed recognition. It does not provide a clear theory as to why SBs appear, whether they result from a failure of the patent system or they are a natural consequence of technological development. This study only presents a first, descriptive approach to SBs and their features through the analysis of some empirical trends of their population.

The examination of SBs can be expanded in several interesting directions. The exploratory analysis could be expanded by a social network analysis of the authors of SBs, to identify potential patterns in their position on the overall network of authors. Moreover, a further step could be made in categorizing the citing patents of SBs. For example, SBs can be the result of migrating technologies, those that got an unexpected application in a field different from the one they were oriented to at first. Analyzing the citing patents can lead to an interesting further development: the study of the "awakening" patent, what is called the prince of the SB in scientific literature (van Raan, 2004). A network analysis both of this patent (in the network of citations) and of the authors (in the network of coauthorships) can point out whether the characteristics of the prince are crucial for the awakening of the SB. If not, it may be that SBs are a product of fashions or trends in the technological direction of industrial development.

Ke et al (2015) have suggested some kind of a "SB index", a measure of how beautiful and sleepy a document is. According to such an index, most of the patents in our study would barely be SBs, while some of them (like the patent that received 55 citations after 36 years sleeping, or the one with 138 citations that slept for 24 years) would score very high. As one changes the definition of SBs to higher degrees of this index, less documents fulfill the SB condition, which accounts for the darker top-left corner of Figure 1. In fact, Ke et al (2015) found that this distribution of quantity of SBs in science depending on the index follows a power-law distribution. For a further study we could further group SBs in several levels, according to the total number of citations and the length of the sleep.

References

Ahuja G, Lampert CM (2001) Entrepreneurship in the large corporation: a longitudinal study of how established firms create breakthrough inventions. Strategic Management Journal 22(6-7):521–543

Albert M, Avery D, Narin F, McAllister P (1991) Direct validation of citation counts as indicators of industrially important patents. Research Policy 20(3):251–259

Azagra-Caro JM, Yegros-Yegros A, Archontakis F (2006) What do university patent routes indicate at regional level? Scientometrics 66(1):219–230

Barber B (1961) Resistance by Scientists to Scientific Discovery: This source of resistance has yet to be given the scrutiny accorded religious and ideological sources. Science 134(3479):596–602

Castaldi C, Frenken K, Los B (2015) Related Variety, Unrelated Variety and Technological Breakthroughs: An analysis of US State-Level Patenting. Regional Studies 49(5):767–781

Cole S (1970) Professional Standing and the Reception of Scientific Discoveries. American Journal of Sociology 76(2):286–306

David PA (1985) Clio and the Economics of QWERTY. The American Economic Review 75(2):332–337

Fleming L (2001) Recombinant Uncertainty in Technological Search. Management Science 47(1):117–132

Foray D (2004) Economics of knowledge. MIT Press, Cambridge, Mass

Hall BH, Jaffe A, Trajtenberg M (2005) Market Value and Patent Citations. The RAND Journal of Economics 36(1):16–38

Harhoff D, Narin F, Scherer FM, Vopel K (1999) Citation Frequency and the Value of Patented Inventions. Review of Economics and Statistics 81(3):511–515

Harhoff D, Scherer FM, Vopel K (2003) Citations, family size, opposition and the value of patent rights. Research Policy 32(8):1343–1363

Hughes TP (2004) American genesis: a century of invention and technological enthusiasm, 1870-1970. University of Chicago Press, Chicago

Jewkes J, Sawers D, Stillerman R (1969) The sources of invention, 2nd edn. The Norton library N502, W. W. Norton, New York

Ke Q, Ferrara E, Radicchi F, Flammini A (2015) Defining and identifying Sleeping Beauties in science. Proceedings of the National Academy of Sciences 112(24):7426–7431

King G, Zeng L (2001) Logistic Regression in Rare Events Data. Political Analysis 9:137–163

Lachance C, Lariviere V (2014) On the citation lifecycle of papers with delayed recognition. Journal of Informetrics 8(4):863–872

Li S, Yu G, Zhang X, Zhang Wf (2014) Identifying princes of Sleeping Beauty - knowledge mapping in discovering princes. In: 2014 International Conference on Management Science & Engineering (ICMSE), IEEE, pp 912–918

Malva AD, Kelchtermans S, Leten B, Veugelers R (2015) Basic science as a prescription for breakthrough inventions in the pharmaceutical industry. The Journal of Technology Transfer 40(4):670–695

Melese T, Lin SM, Chang JL, Cohen NH (2009) Open innovation networks between academia and industry: an imperative for breakthrough therapies. Nature Medicine 15(5):502–507

Mullen B, Johnson C, Salas E (1991) Productivity Loss in Brainstorming Groups: A Meta-Analytic Integration. Basic and Applied Social Psychology 12(1):3–23

Paulus PB, Nijstad BA (2003) Group Creativity. Oxford University Press

van Raan AFJ (2004) Sleeping Beauties in science. Scientometrics 59(3):467–472

Sampat BN, Ziedonis AA (2005) Patent Citations and the Economic Value of Patents. In: Moed HF, Glanzel W, Schmoch U (eds) Handbook of Quantitative Science and Technology Research, Kluwer Academic Publishers, Dordrecht, pp 277–298

Schumpeter JA (1934) The theory of economic development: an inquiry into profits, capital, credit, interest, and business cycle. Harvard University Press, Cambridge

Singh J, Fleming L (2010) Lone Inventors as Sources of Breakthroughs: Myth or Reality? Management Science 56(1):41–56

Sobrero M, Roberts EB (2001) The Trade-off Between Efficiency and Learning in Interorganizational Relationships for Product Development. Management Science 47(4):493–511

Stent G (1972) Prematurity and uniqueness in scientific discovery. Scientific American 227(6):84–93

Trajtenberg M (1990) A Penny for Your Quotes: Patent Citations and the Value of Innovations. The RAND Journal of Economics 21(1):172

Valentin F, Jensen RL (2002) Reaping the Fruits of Science: Comparing Exploitations of a Scientific Breakthrough in European Innovation Systems. Economic Systems Research 14(4):363–388