Sleeping beauties gain impact in overdrive mode

Anthony F. J. van Raan¹

Received: 13 October 2020 / Accepted: 5 February 2021 / Published online: 2 March 2021
© The Author(s) 2021

Abstract

In this study we focus on characteristics of SBs that have not or hardly been investigated previously. We find that the choice of the awakening period in the selection of SBs has consequences for the measured citation patterns. Focusing on medical SBs we analyze patterns in the time-development of the citation impact of SBs; the influence of self-citations on the awakening process; and the occurrence of medical research fields to which the SBs and their citing papers belong. An important finding is that SBs are generally characterized by a sleep that becomes less and less deep instead of a permanent deep sleep. The sleeping period is followed by a phase-transition-like jump as a start of the awakening period and a remarkable regularity is found for the citation impact immediately before and after the jump.

Keywords Sleeping beauties · Citation patterns · Choice of awakening period · Jump to awakening · Occurrence of research fields · Self-citations

JEL Classification 033

Mathematics Subject Classification 62B86 · 62P25 · 91C99 · 94A99

Introduction and background

In science, a Sleeping Beauty is a publication that goes largely or completely unnoticed (‘sleeps’) for a long time and then, almost suddenly, attracts a lot of attention (‘is awakened by a prince’). Garfield focused the attention on the phenomenon of ‘delayed recognition’ (Garfield 1970, 1980, 1989, 1990) which was linked to ‘premature discovery’ or ‘being ahead of time’ (Stent 1972). We refer to our earlier work (van Raan 2004, 2015, 2017a, b) for an overview of the literature on Sleeping Beauties (SBs) and for an extensive analysis of SBs in physics, chemistry, engineering and computer science. Often it is thought that SBs relate to ‘exotic’, rather other-worldly research. But the opposite is the case. More than half of the SBs appeared to be application-oriented. In follow-up work (van Raan and Winnink 2018, 2019) we investigated whether the Sleeping Beauties in

✉ Anthony F. J. van Raan
vanraan@cwts.leidenuniv.nl

¹ Centre for Science and Technology Studies, Leiden University, Leiden, The Netherlands

physics, chemistry, engineering, computer science and also in medical research are cited in patents, i.e., SBs that appear as scientific non-patent references (SNPR) in patents (van Raan 2017a, b). We found that the time lag between the publication year of the SB-SNPRs and their first citation in a patent was becoming shorter in recent years. After our work on SBs in physics, chemistry, engineering, computer science we continued our research on medical SBs. Medical research is characterized by a strong interdisciplinarity and therefore it is important to investigate the role of medical SBs in this context. This issue is discussed in Section “In which medical research fields are SBs published?”.

The study of Sleeping Beauties has attracted considerable attention in recent times. We refer to a number of publications that are relevant to the context of this paper, particularly studies on the characterization of SBs on the basis of their citation patterns. Such studies have been conducted by Braun et al. (2010) and particularly by Li and Ye (2012) and Li (2014) who analyzed the phenomenon of ‘all-elements-sleeping-beauties’ which includes a short awakening immediately after publication followed by a sleeping period and then a second awaking period. Li et al. (2014) developed the idea of a “heart-beat spectrum” (series of annual citation numbers) for SBs, with results that are relevant for our work; we will return to this topic in Section “Citation impact of SBs in the years after publication”.

Specific attention to the citation dynamics in a multidisciplinary context is given by Ke et al. (2015) who introduced a parameter-free measure (the ‘beauty coefficient’) that quantifies the extent to which a specific paper can be considered an SB. Li and Ye (2016) discussed three criteria to distinguish SBs in which early citations are ‘penalized’. Li and Shi (2016) studied the awakening of SBs in ‘genius work’, i.e., scientific articles of which the citations grow exponentially in a very long period, for example, 50 years. He et al. (2018) studied atypical articles whose citation trajectories do not follow the normal rise-and-fall pattern. Ye and Bornmann (2018) and Bornmann et al. (2018) investigated the differences between papers that received many citations shortly after publication and SBs. Fang (2018) characterizes the citation histories of SBs by a derivative analysis of citation curves. El Aichouchi and Gorry (2018) describe on the basis of a remarkable SB the importance of continuity in research in ensuring the acceptance of unpopular hypotheses and the development of new research fields. Du and Wu (2018) focused on the characteristics of not-so-highly cited SBs. A new development is the analysis of sleeping-beauty-like behavior of patents, see Hou and Yang (2019). Also altmetrics is now used in the study of SBs, see the recent work of Hou et al. (2020) and Hou and Yang (2020). In most cases, SBs do not represent ‘isolated’ work, Zhang and Ye (2020) study the phenomenon of groups of publications associated with SBs.

In this study we focus on medical SBs and analyze characteristics of SBs that have not or hardly been investigated previously: patterns in the time-development of the citation impact of SBs, particularly the influence of the chosen awakening period on the citation patterns; characteristics of the start of the awakening period; influence of self-citations on the awakening process; and the occurrence of medical research fields to which the SBs and their citing papers belong.

The structure of this paper is as follows. First we discuss our measurement method and parameter calculation. Next we address the development of the number of SBs from 1980. We continue our paper with a discussion of the characteristics of the citation impact of SBs in the years after publication, including the role of self-citations. A further section focuses on the medical research fields in which the SBs and their citing papers are published. We conclude with an overview of our findings and issues for further research.

Method of measurement and parameter calculation

For the identification of SBs different approaches exist. For instance, the above mentioned work of Ke et al. (2015) is an important contribution to the study of SBs. Their approach with the ‘beauty coefficient’ allows to study SBs for a very long time period. The advantage of this approach is that it is parameter-free, entirely dependent on the complete citation history of a paper and thus independent of specific rules. Our SB detection method is based on specific rules but this makes it possible to accurately determine the influence of specific sleep and awakening periods and particularly the transition from sleep to awakening, a major issue in this paper.

How do we find Sleeping Beauties (SBs) with specified sleeping periods and specified awakening intensities? Our measurement and calculation procedure is as follows. With a fast and efficient search algorithm written in SQL applied to the CWTS enhanced Web of Science (WoS) database (van Raan 2015) we measure for all publications (article types: article, review and letters) in the database, starting with publication year 1980 and published in journals of the medical fields (in total around 15,000,000 publications), the annual number of citations $C(t)$ corrected for self-citations after publication until 2019 and calculate for all these publications the following parameters:

1. The average number of citations c_s during the time interval $[t=0, t=s]$ where $t=0$ is the publication year; we take $s=5, 6, 7, \dots, 20$:

$$c_s = \left[\sum_{t=0}^s C(t) \right] / s \quad (1)$$

This is the citation rate during the given time interval of s years immediately after publication;

2. The average number of citations c_a during the time interval $[t=s+1, t=s+a]$, in most cases we take $a=5$ or 10 :

$$c_a = \sum_{(t=s+1)}^{(s+a)} C(t) / a \quad (2)$$

This is the citation rate during the given time interval of a years after the initial time interval of s years.

So far the procedure with *all* medical publications (of the above specified article types) in the database. The next step is the identification of SBs. This goes as follows. For all time intervals $[t=0, t=s]$ with $s=5, 6, 7, \dots, 20$ and subsequent time intervals $[t=s+1, t=s+a]$, for $a=5$ as well as $a=10$, we identify the papers with the threshold values $c_s \leq 1.0$ and $c_a \geq 5.0$. The papers that meet these conditions are characterized by a very low citation rate during a period immediately after publication of s years, followed by a sudden rise of the citation rate during a period of a years. These are our Sleeping Beauties. We call the time interval $[t=0, t=s]$ of s years the sleeping period and the time interval $[t=s+1, t=s+a]$ of a years the awakening period. We can also choose other threshold values for the citation rates, for instance $c_s \leq 0.5$ and $c_a \geq 10.0$. Also we can apply maximum and minimum values for the awakening period a , i.e., $a(\min)$ and $a(\max)$. The above procedure means that we can adjust the SB search algorithm with 4 parameters:

Fig. 1 Number of SBs by publication year for sleeping periods $s=5$ (panel A) and 10 (panel B), for both awakening periods $a=5$ and 10, as well as for the two sleeping citation intensities $c_s \leq 1.0$ and ≤ 0.5 ; the black dotted line marks the increase of the total medical literature on a relative scale (see main text). Panel C: number of medical publications covered by the WoS from 1980

1. Length of the sleep period in years immediately after publication: s .
2. Sleep intensity in terms of the citation rate during the sleeping period: c_s .
3. Length of the awakening period in years after the sleeping period: a .
4. Awakening intensity in terms of the citation rate during the awakening period: c_a .

We define c_s between 0.0 and 0.5 as a very deep sleep and c_s between 0.6 and 1.0 as a deep sleep. If we take $c_s \leq 1.0$ we cover the whole range from $c_s = 0$ to $c_s = 1.0$. In this study we use 5- and 10-years awakening periods, i.e., $a(\min)=a(\max)=5$ or 10 (for the sake of brevity we write in the rest of this paper $a=5$ instead of $a(\min)=a(\max)=5$, etc.). Obviously, the longer the sleeping period, the older the SBs will be. For instance, if $s=20$ and $a=5$, we need a total time period of $20+5=25$ years and given that 2019 is the last year of the citation measurement, only SBs with $s=20$ published between 1980 and 1995 can be identified. On the other hand for SBs with $s=5$ and $a=5$ we need $5+5=10$ years and thus 2010 is in this case the last publication year of the measurement. In this study we mainly investigate SBs with sleeping periods between 5 and 10 years.

Development of the numbers of SBs from 1980

In a foregoing publication (van Raan and Winnink 2019) we found that the number of SBs measured with an awakening period of 5 years ($a=5$) did not or hardly increase as compared to the annual exponential growth of the total number of WoS-covered medical research publications. Now we also investigate in this study SBs that are more ‘persistent’ in terms of citation impact, i.e., they meet the average awakening intensity requirement $c_a \geq 5.0$ for a period of 10 years ($a=10$). An intriguing question now is whether the number of these more persistent SBs also keeps pace with the general growth of the medical research literature. Or is their longitudinal development different?

In Fig. 1 we show the number of SBs from 1980 for the sleeping periods $s=5$ (panel A) and 10 (panel B), for both $a=5$ and 10, as well as for both sleeping citation intensities $c_s \leq 1.0$ and ≤ 0.5 . In this time trend analysis we use consecutive overlapping 3 year averages. The results for $s=15$ and 20 are shown in Fig. S1 (Supplementary Information). We see that the number of SBs with $a=10$ is larger than the number of SBs with $a=5$. This is to be expected as an awakening period $a=10$ gives a publication more time to reach the required awakening citation average $c_a \geq 5.0$. In Section “Citation impact of SBs in the years after publication” we show a typical example. We also compare the temporal developments with the general increase of the total number of medical publications in the WoS, indicated with the black dotted lines in Fig. 1 (panels A and B). We distinguish two periods: before 2000 (increase approximately 3% per year) and after 2000 (5% per year). These dotted lines are meant to indicate the (exponential) increase of the entire medical literature, on a relative scale chosen in order to fit the lines as good a possible in the figure. Absolute numbers for the entire medical literature are given in Fig. 1, panel C.

At first sight we see that the overall increase of the number of SBs keeps more or less pace with the increase of the total medical literature. But on the other hand, for $s=5$, in both cases $c_s \leq 0.5$ and ≤ 1.0 , we see after 2000 a relatively strong decrease.

Fig. 2 Number of SBs as a function of sleeping period ($s=5$ – 10) for the two sleeping citation intensities $c_s \leq 1.0$ and ≤ 0.5 and for the two awakening periods $a=5$ (publication years 2001–2005) and 10 (publication years 1996–2000)

This is also the case for $s=10$ and $c_s \leq 1.0$. Such a decrease can be expected for our recent time with its greatly increased accessibility of scientific literature. Given however the wave-like character of the earlier time trend, the conclusion that the number of SBs is decreasing since the beginning of the new century may be premature. In the case of $s=10$ and $c_s \leq 0.5$ the numbers are very low, as a consequence the trend in time is very noisy and there is no clear increase or decrease.

This wave-like pattern is intriguing. We repeated our measurements several times with a break of a few weeks, with different awakening periods and awakening citation thresholds keeping the sleeping period constant and always found the same wave-like pattern. We also analyzed the time-development of the number of medical publications in, particularly, the medical fields with a relatively high occurrence of SBs (see Section “In which medical research fields are SBs published?”) both for the publication years of the SBs as well as for the years in which these SBs were going to be cited; we do not find any reason for the wave-like pattern. More specifically, we find no evidence that it is related to journal coverage and particularly journal coverage of the fields in which the SBs frequently occur. A technical artifact is unlikely because all publications from 1980 are measured in one and the same run of the algorithm.

What is the influence of the sleeping period on the number of SBs? In Fig. 2 we show the number of SBs as a function of sleeping period ($s=5$ to 10) in each case for two different sleeping citation intensities $c_s \leq 1.0$ and ≤ 0.5 and for the two awakening periods $a=5$ (publication years 2001–2005) and 10 (publication years 1996–2000), in all cases $c_a \geq 5.0$.

We see that the number of SBs decreases with the sleeping period in all cases with a steep power law, exponent around -3 . So if the sleeping period doubles, the expected number of SBs will be almost an order of magnitude lower. These measurements confirm our earlier observations (van Raan 2004; van Raan and Winnink 2019).

Citation impact of SBs in the years after publication

Time-developments of the citation impact of SBs for $s = 5$ to 10

Recently several papers discussed the dynamics of citation time-developments for different types of papers. In their aforementioned work, Li and Ye (2012) and Li (2014) investigated a special class of SB-like papers that are cited relatively highly immediately after publication in a short period of time, then go back to oblivion (sleep) and wake up again after a long time ('second-act' papers). Wang et al. (2013) developed a mechanistic approach for the citation dynamics of individual papers based on the preferential attachment model in network theory, using parameters from the observed citation history of paper to predict future citations (WSB model). Preferential attachment implies that highly cited papers are more likely to be cited again than less-cited papers. In other words, the probability for a paper to be cited again is proportional to the total number of citations already received by the paper. This phenomenon is also referred to as cumulative advantage. He et al. (2018) found that the WSB model cannot fully explain the observed citation patterns of 'second act' papers and they developed a new model. However, these models cannot predict the citation dynamics of papers where typical 'exogenous' events cause a sudden boost of citations, which is the case in our definition of SBs.

What kind of long-term citation patterns do we see for the papers that we have identified as SBs? In order to investigate this we analyze the time developments of the citation impact of SBs for $s = 5$ to 10 ($a = 5$, $c_a \geq 5.0$). In the case of $c_s \leq 1.0$ we take for $s = 5, 6$ and 7 publication year 2005 and, because of the decreasing numbers of SBs for longer sleeping periods, for $s = 8, 9$ and 10 the publication years 2004–2005. In the case of $c_s \leq 0.5$ we take for $s = 5, 6, 7$ and 8 publication years 2001–2005 and for $s = 9$ and 10 the publication years 2000–2005. Figure 3 shows the trends in the number of citations from the year of publication as the first citation year to 2019 as the last citation year. To enable direct comparison of the citation trends for the different sleeping periods, we have calculated the average numbers.¹ We see three general patterns.

First, the time-dependent citation characteristics are very similar for the SBs with $c_s \leq 1.0$ and those with $c_s \leq 0.5$. A second, remarkable observation is that during the sleeping period we see a slow, approximately linear increase of citations and not a constant zero or constant very low number of citations and also not a random distribution of the very few citations during sleeping period. This finding has an important implication: SBs show a 'weak awakening' already during the sleeping period. This is consistent with the work of Li et al. (2014) on 'heartbeat spectra' (series of annual citation numbers) who found that SBs with most of their citations during the sleeping period in the second half of sleeping period ('late heartbeats') have higher awakening probability than those with most citations during the sleeping period received in the first half of sleeping period ('early heartbeats').

¹ Citation year 1 is equal to the publication year. So for the SBs ($c_s \leq 1.0$) with publication year 2005, the year 2005 is citation year 1, 2019 is citation year 15. In the cases where we combine publication years (for instance for the SBs with $c_s \leq 0.5$ we have taken 2001–2005) the number of citations in citation year 1 is calculated by taking the sum of the number of citations in 2001 for the SBs published in 2001, the number of citations in 2002 for the SBs published in 2002 and so on until the number of citations in 2005 for the SBs published in 2005. Similarly, citation year 15 is calculated by taking the sum of the number of citations in 2015 for the SBs published in 2001, the number of citations in 2016 for the SBs published in 2002 and so on until the number of citations in 2019 for the SBs published in 2005.

Fig. 3 Number of citations for SBs by citation year for $s=5$ to 10 for $c_s \leq 1.0$ and ≤ 0.5 ; in all cases $a=5$

Third, the transition to the awakening period is characterized by a sudden ‘jump’ followed by an also approximately linear increase but at a higher rate as compared to the linear increase during the sleeping period. We see this phenomenon for both sleeping citation intensities $c_s \leq 1.0$ and ≤ 0.5 . It is as if the SBs gain impact in an overdrive mode after a kind of phase transition. In other words, the knowledge diffusion process, at least insofar as measured by citations, is proceeding with a step change. This phenomenon is strongly reminiscent of knowledge diffusion as a percolation process (Bogaert et al. 2000; Barros et al. 2018; Tang et al. 2020) within the complex network of researchers in fields related to the SB; $t=s$ is the critical point on the time scale (in citation years after publication).

As we discussed in our foregoing paper (van Raan and Winnink 2019), it is remarkable that SBs belong to the most cited papers in medical research. For instance, the SBs with $s=5$, $c_s \leq 1.0$ and publication year 2005 ($n=246$) receive in total 20,766 citations, which is on average 84. Given the distribution of citations over the almost 400,000 medical papers in 2005 we find that these SBs are within the top-12%.

Influence of the chosen awakening period on citation patterns

In our earlier publications and also in the previous section we used a standard awakening period of 5 years after the sleeping period, i.e., $a=5$. However, we find that SBs may have a ‘slower’ awakening in the sense that they just do not reach the threshold value $c_a \geq 5.0$ within the first five years after awakening (so they are not found when we use $a=5$ as selection criterion) but easily within 10 year after awakening (and therefore they are found when using $a=10$ as selection criterion). Thus, in that case an SB cannot meet $a=5$ but does meet $a=10$. A typical example is the famous oral surgery very deep and long sleep SB ($s=10$) by Boyne and James (1980), see Fig. S2 (Supplementary

Table 1 Distribution of the three cases over the four sets of publications

	Set A: $s=5, a=5, c_a \geq 5.0, c_s \leq 1.0$ publication year 2005	Set B: $s=5, a=10, c_a \geq 5.0, c_s \leq 1.0$ publication year 2005
Case 1	109 (44%)	
Case 2	137 (56%)	137 (59%)
Case 3		94 (41%)
	Set C: $s=5, a=5, c_a \geq 5.0, c_s \leq 0.5$ publication years 2001–2005	Set D: $s=5, a=10, c_a \geq 5.0, c_s \leq 0.5$ publication years 2001–2005
Case 1	62 (37%)	
Case 2	106 (63%)	106 (63%)
Case 3		140 (43%)

Information). In order to further analyze the influence of the chosen awakening period on citation patterns of SBs we distinguish the following three cases:

1. SBs selected with $a=5$, but they do not meet $a=10$: SBs with a fast starting but relatively short awakening period;
2. SBs selected with $a=5$ and they also meet $a=10$: SBs with a fast starting and relatively long awakening period;
3. SBs selected with $a=10$, but they do not meet $a=5$: SBs with a slow starting but relatively long awakening period, like the Boyne paper.

SBs selected with $a=10$ and also meeting $a=5$ are the same as in case 2. We created four sets of SBs ($s=5, c_a \geq 5.0$): those with $a=5$ and those with $a=10$, each for $c_s \leq 1.0$ and for ≤ 0.5 . For the SBs with $a=10$ the last publication year is 2005. We take this year for our analysis for the SBs with $c_s \leq 1.0$. Because the number of SBs with $c_s \leq 0.5$ is much lower we take for these SBs the publication years 2001–2005. Table 1 shows the distribution of the three cases over the four SB sets and in Fig. 4 we present how the SBs are cited from their publication year to 2019. The data are the averages for the total number of SBs (these total numbers are given in Table 1). Our search algorithm also provides the data for self-citations. We include in Fig. 4 the average number of self-citations, again from publication year to 2019.

On the basis of Table 1 and Fig. 4 we make the following observations.

1. More than half of the SBs belongs to case 2, which means that the majority of SBs selected with $a=5$ also meet $a=10$ and the majority of SBs selected with $a=10$ also meet $a=5$.
2. SBs in case 1 and case 2 make more or less the same jump. But about 3 years after the jump, the citations for case 1 SBs drop, which means that these SBs rapidly loose impact. In contrast to this, case 2 SBs remain at a high impact level: these are the SBs with a lasting high impact.
3. SBs in case 3 shows the smallest jump, particularly for the SBs with the higher sleep intensity $c_s \leq 1.0$. This suggests that for higher c_s values the time-dependent development

Fig. 4 Average number of citations for SB with $s=5$ for cases 1, 2 and 3 as described in the main text. Panel A: $c_s \leq 1.0$, $a=5$; panel B: $c_s \leq 1.0$, $a=10$; panel C: $c_s \leq 0.5$, $a=5$; panel D: $c_s \leq 0.5$, $a=10$. In all cases $c_0 \geq 5.0$. Panel letter indications correspond to the set letter indications in Table 1

of the number of citations becomes more and more continuous and that these papers cannot be considered as SBs anymore.

4. A remarkable finding is that in all cases the average self-citation rate is more or less the same and also shows a similar development. During the beginning of the sleeping period, the average self-citation rate is comparable to the average citation rate. But in the remaining sleeping period the average self-citation rate becomes smaller than the average citation rate, particularly for the SBs with $c_s \leq 1.0$. For SBs with $c_s \leq 0.5$ the average self-citation rate stays on the same level as the average citation rate. It appears that the self-citation behavior is more or less independent of the depth of the sleep (the value of c_s) which is understandable as authors do not know how their paper will be cited in the years immediately after publication. On average there is no excessive self-citation that could be made responsible for the awakening. We further discuss the role of self-citations in subsection “Influence of Self-Citations on the Awakening Process”.

Linear approach of the phase-transition-like jump

To further analyze the phenomenon of the ‘jump to awakening’ in a quantitative sense, we use the same data as in Fig. 4 but now with the awakening period restricted to the first 3 years after the jump. We focus on the SBs with $a=5$ (case 1 and 2) because here the jump is most pronounced. In Fig. 5 we show the results for $s=5$ and 10; the results for $s=6, 7, 8$ and 9 are presented in Fig. S3 (Supplementary Information). We are particularly interested in: (1) the slope of the linear increase during the sleeping period, $\text{slope}(s)$; (2) the slope of the jump, $\text{slope}(j)$; (3) the slope of the first three years after the jump, $\text{slope}(a)$; and (4) the ratios $\text{slope}(a)/\text{slope}(s)$ and $\text{slope}(a)/\text{slope}(j)$. These slopes are simply the first derivatives of the equations given in Fig. 5. For instance in the case

Fig. 5 Average number of citations for SBs with $s = 5$ (panel A: $c_s \leq 1.0$, panel B: $c_s \leq 0.5$) and $s = 10$ (panel C: $c_s \leq 1.0$, panel D: $c_s \leq 0.5$) for the three years after the jump (in all cases $a = 5$)

of $s = 5, c_s \leq 0.5$ (panel B in Fig. 5) the average number of citations as a function of time during the sleeping period is given by the equation:

$$c_s(t) = 0.15t - 0.14, \text{ hence } \text{slope}(s) = dc_s(t)/dt = 0.15.$$

Similarly we find for the first three years after awakening $\text{slope}(a) = da(t)/dt = 1.31$. Thus, the ratio of the ‘awakening slope’ and the ‘sleeping slope’ is $Q = \text{slope}(a)/\text{slope}(s) = 8.73$. For the jump we find $\text{slope}(j) = 3.28$ and the ratio of the ‘awakening slope’ and the ‘jump slope’ $R = \text{slope}(a)/\text{slope}(j) = 0.40$. We present the values of the slopes and the ratios in Table 2.

We see that $\text{slope}(s)$ and also $\text{slope}(j)$ tend to decrease for larger s , whereas $\text{slope}(a)$ tend to increase for larger s . The ratios Q and R increase with s in both cases, they are lower for $c_s \leq 1.0$ as compared to those for $c_s \leq 0.5$. In Fig. 6 we show the graphical representation of the data in Table 2. We find that the best fit for the ratios of $\text{slope}(a)$ and $\text{slope}(s)$ as a function of s is given by a power law equation

$$Q = As^\beta \quad (3)$$

where $A = 0.43$ and $\beta = 1.79$ in the case of $c_s \leq 0.5$; and 0.10 and 1.89, respectively in the case of $c_s \leq 1.0$. For ratio R we find a similar power law dependence, be it with a considerably smaller exponent.

Considering all of the above observations together, we find a general picture for the change of the citation impact trend pattern if we go from a relatively short sleep to a longer sleeping period, see Fig. 7. What essentially happens is when the sleeping period gets longer, the phase-transition-like jump practically disappears and merges gradually with the awakening citation trend.

Table 2 The slopes and the ratios of slopes as a function of sleeping period s , for both $c_s \leq 0.5$ and ≤ 1.0 . The number of SBs and their publication years are given by n and py , respectively

s	slope(s)	slope(j)	slope(a)	Q	R	n	py
$c_s \leq 0.5$							
5	0.15	3.28	1.31	8.73	0.40	168	2001–2005
6	0.16	3.58	1.50	9.38	0.42	105	2001–2005
7	0.11	2.48	1.78	16.18	0.72	40	2001–2005
8	0.09	3.02	1.18	13.11	0.39	42	2001–2005
9	0.08	2.53	1.83	22.88	0.72	30	2000–2005
10	0.06	2.26	1.87	31.17	0.83	27	2000–2005
$c_s \leq 1.0$							
5	0.38	3.07	0.91	2.39	0.30	246	2005
6	0.31	2.94	0.96	3.10	0.33	131	2005
7	0.27	2.91	0.78	2.89	0.27	74	2005
8	0.20	2.35	1.22	6.10	0.52	84	2004–2005
9	0.17	2.68	1.06	6.24	0.40	66	2004–2005
10	0.13	2.66	1.13	8.69	0.42	47	2004–2005

Fig. 6 Slopes (panel A) and ratios of slopes (panel B) as a function of sleeping period for $c_s \leq 0.5$ and ≤ 1.0

Fig. 7 Change of the citation impact trend pattern: when the sleeping period gets longer (from the black to the red curve), the phase-transition-like jump practically disappears and merges gradually with the awakening citation trend

Influence of self-citations on the awakening process

Do self-citation play an important or perhaps even a decisive role in the awakening of SBs? In subsection “Influence of the chosen awakening period on citation patterns” we concluded that there are no indications that excessive self-citation could be made responsible for the awakening. Because ‘self-awakening’ could be a trigger for the awakening process, we now look to self-citations in more detail by comparing the SBs with a short and a long sleeping period. As an example we take all the SBs with a short sleeping period ($s=5$) and a longer awakening period ($a=10$), which is the data set in Fig. 4, panel B, case 2 and 3 together, publication year 2005 ($n=231$). For the long sleeping period we take the SBs with the same parameters, only now $s=10$ and publication year is 2000 ($n=44$) because we need a 5 year longer total measuring period. We show in Fig. 8 (panels A and B) the citation trend of these SBs from the year of publication until 2019. Most remarkable is that in both cases the average citation rate and the average self-citation rate are more or less the same during the beginning of the sleeping period. But in the remaining sleeping period the average self-citation rate becomes smaller than the average citation rate, a phenomenon we noted earlier in subsection “Influence of the chosen awakening period on citation patterns”.

Again we conclude that there is, in general, no excessive self-citation which could be made responsible for the awakening. But are there exceptions? By analyzing the annual number of self-citations for each SB we find the distribution of the number of self-citations over the SBs during the sleeping period. In the case of $s=5$ (publication year 2005, $n=231$) we find that 13% of the SBs has a higher average self-citation during the sleeping period (5 years) than the threshold value for citations ($c_s \leq 1.0$). For $s=10$ (publication year 2000, $n=44$) we have a similar result, here we find that 11% of the SBs has a higher average self-citation during the sleeping period (10 years) than the threshold value for citations ($c_s \leq 1.0$). Twice or more self-citations than the threshold value for citations during the sleeping period is found for only 3% and 2% of the SBs with $s=5$ and $s=10$, respectively. For the entire distribution function see Fig. 8, panel C.

Has the self-citation behavior of SB authors changed over time, for instance are authors more inclined to ‘promote’ their work through increasing numbers of self-citations in recent years? In order to investigate this we analyzed from 1980 the self-citations of the SBs with the short awakening period in order to make the analysis as recent as possible. Taking the SBs with $s=5$, $a=5$, $c_a \geq 5.0$, $c_s \leq 1.0$ we are able to extend the analysis up to

Fig. 8 Average number of citations and self-citations for SBs by citation year. Panel A: $s=5$, $c_s \leq 1.0$, $a=10$, $c_a \geq 5.0$, publication year 2005; panel B: $s=10$, $c_s \leq 1.0$, $a=10$, $c_a \geq 5.0$, publication year 2000; panel C: distribution of self-citations in both cases

and including 2010. We calculated the average number of self-citations during the sleeping period every five years from 1980 to 2010, see Fig. 9, panel A. We do not find a significant evidence of higher self-citing in recent years and this finding is confirmed by the frequency distribution of self-citations over SBs for 1980–2010, see Fig. 9, panel B.

In which medical research fields are SBs published?

Frequency-rank distribution for the SBs and their citing papers

So far there has been no research into the medical fields in which the SBs have been published. In total there are 58 WoS-covered medical fields (WoS subject categories). Our search algorithm also identifies the medical fields to which a SB belongs. We measure the number of times a specific medical field occurs in the selected sets of SBs and thus we determine the frequency distribution of fields. Because SBs can be assigned to more than one field (maximum three), the number of times SBs are assigned to a specific medical field will be larger than the actual number of SBs.

First we identified the medical fields for the same set of SBs as used in the forgoing section with the short sleeping period $s=5$ and a longer awakening $a=10$ ($c_s \leq 1.0$, $c_a \geq 5.0$). In publication year 2003 there is a relatively large number of these SBs ($n=380$), so we take this year as an example. Second, we compare the occurrence of these SB fields with the occurrence of these fields in the entire WoS-covered medical literature published in 2003. Third, we collect all papers citing these SBs in the period 2003–2019 and identify their research fields. Figure 10, panel A, shows the frequency-rank distribution for the fields with occurrence frequency ≥ 10 . We see that the ten fields with the highest occurrence are pharmacology and pharmacy; public, environmental and occupational health; psychiatry; nutrition and dietetics; surgery; sport sciences; neurosciences; medicinal chemistry; radiology, nuclear medicine and medical imaging; dentistry and oral surgery. The best fit of the measured frequency-rank distribution appears to be a logarithmic function:

$$f(i) = -A[\ln \{r(i)\}] + B \quad (4)$$

where $f(i)$ is the measured frequency of field i and $r(i)$ is the rank of field i according to its frequency. The coefficients A and B follow from the measurement. Frequency-rank distributions of the type given in Eq. 4 have cumulative distribution functions characterized by a stretched exponential. Such stretched exponential functions are claimed to describe many distribution functions in nature, economy (Knopoff and Sornette 1995; Laherrère and Sornette 1998; Sornette 2000) and some in bibliometrics (for instance ranking of universities by field-normalized impact).

The ratio of the occurrence of the SB fields with the occurrence of these fields in the entire WoS-covered medical literature (2003) gives us an indication of the relative occurrence of medical fields in the selected SBs. We calculated this ratio for all fields with an occurrence frequency ≥ 10 . If the ratio for a specific field is above 1, then this field is more present in the SBs as compared to the total medical literature. Given the relatively low numbers of the SBs, we estimate that only a ratio > 2 is significant. As we see in Fig. 10,

Fig. 10 Occurrence frequency (panel A) and relative occurrence (panel B) of fields for SBs with $s=5$ (►) ($a=10$, $c_a \geq 5.0$, $c_s \leq 1.0$), published in 2003 (the ratio value 2 is indicated with the blue horizontal line); panel C: occurrence frequency of fields for all papers published in 2003–2019 citing these SBs

panel B, this is the case for nursing; rehabilitation; medicinal chemistry; nutrition and dietetics; sport sciences; dentistry and oral surgery. Our first impression is that particularly non-core medical fields are overrepresented in (at least the selected set of) SBs. In follow-up work we will investigate this in detail by analyzing more publication years and more different sleeping times. In Fig. 10, panel C, we show the frequency-rank distribution for the fields of the 39,206 papers citing the above SBs in the period 2003–2019. Also this distribution is a similar logarithmic function as Eq. 4. But evidently, much more fields are involved in the case of the citing papers (we show the fields that contribute more than 2% within the total number of fields of the citing papers). As a consequence, the distribution for the citing papers is flatter than the distribution for the SBs which means much larger values for the coefficients A and B . Table 3 gives an overview of results.

Fig. 9 Self-citations during the sleeping period every five years from 1980 up to and including 2010, for SBs with $s=5$, $a=5$, $c_a \geq 5.0$, $c_s \leq 1.0$. Panel A: average number of self-citations; panel B: distribution of self-citations over the SBs

Table 3 Occurrence frequency of fields for of SBs with $s=5$ ($a=10$, $c_a \geq 5.0$, $c_s \leq 1.0$), published in 2003 and for all papers published in 2003–2019 citing the above SBs

SBs ($N=380$)	n	%	Papers citing the SBs ($N=39,206$)	n	%
Pharmacology, pharmacy	38	10.0	Pharmacology, pharmacy	2109	5.4
Public environm occupat health	30	7.9	Neurosciences	2028	5.2
Psychiatry	25	6.6	Psychiatry	1823	4.7
Nutrition, dietetics	23	6.1	Public environm occupat health	1785	4.6
Surgery	22	5.8	Surgery	1760	4.5
Sport sciences	19	5.0	Clinical neurology	1649	4.2
Neurosciences	18	4.7	Sport sciences	1629	4.2
Med chemistry	16	4.2	Oncology	1390	3.5
Radiology, nucl med, med imag	16	4.2	Orthopedics	1346	3.4
Dentistry, oral surgery	15	3.9	Rehabilitation	1332	3.4
Genetics, heredity	15	3.9	General internal med	1320	3.4
Clinical neurology	15	3.9	Dentistry, oral surgery	1239	3.2
Oncology	14	3.7	Cardiac cardiovascular systems	1136	2.9
General internal med	14	3.7	Radiology, nucl med, med imag	1120	2.9
Cardiac cardiovascular systems	13	3.4	Biochem, mol biology	1094	2.8
Orthopedics	13	3.4	Food science technology	1045	2.7
Rehabilitation	13	3.4	Multidisciplinary sciences	1036	2.6
Endocrinology, metabolism	10	2.6	Nutrition, dietetics	887	2.3
Medicine research experimental	10	2.6	Ophthalmology	807	2.1
Nursing	10	2.6	Clinical psychology	769	2.0

N is the total number of SBs and papers citing the SBs, respectively; n is the number in the given field and % is the percentage

Difference in medical fields for short versus long sleeping SBs

In our next analysis we compare the fields for the $s=5$ SBs with the $s=10$ SBs. In order to make the comparison as recent as possible, we take $a=5$. In this case, the last publication year for the $s=10$ SBs ($c_a \geq 5.0$, $c_s \leq 1.0$) is 2005. Given the low annual number of $s=10$ SBs we take the five year period 2001–2005 ($n=136$). For the $s=5$ SBs we take the middle year 2003 ($n=316$). Figure 11 shows the frequency-rank distribution for both cases. The ten fields with the highest occurrence in the $s=5$ case are (Fig. 11, panel A): pharmacology and pharmacy; surgery; public, environmental and occupational health; neurosciences; nutrition and dietetics; clinical neurology; psychiatry; sport sciences; medicinal chemistry; dentistry and oral surgery. In the case of $s=10$ (Fig. 11, panel B) we find largely the same fields in the top-10: public, environmental and occupational health; neurosciences; pharmacology and pharmacy; nutrition and dietetics; oncology; sport sciences; immunology; rehabilitation; surgery; clinical neurology. Also in this analysis we find that the frequency-rank distribution of fields within sets of specific SBs is best described by a logarithmic function as given by Eq. 4. The frequency-rank distribution function for $s=10$ is somewhat steeper than for $s=5$ which can be expected given the lower number of SBs.

Fig. 11 Occurrence frequency of medical fields for of SBs with $s=5$ published in 2003 (panel A) and for SBs with $s=10$ published in 2001–2005 (panel B). In both cases $a=5$, $c_a \geq 5.0$, $c_s \leq 1.0$

Conclusions

In this study we investigate several characteristics of SBs that have not or hardly been investigated previously. We find that the choice of the awakening period in the selection of SBs has consequences for the measured citation patterns. In foregoing publications we used a standard awakening period of five years after the sleeping period. However, SBs

may have a ‘slower’ awakening in the sense that they just do not reach the citation threshold value within the first five years after awakening but easily within ten years after awakening. On the basis of this phenomenon we distinguish three cases with different citation patterns related to the chosen awakening period. The analysis of these cases leads to a number of interesting findings. A first remarkable observation is that generally we see during the sleeping period a slow, approximately linear increase of citations and not a constant zero or constant very low number of citations and also not a random distribution of the very few citations during sleeping period. This finding has an important implication: most SBs show a ‘weak awakening’ already during the sleeping period.

A further remarkable observation is that the transition to the awakening period is characterized by a sudden ‘jump’ and followed by an approximately linear increase of the number of citations but at a much higher rate as compared to the linear increase during the sleeping period. It is as if the SBs gain impact in an overdrive mode after a kind of phase transition. In other words, the knowledge diffusion process, at least insofar as measured by citations, is proceeding with a step change. In order to analyze this ‘jump to awakening’ in a quantitative sense, we measured the slope of the linear increase during the sleeping period; the slope of the jump; and the slope of the first three years after the jump. The slopes and their ratios provide new insight into the development and structure of the SB citation patterns. If we go from a relatively short sleep to a much longer sleeping period we find that when the sleeping period gets longer, the phase-transition-like jump practically disappears and more or less merges with the awakening citation trend. In all cases the average citation rate and the average self-citation rate are more or less the same during the beginning of the sleeping period, followed by a smaller average self-citation rate in the remaining sleeping period. We conclude that there is no excessive self-citation which could be made responsible for the awakening.

Our search algorithm also identifies the medical fields to which a SB belong. In this way, we can find out which fields frequently occur in a set of SBs, whether these fields are overrepresented in SBs and how these SB fields compare with the fields of the publications citing the SBs. The best fit of the measured frequency-rank distribution appears in all cases to be logarithmic. We find that particularly nutrition and dietetics; sport sciences; medicinal chemistry; radiology, nuclear medicine and medical imaging; dentistry and oral surgery; rehabilitation; and nursing are overrepresented in SBs. Therefore our first impression is that particularly non-core medical fields are overrepresented in (at least the selected set of) SBs. In follow-up work we will investigate this in more detail by analyzing more publication years and more different sleeping times.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11192-021-03910-5>.

Availability of data and material See OSF repository <https://osf.io/4ru96/>.

Compliance with ethical standards

Conflict of interest The author declares that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not

permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Barros, B., Fernandez-Zubieta, A., Fidalgo-Merino, R., & Triguero, F. (2018). Scientific knowledge percolation process and social impact: A case study on the biotechnology and microbiology perceptions on Twitter. *Science and Public Policy*, 45(6), 804–814.
- Bogaert, J., Rousseau, R., & van Hecke, P. (2000). Percolation as a model for informetric distributions: Fragment size distribution characterized by Bradford curves. *Scientometrics*, 47(2), 195–206.
- Bornmann, L., Ye, A. Y., & Ye, F. Y. (2018). Identifying “hot papers” and papers with “delayed recognition” in large-scale datasets by using dynamically normalized citation impact scores. *Scientometrics*, 116(2), 655–674.
- Boyne, P. J., & James, R. A. (1980). Grafting of the maxillary sinus floor with autogenous marrow and bone. *Journal of Oral Surgery*, 38(8), 613–616.
- Braun, T., Glänzel, W., & Schubert, A. (2010). On Sleeping Beauties, Princes and other tales of citation distributions. *Research Evaluation*, 19(3), 195–202.
- Du, J., & Wu, Y. (2018). A parameter-free index for identifying under-cited sleeping beauties in science. *Scientometrics*, 116(2), 959–971.
- El Aichouchi, A., & Gorry, P. (2018). Delayed recognition of Judah Folkman’s hypothesis on tumor angiogenesis: when a prince awakens a sleeping beauty by self-citation. *Scientometrics*, 116(1), 385–399.
- Fang, H. (2018). Analysing the variation tendencies of the numbers of yearly citations for sleeping beauties in science by using derivative analysis. *Scientometrics*, 115(2), 1051–1070.
- Garfield, E. (1970). Would Mendel’s work have been ignored if the Science Citation, Index was available 100 years ago? *Essays of an Information Scientist* 1, 69–70; also in *Current Contents* 2, 69–70.
- Garfield, E. (1980). Premature discovery or delayed recognition –why? *Essays of an Information Scientist* 4, 488–493; also in *Current Contents* 21, 5–10, 1980.
- Garfield, E. (1989). Delayed recognition in scientific discovery: Citation frequency analysis aids the search for case histories. *Current Contents*, 23, 3–9.
- Garfield, E. (1990) More delayed recognition. Part 2. From Inhibin to Scanning electron microscopy. *Essays of an Information Scientist* 13, 68–74; also in *Current Contents* 9, 3–9, 1990.
- He, Z., Lei, Z., & Wang, D. (2018). Modeling citation dynamics of “atypical” articles. *Journal of the Association for Information Science and Technology* 69(9) 1148–1160.
- Hou, J., & Yang, X. (2019). Patent sleeping beauties: evolutionary trajectories and identification methods. *Scientometrics*, 120(1), 187–215.
- Hou, J., & Yang, X. (2020). Social media-based sleeping beauties: Defining, identifying and features. *Journal of Informetrics*, 14(2), 101012.
- Hou, J., Li, H., & Zhang, Y. (2020). Identifying the princes base on Altmetrics: An awakening mechanism of sleeping beauties from the perspective of social media. *PLoS ONE*, 15(11), e0241772.
- Ke, Q., Ferrara, E., Radicchi, F., & Flammini, A. (2015). Defining and identifying sleeping beauties in science. *Proceedings of the National Academy of Sciences of the United States of America*, 112(24), 7426–7431.
- Knopoff, L., & Sornette, D. (1995). Earthquake Death Tolls. *Journal de Physique*, I(5), 1681–1688.
- Li, J. (2014). Citation curves of “all-elements-sleeping-beauties”: “flash in the pan” first and then “delayed recognition.” *Scientometrics*, 100(2), 595–601.
- Li, J., & Shi, D. (2016). Sleeping beauties in genius work: When were they awakened? *Journal of the Association for Information Science and Technology*, 67(2), 432–440.
- Li, J., & Ye, F. Y. (2012). The phenomenon of all-elements-sleeping-beauties in scientific literature. *Scientometrics*, 92(3), 795–799.
- Li, J., & Ye, F. Y. (2016). Distinguishing sleeping beauties in science. *Scientometrics*, 108(2), 821–828.
- Li, J., Shi, D., Zhao, S. X., & Ye, F. Y. (2014). A study of the “heartbeat spectra” for “sleeping beauties.” *Journal of Informetrics*, 8(3), 493–502.
- Sornette, D. (2000). *Critical Phenomena in Natural Sciences. Chaos, Fractals, Selforganization and Disorder: Concepts and Tools*. Berlin, Heidelberg, New York: Springer Verlag, pp. 148–149.
- Sornette, D., Knopoff, L., Kagan, Y. Y., & Vanneste, C. (1996). Rank-ordering statistics of extreme events: Application to the distribution of large earthquakes. *Journal of Geophysical Research-Solid Earth*, 101, 13883–13893.

- Stent, G. S. (1972). Prematurity and uniqueness in scientific discovery. *Scientific American*, 227(6), 84–93.
- Tang, M. C., Teng, W., & Lin, M. (2020). Determining the critical thresholds for co-word network based on the theory of percolation transition: A case study in Buddhist studies. *Journal of Documentation*, 76(2), 462–483.
- van Raan, A. F. J. (2004). Sleeping Beauties in Science. *Scientometrics*, 59(3), 461–466.
- van Raan, A. F. J. (2015). Dormitory of physical and engineering sciences: Sleeping beauties may be sleeping innovations. *PLoS ONE*, 10(10), e0139786.
- van Raan, A. F. J. (2017a). Sleeping beauties cited in patents: Is there also a dormitory of inventions? *Scientometrics*, 110(3), 1123–1156.
- van Raan, A. F. J. (2017b). Patent citations analysis and its value in research evaluation: A review and a new approach to map technology-relevant research. *Journal of Data and Information Science*, 2(1), 13–50.
- van Raan, A. F. J., & Winnink, J. J. (2018). Do younger sleeping beauties prefer a technological prince? *Scientometrics*, 114, 701–717.
- van Raan, A. F. J., & Winnink, J. J. (2019). The occurrence of ‘Sleeping Beauty’ publications in medical research: Their scientific impact and technological relevance. *PLoS ONE*, 14(10), e0223373.
- Wang, D., Song, C., & Barabási, A.-L. (2013). Quantifying long-term scientific impact. *Science*, 342(6154), 127–132.
- Ye, F. Y., & Bornmann, L. (2018). “Smart girls” versus “sleeping beauties” in the sciences: The identification of instant and delayed recognition by using the citation angle. *Journal of the Association for Information Science and Technology*, 69(3), 359–367.
- Zhang, H. H., & Ye, F. Y. (2020). Identifying “associated-sleeping-beauties” in “swan-groups” based on small qualified datasets of physics and economics. *Scientometrics*, 122(3), 1525–1537.