Jointly published by Akadémiai Kiadó, Budapest
and Kluwer Academic Publishers, Dordrecht

Scientometrics,
Vol. 59, No. 3 (2004) 461–466

Short communication

Sleeping Beauties in science

ANTHONY F. J. VAN RAAN

Centre for Science and Technology Studies, Leiden University, Leiden (The Netherlands)

A 'Sleeping Beauty in Science' is a publication that goes unnoticed ('sleeps') for a long time and then, almost suddenly, attracts a lot of attention ('is awakened by a prince'). We here report the -to our knowledge- first extensive measurement of the occurrence of Sleeping Beauties in the science literature. We derived from the measurements an 'awakening' probability function and identified the 'most extreme Sleeping Beauty so far'.

Introduction

'Being ahead of one's time' has always both fascinated and frightened scientists. Their dearest publication may become a 'Sleeping Beauty', an article that goes unnoticed ('sleeps') for a long time and then, almost suddenly, attracts a lot of attention. It is our experience in the application of bibliometric methods in research evaluation (MOED et al., 1995; VAN RAAN, 1996) that on quite a few occasions, scientists claimed that one or more of their publications will not be picked up for a while, as they are 'ahead of time'. We call this the 'Mendel syndrome', mentioned after Gregor Mendel (MENDEL, 1865) whose discoveries in plant genetics were so unprecedented that it took thirty-four years for the scientific community to catch up to it. So the search for sleeping beauties is not just an exotic whim, but a necessity in order to have an answer to Mendel-like claims. In this paper we present the – to our knowledge – first measurements of the occurrence of Sleeping Beauties in the science literature. The 'most extreme Sleeping Beauty so far' is identified and an attempt is made to derive from the measurements an 'awakening' probability function.

Received December 16, 2003

Address for correspondence:
ANTHONY F. J. VAN RAAN
Centre for Science and Technology Studies, Leiden University
P.O. Box 9555, 2300 RB Leiden, The Netherlands
E-mail: vanraan@cwts.leidenuniv.nl

0138–9130/2004/US $ 20.00

Method

We studied scientific sleeping beauties with a bibliometric approach. There are three main variables: (1) *depth of sleep*, we take two modalities: the article receives at most 1 citation on average per year (deep sleep), or between 1 to 2 citations on average per year during a specific period (less deep sleep); (2) *length of the sleep*, duration of the above period; and (3) *awake intensity*: number of citations per year, during four years following the sleeping period.

Using our very large CWTS scientific publication data-system* with about 20,000,000 articles from 1980 up till now and a total volume of about 300,000,000 citations, we carried out the following measuring procedure. For 6 sleeping periods with length s = 5, 6, 7, 8, 9, and 10 years, respectively, all with publication years starting in 1980, we identified in each of the six sleeping periods all articles either 'in deep sleep' or 'in less deep sleep' as defined above. For instance, in the case of s = 10 the number of citations c_s (self-citations excluded) is between 0 and 10 (indicated throughout this paper as [0,10]) for deep sleep, and c_s = [11,20] for less deep sleep.

Next we investigated for all of the above articles with different s and c_S the 'awake intensity', i.e., the extent to which these articles have been cited (again self-citations excluded) in a four-years period immediately following each of the six sleeping periods. More particularly, we investigated whether they belong to one of 5 'awake intensity classes' c_w = [21,30], [31,40], [41,50], [51,60] and [>60], respectively, i.e., on the average 6, 9, 12, 15, and more than 15 citations per year during the four-year awakening period. In this way we performed in total about 5.10^8 citation analyses within the data-system described above.

Results and discussion of the general findings

We first show in Table 1 as a typical example the results of our measurements for the longest, 'most recent' sleeping period (s = 10, articles from 1988, sleeping period 1988–1997), for both 'deep sleep' (c_s = [0,10]), as well as 'less deep sleep' (c_s = [11,20]), in numbers of publications (N):

* Based on the *Science Citation Index* (SCI), the *Social Science Citation Index* (SSCI), the *Arts & Humanities Citation Index* (AHCI) and all related indexes, produced and published by the Institute for Scientific Information (ISI) in Philadelphia.

Table 1. Sleeping Beauties with sleeping period 1988–1997

c_w	N, less deep	N, deep
[21,30]	276	41
[31,40]	29	6
[41-50]	5	0
[51-60]	0	1
[> 60]	1	0

For instance, we found for 1988 in total 41 articles (from about 1,000,000 articles published in 1988) that after a 'deep sleep' of 10 years received during the next four years (the 'awakening period') between 21 and 30 citations (about 6 to 7 citations on average per year).

A substantial part of all our measurements is represented in Figure 1. We show the number of publications N as a function of four awakening intensity classes c_w = [21,30], [31,40], [41,50], [51,60] (centered on the logarithmic x-axis scale at 25, 35, 45 and 55, respectively) in four cases:

a: shortest sleeping period: $s = 5$; deep sleep: c_s = [0,5];
b: shortest sleeping period: $s = 5$; less deep sleep: c_s = [6,10];
c: longest sleeping period: $s = 10$; deep sleep: c_s = [0,10];
d: longest sleeping period: $s = 10$; less deep sleep: c_s = [11,20].

Each of these cases has a separate 'panel' of one decade in Figure 1 with c_w running from 10 to 100.

The total number of publications in the data-system increases from 656,991 in 1980 to 1,046,839 in 2000. It is clear that the more publications are *published* in a given year, the higher the probability of Sleeping Beauties. This higher probability also works at the 'citing side': the more publications are available as *citing* papers in later years, the higher the chance that a publication of an earlier year will be cited. Thus, we normalized the measured numbers in order to take the increasing number of cited and citing publications into account.

The results are striking: in all four cases a similar, steep power-law decrease of N as a function of c_w is found. Comparison of the N values of the 'deep sleep' cases with those of the 'less deep sleep' cases for the same c_w values (case [a] compared with [b], and [c] with [d]), and comparison of the N values of the 'short sleep' with the 'long sleep' cases for the same c_w values (case [a] compared with [c], and [b] with [d]), reveals first estimations of the dependencies of N on s and c_s.

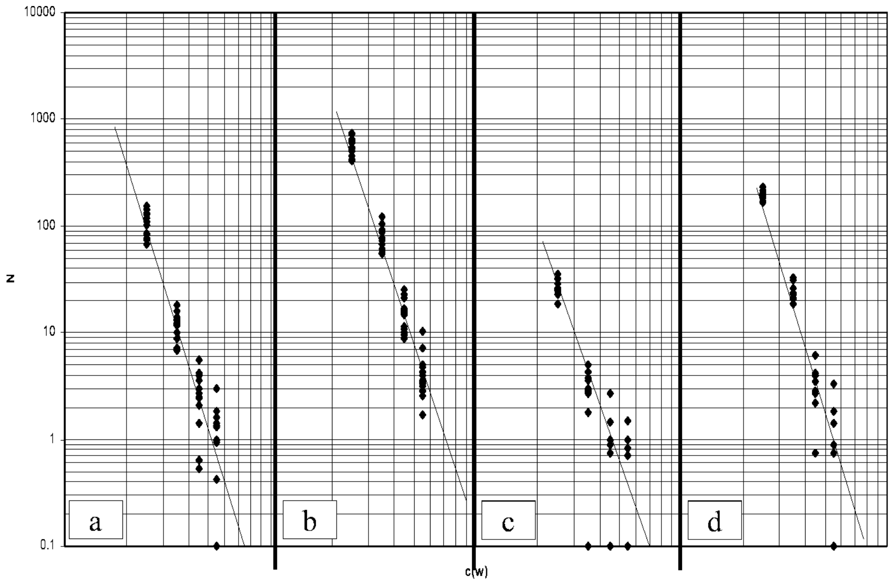

Figure 1. Sleeping Beauty characteristics
♦: Number of publications N is presented as a function of four awakening intensity classes c_w = [21,30], [31,40], [41,50], [51,60] (centered on the logarithmic x-axis scale at 25, 35, 45 and 55, respectively) in four cases: (a) shortest sleeping period: $s = 5$; deep sleep: $0 \leq c_s \leq 5$; (b) shortest sleeping period: $s = 5$; less deep sleep: $6 \leq c_s \leq 10$; (c) longest sleeping period: $s = 10$; deep sleep: $0 \leq c_s \leq 10$; (d) longest sleeping period: $s = 10$; less deep sleep: $11 \leq c_s \leq 20$. Each of these cases has a separate 'panel' of one decade in the figure with c_w ranging from 10 to 100

Thus, from our total set of measurements we were able to derive a general 'Grand Sleeping Beauty Equation':

$$N = f\{s, c_s, c_w\} \sim s^{-2.7} \,.\, c_s^{+2.5} \,.\, c_w^{-6.6}$$

which (after determination of a constant factor) gives the *number* of Sleeping Beauties for any sleeping time, sleep intensity and awake intensity, and particularly the *dependency* on these variables. The dependence on the awakening intensity has the steepest slope ever found in bibliometric analysis.

Formulated in daily language, we find on the basis of the above observations the following general characteristics of Sleeping Beauties in scientific literature:

- The probability of awakening after a deep sleep is smaller for longer sleeping periods;
- For a less deep sleep, the length of the sleeping period matters less for the probability of awakening;
- The probability for higher awakening intensities decreases extremely rapidly in all cases (the very steep slope indicated in Fig. 1 for all four cases [a], [b], [c], and [d]). This probability is independent of both length as well as depth of sleep!

Extreme case and concluding remarks

The most extreme case (in this investigation), i.e., with the longest sleeping period (s = 10), the deepest sleep (c_s = 0 in all 'sleeping years'), and the highest awake intensity ($c_w > 60$), we find just *one* Sleeping Beauty, published in 1986 (as can be seen in Table 1, for the publication year 1988 is there no Sleeping Beauty of this kind). This extraordinary paper is: L. J. ROMANS, *Physics Letters B*, 169 (1986) 374–380, "Massive N = 2a supergravity in ten dimensions", with $c_w = 64$).

The prince of this Sleeping Beauty is the (author of the) first citing paper after ten years: J. POLCHINSKI, *Physical Review Letters* 75 (1995) 4724–4727. On the basis of the comments of both authors it became clear that we indeed deal with a classical case of 'being ahead of one's time'. The Romans' paper presented a model of supergravity in string theory leading to a new phenomenon (breaking supersymmetry). At the time of publication (1986) the entire theoretical physics field on supergravity was in a sense a sleeping beauty. Many years later (1995), the string theory community was 'more ready' for it as the phenomenon was rediscovered by the 'prince'. A colleague of the prince then remembered the Romans paper, probably helped by the fact that they were at the same affiliation as the Sleeping Beauty, namely the Institute for Theoretical Physics, University of California at Santa Barbara, USA. The link (the citation) was made, and the Sleeping Beauty awoke.

Of course, there are many more Sleeping Beauties if we decrease (1) the sleeping time, and/or (2) the depth of the sleep, and/or (3) the awake intensity, as can be seen immediately from the above derived equation which clearly represents a continuous distribution. Thus, we can find 100 to even 1,000 'less prominent' Sleeping Beauties per year, which is, however, still not much in an annual population of about 1,000,000.

The search algorithm is not designed to detect articles that attract a substantial number of citations immediately or soon after publication and followed by a considerable increase of citations in later years for a very long time. Such articles cannot be considered as 'Sleeping Beauties', they are typical 'Classics'.

We are currently updating our study by investigating the occurrence of the most recent Sleeping Beauties. Further work is necessary to analyze the statistics of Sleeping Beauties for different fields and to study the possible influence of specific (types of) journals.

*

I thank Peter Negenborn for his extensive data-analytical and programming work. I am grateful to Professor Eric Bergshoeff (University of Groningen), Professor Joseph Polchinsky (University of California at Santa Barbara) and Dr Larry Romans (Jet Propulsion Laboratory) for their stimulating comments.

References

MENDEL, G. (1865), Versuche über Pflanzen-Hybriden (Experiments with Plant Hybrids), *Proceedings of the National History Society of Brunn* (Bohemia, now Czech Republic).

MOED, H. F., DE BRUIN, R. E., VAN LEEUWEN, TH. N. (1995), New bibliometric tools for the assessment of national research performance: Database description, overview of indicators and first applications. *Scientometrics,* 33 : 381–422.

VAN RAAN, A. F. J. (1996), Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises. *Scientometrics,* 36 : 397–420.