Rediscovering Forgotten Research:
Sleeping Beauties at the University of Waterloo

Jeffrey Demaine*

Dana Porter Library, University of Waterloo,
Waterloo, ON, Canada
E-mail: jdemaine@outlook.com

ABSTRACT

An academic article is normally cited within a few years of publication, after which interest falls off as the research field moves on. However, an article is sometimes ignored for many years only to attract interest after a long period of dormancy. Such articles are called "Sleeping Beauties." A general characterization of this pattern has recently been defined and is used in this study to identify five Sleeping Beauties that were published by researchers at the University of Waterloo in the 1970s and 1980s. While a handful of studies have examined the occurrence of such Sleeping Beauties in specific fields of research or in a particular journal, none has yet identified these unusual articles in the context of the lasting impact of a university's research. This study is therefore a novel application of the latest technique for identifying Sleeping Beauties. The possibilities for using this unusual citation pattern in raising the profile of a university's research are discussed.

Keywords: bibliometrics, sleeping beauties, beauty coefficient, research impact

Open Access

1. INTRODUCTION

A university's ability to advance a research front can be quantified by counting the citations to its published research. But it is not only the recent research coming from a university that advances a research front. Some papers are able to influence current research many years after they first appear. Papers published decades ago may find new relevance and become highly cited after many years of dormancy, suggesting that these ideas were ahead of their time. Publications that exhibit this pattern of delayed recognition are known as "Sleeping Beauties" (SBs) (van Raan, 2004). Although rare, they have been identified in such diverse research areas as physics (Redner, 2005), pediatrics (Završnik & Kokol, 2016), medicine and biological engineering (Huang, Hsu, & Ciou, 2015), and psychology (Lange, 2005; Ho & Hartley, 2017). This raises the question of whether SBs can also be found at specific institutions. The current study seeks to identify SBs in the research papers published by the faculty of the University of Waterloo located in Ontario, Canada.

In this paper we present a review of the literature about SBs and explore some of the reasons behind this unusual citation pattern. We implement the most advanced algorithm for evaluating the 'surprisingness' of an article's rediscovery and use it in a case study of SBs published by researchers at the University of Waterloo.

One of the original SBs to be studied is the work of Einstein, Podolsky, and Rosen published in 1935. Known as the EPR paper, it was not extensively cited until some 60 years after it appeared (Fig. 1). Although Redner (2005) notes that the EPR paper was cited 36 times before 1980, the explosive growth of interest in this paper since 1990 is a hallmark of an idea that was ahead of its time. Though the concept of quantum entanglement presented in the EPR paper may have been of some theoretical interest throughout the twentieth century, it is only with the technological advances in quantum computing in recent years that this article has found new relevance. Although it is an old paper, it has become current and is now a central part of the evolving research front. A history of the implications of the EPR paper make clear its relevance to quantum physics: "Due to its role in the development of quantum information theory, it is also near the top in [the] list of currently 'hot' papers" (Fine, 2017).

1.1. Drivers Behind the SB Citation Pattern

There are several explanations as to why an article would exhibit such an unusual pattern of citations. In some cases, SBs appear because the research in the article finds relevance in another discipline where it has an impact far greater than in its substantive field. Examining the metadata of the articles that cited the SBs and caused them to 'awaken' (so called "Prince" articles), Braun, Glänzel, and Schubert (2010) as well as Teixeira, Vieira, and Abreu (2017) both found that 40% of the Princes were from a research field different from that of the SB that they awakened. In such a situation
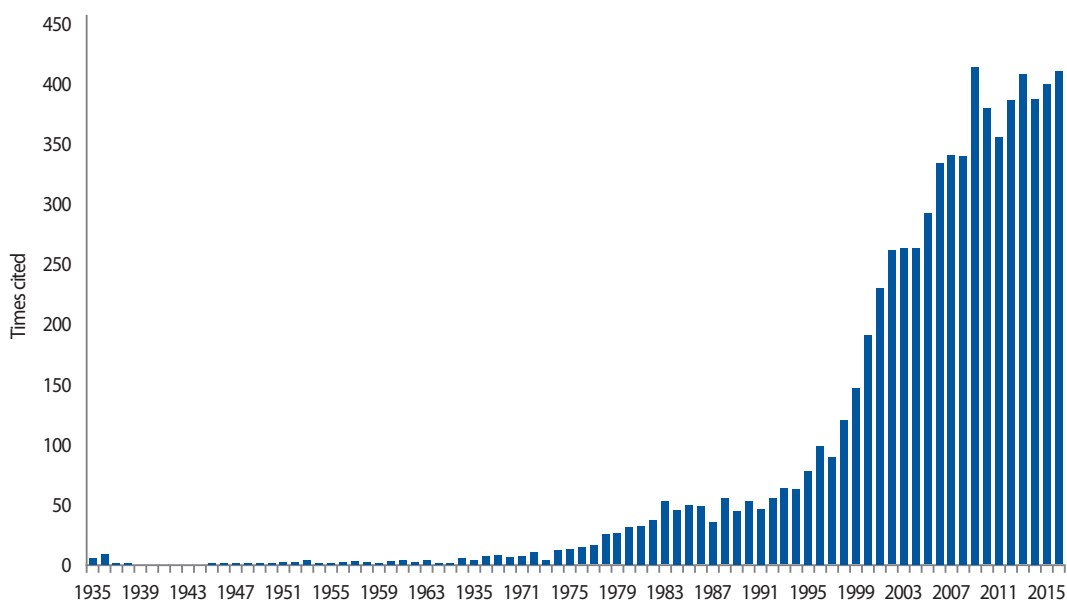

Fig. 1. Citations of the Einstein, Podolsky, and Rosen (1935) paper.

the SB pattern of citations can be seen as a snapshot of the transfer of ideas from one domain of knowledge to another. They also found that the Princes were consistently from journals with twice the journal impact factor of the journals in which the SBs appeared. Thus it appears that in many cases the dormancy of SBs is due in some respects to the relative obscurity of the journal in which it was published. It is only when subsequent research in higher-profile journals and/or fields picks up on the dormant article that the SB is awakened.

Secondly, the concepts outlined in a paper may be ahead of their time or run counter to the prevailing consensus of the research field. A study of SBs in the field of innovation studies found that the reasons for their delayed recognition varies and is as much due to the content of the SB as to the characteristics of the Prince (Teixeira et al., 2017). A long dormancy is sometimes due to resistance within the scientific community to the ideas described in the SB, and its awakening is sometimes attributable to the development of new conceptual models that can leverage the ideas of the SB. This aligns with the concept of a "paradigm shift" as described by Thomas S. Kuhn in his landmark book *The structure of scientific revolutions* (Kuhn, 1962). In this scenario, the SB serves as an indicator of the rapid evolution of a research field as it overturns outmoded ideas.

A third explanation for the sudden interest in a long-dormant paper is one of technological readiness. It may be that the ideas discussed in an SB paper are correct and/or relevant to the field, but the equipment required to test or implement those ideas is too expensive to be widely available or simply does not exist. The EPR paper illustrates just how it may take decades for the right combination of ideas and technological advancement to come together. Certainly few would have dismissed an article by Albert Einstein as being of no value. Indeed it was not actually dormant and it received a modest yet steady number of citations every year during the 1950s, 1960s, and 1970s. The discussion around this paper became known as the "EPR paradox" (the possibility of faster-than-light communication between two particles). Yet it was only in the late 1980s that the EPR paper began to be highly cited, coinciding with the ability to exploit quantum entanglement in the context of quantum computing that these ideas became applicable.

Thus there are at least four reasons why an article should become an SB: It is hidden in a relatively obscure journal, it finds traction in a different field, it is too unorthodox to be immediately integrated into its proper field of research, or it is ahead of its time in terms of the technology required to apply the concepts it contains.

2. METHODOLOGY

There are a number of approaches to identifying SBs. When the concept was originally characterised, articles were evaluated according to features of their citation history. Glänzel and Garfield (2004) defined these "delayed recognition" papers as having been uncited for at least five years after publication, and then subsequently being cited at least 50 times in the following 15 years. Redner (2005) offered a simple rule-of-thumb for determining which papers qualify as SBs. He defines a "revived classic as a nonreview Physical Review article, published before 1961, that has received more than 250 citations and has a ratio of the average citation age to the age of the paper greater than 0.7." This approach to describing SBs lends itself well to their identification in large databases by defining a few search parameters.

A more general technique for identifying SBs that does not rely on rule-of-thumb thresholds has been recently proposed by Ke, Ferrara, Radicchi, & Flammini (2015). Instead, the algorithm they propose (Equation 1) expresses how surprising the citations to an article are in relation to the number of years it has been dormant. The resulting number is called the "Beauty Coefficient" (BC).

$$B = \sum_{t=0}^{t_m} \frac{\frac{c_{t_m} - c_0}{t_m} \cdot t + c_0 - c_t}{\max\{1, c_t\}}$$

Equation 1. The Beauty Coefficient as described by Ke, Ferrara, Radicchi, and Flammini (2015).

This approach takes as its input five parameters:

The number of years since publication, t
The number of times an article was cited in its year of publication, c_0
The number of citations at year t, c_t
The number of years since publication until the year of maximum citation, t_m
The number of times an article was cited in its most highly-cited year, c_{t_m}

By calculating a sum of these five parameters for every year from publication to the year in which the article in question is most highly cited, a metric of the SB effect is obtained which expresses how surprising the resurgence of citations is. Considering that in the context of an article which has received a steadily-increasing number of citations year after year, yet another year of increased citations is not at all surprising, and consequently the article in question

would have a very low BC. Conversely, a paper that has been dormant for decades only to receive a sudden and large spike in citations is highly unusual and would therefore receive a high BC.

This technique is used in this study to identify SB papers that were published by faculty at the University of Waterloo. To calibrate our implementation of the Ke et al. (2015) algorithm, we quantify the rapid growth in citations to the EPR paper after 1987 and arrive at a BC of 2,333. This is very similar to the score of 2,258 calculated by Ke et al. (2015), the slight increase being due to differences in the journals indexed (and therefore the citations identified) between the Web of Science used by Ke et al. and the Scopus database used in the current study. In addition, the accumulation of new citations to the EPR paper in the three years since Ke et al. collected their data would naturally produce a higher BC.

To identify SBs at the University of Waterloo, citation frequency data (Demaine, 2018) was downloaded from Elsevier's Scopus database in November 2017 for the period 1958 (when the university was founded) to 1998 (inclusively). While the University of Waterloo now publishes thousands of papers every year, its output at its founding was naturally very modest. For example, the first and only paper published by Waterloo in 1958 was "Decay of immediate memory with age" by Fraser. Since then the growth in publications from the university has been impressive with the university publishing 4,341 papers in 2017 (Scopus, 2018).

As no SB articles were found to have been published after 1987, this study will only examine the first 30 years of

Table 1. Papers by year of publication and number of times they have been cited

| Publication year	No. of papers	Total cites	Average cites per paper
1958	1	10	10
1960	4	6	1.5
1961	6	9	1.5
1962	4	65	16.3
1963	21	722	34.4
1964	39	322	8.3
1965	69	911	13.2
1966	83	903	10.9
1967	116	3,855	33.2
1968	156	2,726	17.5
1969	249	6,659	26.7
1970	301	6,141	20.4
1971	385	8,654	22.5
1972	432	8,658	20.0
1973	464	9,288	20.0
1974	524	9,180	17.5
1975	461	8,251	17.9
1976	470	8,250	17.6
1977	489	10,991	22.5
1978	490	8,053	16.4
1979	520	13,221	25.4
1980	553	10,343	18.7
1981	532	10,736	20.2
1982	572	13,845	24.2
1983	655	13,648	20.8
1984	859	18,217	21.2
1985	877	15,024	17.1
1986	921	17,423	18.9
1987	865	19,078	22.1
1988	910	23,326	25.6
Total	12,028	248,515	Average: 18.75
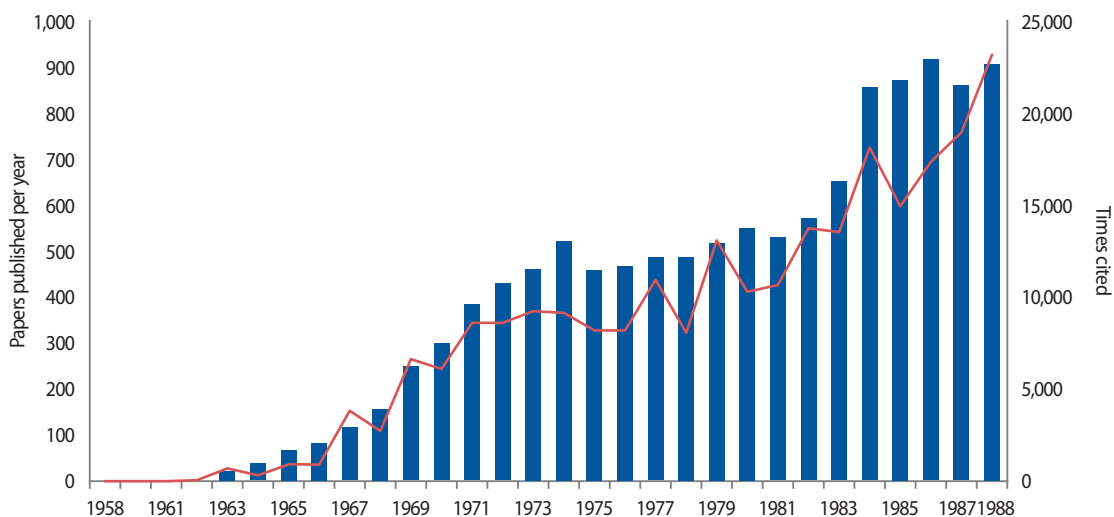
Fig. 2. Publications by the University of Waterloo 1958 to 1988 and times cited.

the university's development up to 1988. To provide some context for this analysis, we see that the university published 12,028 papers from 1958 to 1988. This is illustrated in Fig. 2 (with associated data in Table 1). We see that after a slow start in the early 1960s, the university was publishing a thousand papers per year by the end of the 1980s. On average these twelve thousand papers have been cited 18.75 times, although this statistic is heavily skewed by highly cited outliers.

While much of the earlier work in this field has relied on programmatic approaches employing SQL to scan a local database of citation data for patterns that match certain threshold criteria (for example, in Redner's 2005 study, over a century's worth of publications of the American Physical Society were searched), the much smaller amount of data collected for this study permitted a manual approach to identifying SBs.

Working at the level of an institution (that is to say, for several thousand records), the citation-by-year data of Scopus can be exported as a delimited text file and then sorted using a spreadsheet application such as Microsoft Excel. From this point, the technique for identifying unusual citation patterns is straightforward: With successive columns listing the citations received in each year after publication for the rows of articles, a sorting is defined in which each successive year is a sorting level. With each sorting level ordered from lowest to highest, those articles with the lowest number of citations appear at the top of the list. One visually scans this layout for a long series of years after an article's publication in which there were zero (or nearly zero) citations, followed by a more recent increase. The BC can then be calculated using the technique of Ke et al. (2015) for the small number of articles that are identified as having the characteristics of a SB.

Note that Ke et al. (2015) do not specify a threshold value for determining the significance of the BC: "There are no clear demarcation values that allow us to separate SBs from 'normal' papers: delayed recognition occurs on a wide and continuous range." While this new metric measures the magnitude of the awakening, it does not offer a mechanism for determining whether a paper is a SB or not. This is due to the fact that they find that BCs exhibit a scale-free distribution when calculated for articles in both the Web of Science and American Physical Society databases (Redner, 2005). This implies that there is no characteristic value for the BC and that while it must be a positive value, it may range from null to an arbitrarily large number.

While the BC follows a scale-free distribution, it is also true that the greater the BC, the more sudden and surprising

are the citations to a re-awakened article and the more that article fits the definition of a SB. Thus there must be a practical lower limit below which the recognition, delayed as it may be, is simply too small to signify any meaningful impact on the current research front. Given that Ke et al. (2015) considered a BC value of 30 as being "small," we will use a threshold of 100 for the BC as the lower limit of what constitutes a meaningful SB.

3. RESULTS

We identified five articles that were published by faculty or graduate students of the University of Waterloo that exhibit a clear SB citation pattern (Table 2, Fig. 3). The earliest SBs we discovered were published in 1971, and the most recent was published in 1987. As SBs are known to be quite rare, it is not surprising that there should only be a handful from any given institution.

A statistical description of these articles illustrates just how unusual they are. Besides the BC, the unusual nature of these articles is illustrated by calculating their average citation age. Note that the concept of "citation age" was defined by Redner (2005) as "The age of a citation is the difference between the year when a citation occurs and the publication year of the cited paper." To arrive at an average citation age, the number of citations in a given year is multiplied by the number of years since publication to generate an age-weighted citation count. The overall total of the citation ages for all years is divided by the total number of citations to determine the average citation age.

Redner (2005) looked at a century's worth of articles in the journal *Physical Review* and found an average citation age of 6.2 years. Most articles have most of their impact within a few years of publication. In contrast we see (Table 2) that the average citation age of these five articles from the University of Waterloo is considerably longer and ranges from 21.5 to 40.2 years (calculated up to the year of peak citation). Thus the peak of citations to these SB articles happens decades after most articles have had an impact on their field.

The most striking result is the 1974 article by Horndeski in the *International Journal of Theoretical Physics*. Since 2010 the growth of interest in this paper has been explosive, awakening in 2011 with 16 citations and reaching a peak of 153 citations only five years later. This is an ideal SB citation pattern. Note that this article was cited once in 1976, 1977, and 1983 by other researchers, indicating that the paper was indexed by citation databases and was potentially

Table 2. Five Sleeping Beauty articles published by the University of Waterloo

| Year & field	Author. "Title" Journal	Total citations	Average citation age	Beauty Coefficient
1974 Physics	Horndeski G. W. "Second-order scalar-tensor field equations in a four-dimensional space" *International Journal of Theoretical Physics*	619	40.2	2,434
1980 Physics	Collins C. B., Glass E. N., Wilkinson D. A. "Exact spatially homogeneous cosmologies" *General Relativity and Gravitation*	186	28.9	315
1971 Physics	Lovelock D. "The Einstein tensor and its generalizations" *Journal of Mathematical Physics*	1,180	35.2	286
1981 Computer science	Mark J. W., Todd T. D. "A nonuniform sampling approach to data compression" *IEEE Transactions on Communications*	116	29.9	117
1987 Computer science	Kilgour D. M., Hipel K.W., Fang L. "The graph model for conflicts" *Automatica*	119	21.5	103
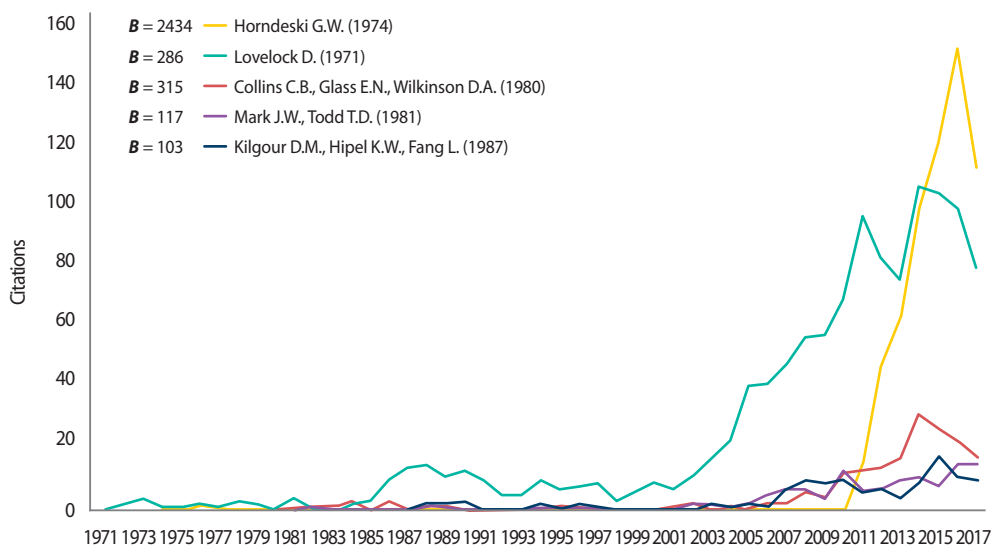
Fig. 3. Citation history of five Sleeping Beauty papers from the University of Waterloo. The Beauty Coefficient for each paper is shown as the value "B."

discoverable. After this, the Horndeski paper did not receive another citation for 28 years. Interestingly, and despite being four decades younger than the paper by Einstein, Podolsky, and Rosen (1935), citations to this article spiked so suddenly after 2010 that its BC of 2,434 even exceeds that of the EPR paper which has a value of 2,333. While the EPR paper has been cited many more times than Horndeski's 1974 paper, it is actually the latter that has had a more sudden and surprising impact on its research field.

Entitled "Second-order scalar-tensor field equations in a four-dimensional space," the Horndeski paper proposes a highly theoretical reimagining of what gravity is. A review of the recent articles that cite it indicates that this paper has become central to the understanding of Galilean gravity models. Indeed the paper has become something of a

name brand within this sub-specialty of cosmology. Of the citing articles, 100 use that researcher's name to represent the associated concept: "Horndeski theories," "Horndeski model," and "Horndeski gravity." A deeper analysis of the evolution of this very complex and theoretical topic in cosmological physics is beyond the scope of this paper. But from a bibliometric perspective, it is sufficient to note that the Horndeski paper lay dormant for a quarter of a century and in the relatively short period since its awakening in 2011 has suddenly became relevant to the understanding of gravity. This is an example of the second type of driver for the emergence of an SB, that of an idea that was ahead of its time. This represents, on a very small scale, what Thomas S. Kuhn described as a "paradigm shift" in science.

The citation history of the Horndeski and another

SB paper, Lovelock's "The Einstein tensor and its generalizations" (1971) are entwined: Horndeski was the graduate student of Lovelock and the Horndeski paper cites Lovelock's paper. Being an extrapolation of Lovelock's ideas, the Horndeski (1974) paper took on new relevance once the Lovelock paper was awakened and it was through the latter that researchers were presumably led to the Horndeski paper. Indeed, these papers have been co-cited 68 times in the Scopus database as of March 2018.

Two articles in computer science from the 1980s are also somewhat surprising: Kilgour, Hipel, and Fang (1987), and Mark and Todd (1981). While they have each been cited more than 100 times, the pattern of citations to these articles is of a more gradual awakening rather than a sudden spike of interest. This moderates just how high their BC score can be. Still, they were both very much dormant for two decades, and the average age of the citations to these is 21.5 and 29.9 years, respectively, so their BCs are greater than 100 and they can both rightfully be considered SBs.

Note that while an article's BC is related to the number of citations it receives, it is not strictly proportional. Consider two articles in physics: Lovelock (1971) has received 1,180 citations (as of November 2017, Scopus) and has a BC of 286. In contrast, Collins, Glass, and Wilkinson (1980) has received only 186 citations and yet has a higher BC of 315. This is because the surprisingness of the spike in citations to the Lovelock paper after 2001 is muted by the modest attention it received in the 1980s. We see that the BC algorithm of Ke et al. (2015) takes into account both the depth of the sleep and the suddenness of the awakening in determining how much of an SB a paper represents.

Given that SBs have been found in a wide range of research fields, and that Glänzel and Garfield (2004) found twice as many SBs in life sciences as in physics, it may seem curious that the delayed recognition papers from the University of Waterloo occur in only physics and computer science. This is no doubt a reflection of the history and research strengths of the university, which has no medical school and which was founded in 1958 with a focus on engineering, math, and computer science. This legacy continues to this day, with the university being ranked as having the 70th best program in engineering and technology (which includes computer science) in the world according to the 2018 QS World University Rankings (https://www.topuniversities.com/subject-rankings/2018). The natural sciences (which includes physics) also do well, and the university is ranked as the 116th best program in the world. It is therefore not surprising that no articles from the social sciences were found amongst the SBs from the University of Waterloo, as this faculty is a smaller and more recent part of the organization.

4. CONCLUSION

By identifying SBs in the publication history of the University of Waterloo, we have uncovered the legacy of some of the research performed there decades ago. Rather than being forgotten these unusual examples of scholarship have found new life, contributing to the research fronts of physics and computer science. Despite being a middle-sized university founded only 60 years ago, the University of Waterloo has produced a handful of SBs, including one even more surprising (in terms of the suddenness of its impact as measured by its BC) than the Einstein, Podolsky, and Rosen paper.

While the University of Waterloo is renowned within Canada for its research in such fields as computer science, nanotechnology, and engineering, it is not an exceptional institution in the context of higher education globally. It is therefore not unreasonable to expect that many other universities around the world have also produced research that has lain dormant for many years and that has recently been rediscovered. The technique outlined here could be used at other institutions to identify researchers who were ahead of their time.

The rationale for doing so is not simply esoteric. The discovery of this unusual citation pattern in the historical publications of an institution presents it with an opportunity to think about the use of bibliometrics in a new way. In contrast to the negative reputation that bibliometrics has gained as a result of its inappropriate use in judging faculty, SBs are a thoroughly positive application of bibliometrics because they celebrate work that has been overlooked. Indeed, the fairy tale analogy implied by the term "Sleeping Beauty" is not simply that the articles have been awakened, but that the story has a happy ending. For faculty who have been conditioned to view bibliometrics as merely a form of accounting, SBs demonstrate that bibliometrics can instead be used to construct a positive story about the use of citations in describing research.

How then can universities such as Waterloo capitalize on this uncommon research legacy? One approach would be to use SBs in a communications plan to highlight the most impactful research in the history of the university. This bibliometric technique is easily applied at any university with access to the appropriate databases. Once identified, the SBs in an institution's publication record demonstrate the legacy of the groundbreaking research that was performed there.

ACKNOWLEDGMENTS

REFERENCES

Braun, T., Glänzel, W., & Schubert, A. (2010). On Sleeping Beauties, Princes and other tales of citation distributions. *Research Evaluation*, 19(3), 195-202.

Collins, C. B., Glass, E. N., & Wilkinson, D. A. (1980). Exact spatially homogeneous cosmologies. *General Relativity and Gravitation*, 12(10), 805-823.

Demaine, J. (2018). DATA.xlsx (version 1). *figshare*. Retrieved Jun 30, 2018 from https://doi.org/10.6084/m9.figshare.6464840.v1.

Einstein, A., Podolsky, B., & Rosen, N. (1935). Can quantum-mechanical description of physical reality be considered complete? *Physical Review*, 47(10), 777-780.

Fine, A. (2017). Einstein-Podolsky-Rosen argument in quantum theory. *The Stanford Encyclopedia of Philosophy*. Retrieved Jun 30, 2018 from https://plato.stanford.edu/archives/win2017/entries/qt-epr/.

Fraser, D. C. (1958). Decay of immediate memory with age. *Nature*, 182(4643), 1163.

Glänzel, W., & Garfield, E. (2004). The myth of delayed recognition. *The Scientist*, 18(11), 8-9.

Ho, Y.-S., & Hartley, J. (2017). Sleeping Beauties in psychology. *Scientometrics*, 110(1), 301-305.

Horndeski, G. W. (1974). Second-order scalar-tensor field equations in a four-dimensional space. *International Journal of Theoretical Physics*, 10(6), 363-384.

Huang, T.-C., Hsu, C., & Ciou, Z.-J. (2015). Systematic methodology for excavating sleeping beauty publications and their princes from medical and biological engineering studies. *Journal of Medical and Biological Engineering*, 35(6), 749-758.

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.

Kilgour, D. M., Hipel, K. W., & Fang, L. (1987). The graph model for conflicts. *Automatica*, 23(1), 41-55.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.

Lange, L. L. (2005). Sleeping beauties in psychology: Comparisons of "hits" and "missed signals" in psychological journals. *History of Psychology*, 8(2), 194-217.

Lovelock, D. (1971). The Einstein tensor and its generalizations. *Journal of Mathematical Physics*, 12(3), 498-501.

Mark, J. W., & Todd, T. D. (1981). A nonuniform sampling approach to data compression. *IEEE Transactions on Communications*, 29(1), 24-32.

Redner, S. (2005). Citation statistics from 110 years of *Physical Review*. *Physics Today*, 58(6), 49-54.

Scopus (2018). [Search string = AF-ID ("University of Waterloo" 60014171)]. Retrieved Aug 27, 2018 from https://www.scopus.com.

Teixeira, A. A. C., Vieira, P. C., & Abreu, A. P. (2017). Sleeping Beauties and their Princes in innovation studies. *Scientometrics*, 110(2), 541-580.

van Raan, A. F. J. (2004). Sleeping Beauties in science. *Scientometrics*, 59(3), 467-472.

Završnik, J., & Kokol, P. (2016). Sleeping Beauties in pediatrics. *Journal of the Medical Library Association*, 104(4), 313-314.