Physica A 603 (2022) 127807

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

On the dynamics of reporting data: A case study of UFO sightings

Fernando J. Antonio [a,*], Andreia S. Itami, Jônatas F. Dalmedico [b], Renio S. Mendes [c]

[a] Departamento Acadêmico de Ciências da Natureza, Universidade Tecnológica Federal do Paraná, Av. Alberto Carazzai, 1640, Cornélio Procópio, 86300-000, PR, Brazil
[b] Instituto de Física e Matemática, Universidade Federal de Pelotas, Capão do Leão, SN, 96160-000, RS, Brazil
[c] Departamento de Física, Universidade Estadual de Maringá, Av. Colombo, 5790, Maringá, 87020-900, PR, Brazil

ARTICLE INFO

Article history:
Received 3 November 2021
Received in revised form 17 May 2022
Available online 29 June 2022

Keywords:
Human reporting events
Phase transition
Rounding numbers
Rank distribution

ABSTRACT

There are a growing number of large databases online. In this study, we used data from 80,332 cases of UFO sightings reported from 1906 to 2014 as a tool for investigating the human behavior in reporting events and how a time lag influences the patterns of reporting events. We identified that new reports were sensitive to media broadcasting and daytime hours. This last pattern was ruled by different reporting mechanisms distinguishing daytime and night hours with an evident phase transition in the rank distribution. Also, we verified that more than 41% of the sightings reported supposedly happened in perfect o'clock hours, indicating that observers have a strong preference for rounder numbers, a result that could be relevant for investigations of experimental data collected by humans in mechanical and analog measuring instruments.

This rounding pattern also led to different mechanisms underlying the rank distribution of round or non-round minutes. Finally, we found numerical evidence that observers have an urgency in reporting the sighting. The probability distribution of reporting time lag revealed a critical reporting lag of 83.5 days after the contact. Additionally, we identified that the month-day rounding mechanism varies with time being connected to the critical reporting lag obtained.

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

Humans have odd behavior. Apart from some exceptions, in general, humans tend to round values even in answer to a simple question such as *'What time is it?'*. Another simple test is to check alarm clocks. Indeed, we will see many alarms set at 6 o'clock, but rarely alarms are programmed to sound at 5:58 h, 5:59 h, 6:01 h, or even 6:02 h. Prevalently, humans prefer rounder number, *i.e.*, numbers ending in 0 or 5. Also, it has been identified that round numbers can act as a goal that influences the behavior of baseball players [1]. Such rounding pattern can be associated with a preferential choice that has already been observed in situations such as clustering on round numbers in trading in global markets [2], in orders of Israeli IPO auctions [3], on round fractions of stock prices [4] and on the income numbers of Indian companies [5]. Nevertheless, this rounding can be originated as an approximation performed because the speaker forgot the exact value. Such a case seems to be true for situations that happened in the past, such as in simple addition problems [6].

* Corresponding author.
E-mail address: fjantonio@utfpr.edu.br (F.J. Antonio).

https://doi.org/10.1016/j.physa.2022.127807
0378-4371/© 2022 Elsevier B.V. All rights reserved.

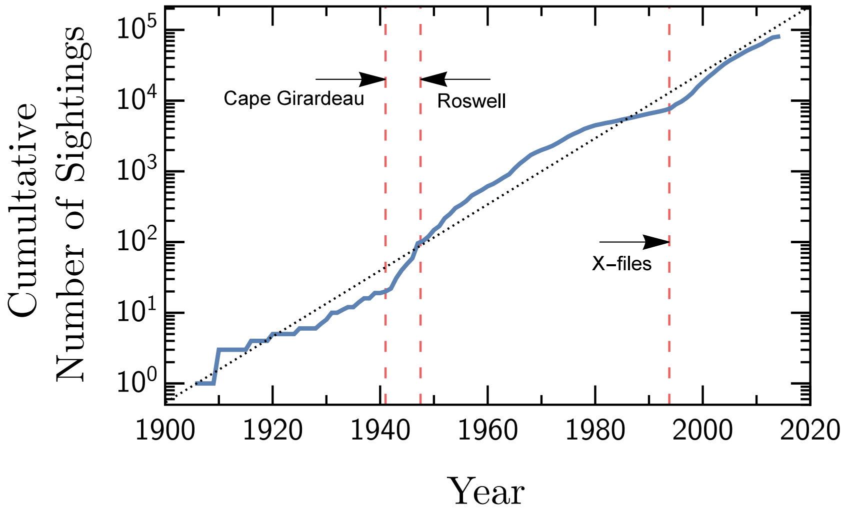

Fig. 1. The accumulated number of sightings since 1906. The vertical axis is in logarithmic scale and the dashed lines represent two UFO sighting events massively broadcasted by the media and the launch of the popular TV series X-files. The straight dotted line represents an exponential growth with coefficient 0.1077 year^{-1} (95% CI: (0.1049, 0.1104), $R^2 = 0.9828$).

As technology advances, information becomes more accessible and more broadcasted. As a result, data on events are more reported and stored than ever. Examples include a large variety of fields, incorporating animal population data [7,8], data on Twitter for predicting epidemics [9] and data on conflicts available [10]. In particular, studies concerning human biological aspects such as comportment [11,12] and performance enhancement [13] have raised great interest in the scientific community in the last years. Among that universe of data already available, there is still a need for datasets related to memory and rounding patterns that cover different periods of time, from a few minutes up to many years. As a consequence, there are not many empirical studies on that subject.

Our primary goal in this study was to investigate human behavior by analyzing data of reported sighting events. We focused on the reporting dynamics rather than the sighting events per se. The first approach was to identify patterns concerning sighting hours reported by observers. Another analysis was to investigate patterns of sighting minutes. We also evaluated a reporting lag (time between the sighting and the reporting date) and how the reported values were connected to a feature of rounding numbers.

2. Data presentation and methods

In this work, we employed data from UFO sightings dating back as far as 1906 up to 2014. The data contain 80,332 records with details about sighting date, geolocation of the event, sighting duration, UFO shape, standardized date and time of the sighting, and the documented date (when the UFO sighting was reported, not to be confused with the sighting date). These data are freely available on the Kaggle platform [14] and were curated by The National UFO Reporting Center (NUFORC) [15]. We used the data as an empirical observation of the human memory of reporting events. In this scenario, we investigated patterns of rounding numbers and their association with ancient memories.

3. Data analysis and discussion

3.1. Profile of data

As expected, the number of sightings cases has not increased constantly over time. It has been growing since 1906, when the first case in our database was reported, as evident in Fig. 1. The continuous line is the number of sighting cases accumulated by year. The straight dotted line was obtained as a fit for the data. It indicates that the general growth in the number of reported cases by year follows an exponential behavior with coefficient 0.1077 year^{-1} (95% CI: (0.1049, 0.1104), $R^2 = 0.9828$). It is a fact that the world's population is increasing but at a much slower rate. It is also evident in Fig. 1 a sharp increase after famously broadcasted cases such as Cape Girardeau and Roswell cases and even the famous TV series X-files.

The Cape Girardeau Missouri Crash was an incident that awoke mass media interest. It happened in October 1941, when a non-identified round metallic object fell, and supposedly three small dead bodies of non-humans were captured [16]. From Fig. 1, there is a remarkable, positive growth tendency in the number of sighting cases reported. Another event highlighted in the figure is The Roswell case which happened in June 1947 near Roswell, New Mexico, and it is usually

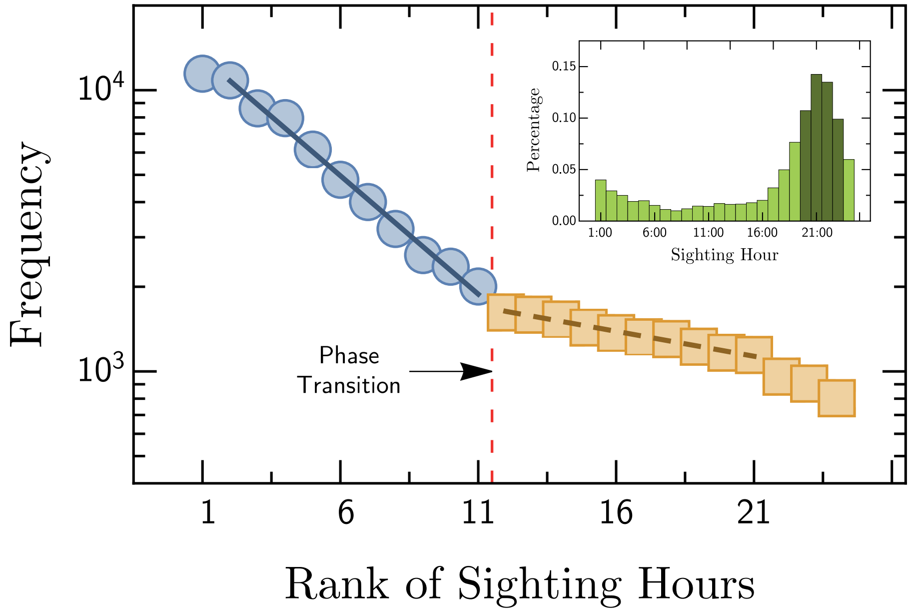

Fig. 2. Rank distribution of the hours of each sighting event report. The vertical axis given in logarithmic scale evidences a phase transition separating daytime and night hours. The first decay is a faster exponential covering hours from 17:00 h to 3:00 h. The slope was found to be −0.194 (95% CI: (−0.207, −0.181], $R^2 = 1.000$) through a non-linear fit to the data. The second decay is slower and it refers to hours from 4:00 h to 16:00 h. The slope of the curve is −0.041 (95% CI: (−0.044, −0.037), $R^2 = 1.000$), obtained through a non-linear fit to the data. The inset shows the percentage of occurrence of truncated values of sighting hours, showing a significantly larger sighting occurrence at night (darker bars).

considered the first accident involving UFO in the United States [17,18]. When case reports were almost stagnated, a new acceleration in growth happened simultaneously with the launch of a very popular TV series called X-Files, which aired first in September 1993 and addressed mystery subjects frequently associated with UFOs [19]. It is also possible that other broadcasted events involving UFO apparitions just after this period reinforced this acceleration. In particular, the *Varginha UFO incident* – a series of events involving the alleged sighting and capture of an extraterrestrial being by the Brazilian military in Varginha, Brazil on 20th January 1996. Such reports got media coverage worldwide [20]. The subject also received global media coverage toward the middle of 1996 with the famous movie *Independence Day* [21].

It is impossible to assure from data if reports or UFO episodes whet new reports, but there is a suggestion that it does influence positively. In some aspects, this pattern is similar to the "the rich gets richer" from network theory explored by Barabási and Albert [22]. Also, it has already been verified that novelties induce a burst of great attention that fades fast [23]. This pattern can be observed by the number of tweets reacting to homicides in London after broadcasting [24]. Curiously, the increase of information flows shortens attention cycles [25]. In particular, novelty within groups decays with a stretched-exponential law, suggesting the existence of a natural time scale over which attention fades.

3.2. Hours of sightings

This section analyzes the frequency of sightings reported in each hour of the day. A sighting peak centered on 21:00 h was identified, as illustrated in the inset of Fig. 2, where the bars represent the percentage of cases reported per hour, truncating the minutes. Almost half of the reported sightings (48.4%) occurred between 20:00 h and 23:59 h (see the darker bars in the inset of Fig. 2).

A rank–size histogram of the frequency of sighting hours (see Fig. 2) reveals a phase transition distinguishing the two regimes. The first regime is an exponential decay covering the period from 17:00 h up to 3:00 h, usually nighttime. A non-linear fit to these data led to a straight line whose slope is −0.194 (95% CI: (−0.207, −0.181], $R^2 = 1.000$). The second regime fits well with slower exponential decay, referring to daytime hours and concerning hours varying from 4:00 h to 16:00 h. The right dashed line has to slope −0.041 (95% CI: (−0.044, −0.037), $R^2 = 1.000$) and was obtained by a non-linear fit to the data.

Depending on the sunlight, the presence of two different regimes might be connected to the circadian rhythm rather than to a "UFO visiting pattern". This assumption is supported by the fact that the most-reported UFO shape (20.6% of all the cases) was *lights*. In fact, it is easier to identify lights during nighttime.

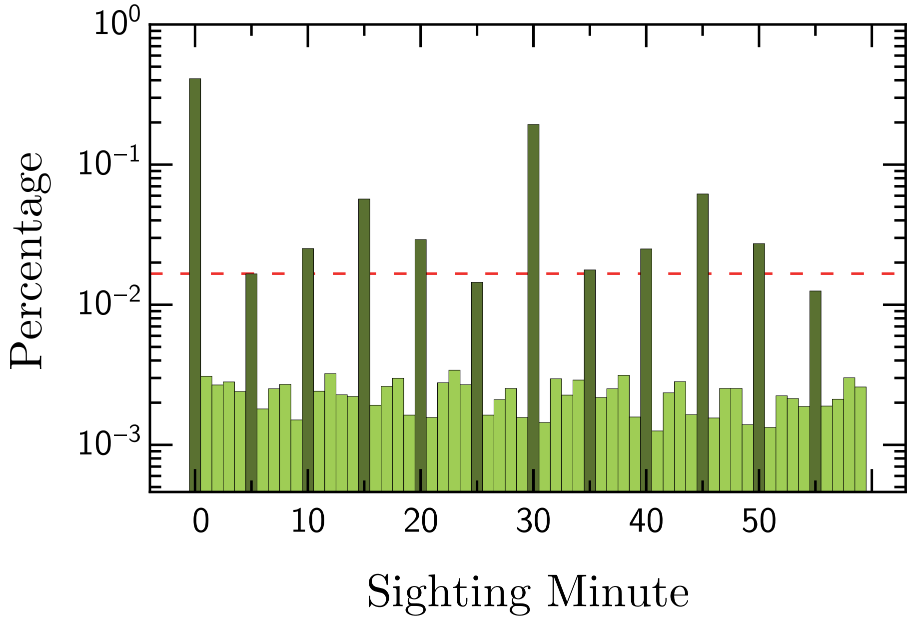

Fig. 3. Normalized histogram of the minutes of the reported cases shown in logarithmic scale on the vertical direction. All the minutes ending in 0 or 5 were darkened to evidence the different behavior of round minutes. The dashed line represents an expected equiprobable value.

3.3. Minutes of sightings

The analysis of raw data exposed that more than 41% of the sightings reported supposedly happened in perfect o'clock hours. Another sharp peak for cases that occurred in half hours is also noticeable. This preferential tendency for o'clock or half hours sums up to a remarkable amount of 60.4% of the cases reported. In contrast, for a completely random scenario, it should be expected to be less than 1.67% in each one of these groups. A possible explanation is that it is a consequence of simplicity to optimize the relevance of the information [26].

Aiming to check the rounding tendency on our data deeply, we have built a normalized histogram of the minutes present in each sighting report. As depicted in Fig. 3, a rounding preference became evident in comparison with the red dashed line, which represents a distribution for equiprobable values.

It is also visually noticeable in Fig. 3 not only the primary preference for integer or half hours but also some taste for minutes ending in 0 or 5 to the detriment of other ending digits. The zero-five rounding tendency just identified leads to another result related to the last digit of the sighting minutes. As depicted in the inset of Fig. 4, minutes ending in 0 were the most reported, appearing in more than 71% of the reported cases. The second most popular ending digit was 5, whose frequency was 18% in the data. In contrast, minutes ending in any digit different from 0 or 5 were found in less than 11% of the reported cases.

The rank distribution of sighting minutes in the logarithmic scale on both axes evidences a phase transition between the pattern for round minutes (ending in 0 or 5) from all the others. The first decay is a fast power law with slope -1.40 (95% CI: $(-1.56, -1.24]$, $R^2 = 0.97$), as illustrated by the straight line in Fig. 4. This regime covers the minutes 0, 30, 45, 15, 20, 50, 10, 40, 35, 5, 25, and 55 respectively. The second regime covers all the minutes ending in values different from 0 and 5, being much slower than the first regime, following an exponential decreasing tendency, as shown by the dashed line in Fig. 4.

3.4. Reporting lag of sightings

Let us define the sightings' reporting lag τ as the number of days between the sighting event and its reporting date. The probability density distribution of τ for the data is available in Fig. 5, pointing toward a phase transition around a critical reporting value τ_c. The critical time even truncates a logarithmic decay distribution into an almost zero-probability distribution.

A non-linear fit to the data leads to a probability distribution function $\mathcal{P}(\tau)$ given by

$$\mathcal{P}(\tau) \sim a \ln\left(\frac{\tau_c}{\tau}\right) \tag{1}$$

in the range $1 < \tau < \tau_c$. This regime corresponds to more than 72% of the reported cases (see the straight line in Fig. 5) and the parameters were found to be $\tau_c = 83.5$ days (95% CI: (10.3, 157), $R^2 = 0.949$) for the critical reporting lag and

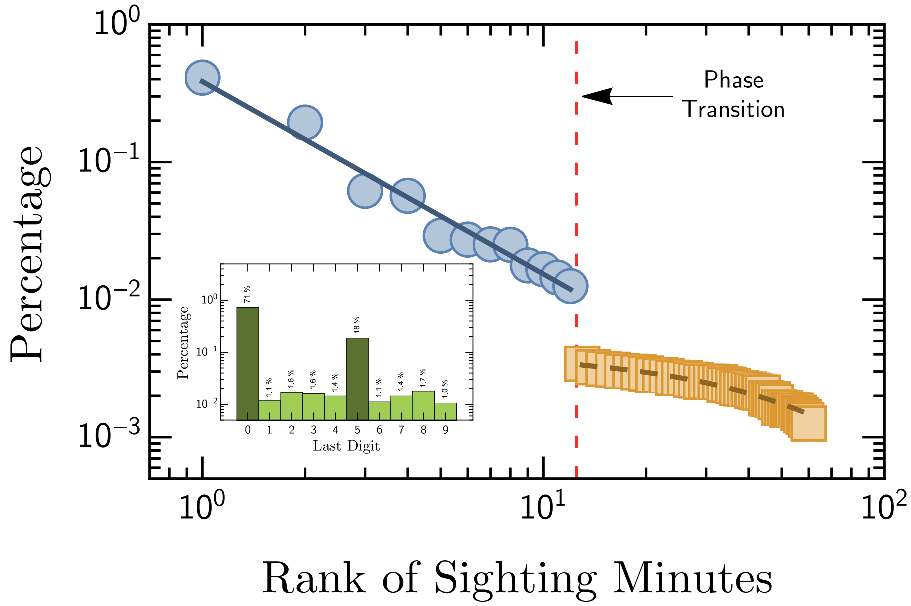

Fig. 4. Rank distribution of the minutes of each sighting case. The curve is shown in logarithmic scale on both axes evidencing two regimes. The first one is a fast power-law decay with slope −1.40 (95% CI: (−1.56, −1.24], $R^2 = 0.97$) obtained by a non-linear fit to the data. This behavior corresponds to minutes ending in 0 or 5, illustrated by the straight line. The second regime is much slower, following an exponential decay, as depicted by the dashed line. This last regime covers the minutes ending in values different from 0 and 5. The inset shows the frequency of the last digit of the minutes reported in each sighting event. The minutes endings in 0 or 5 were darkened.

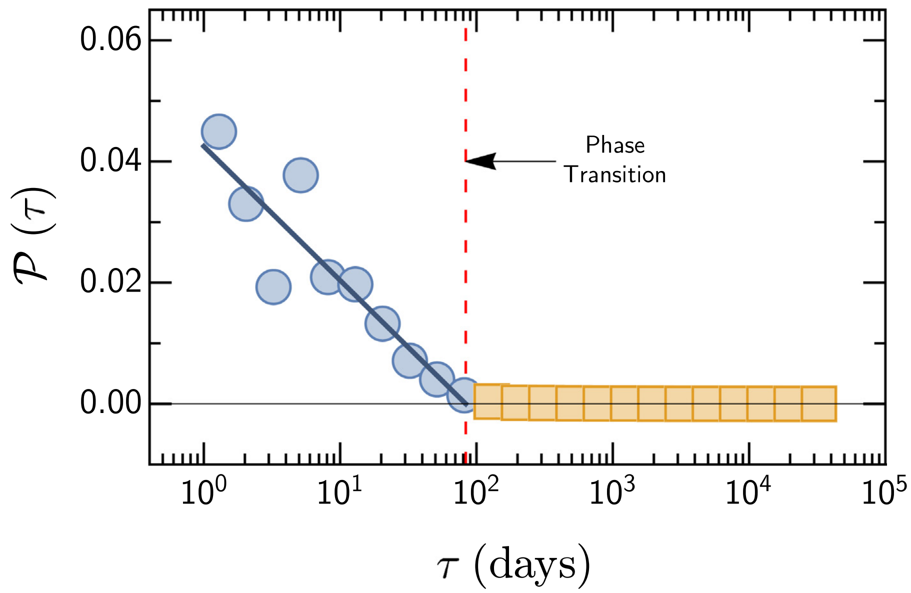

Fig. 5. Probability distribution of reporting lag, the time interval between the sighting date. The horizontal axis is shown on a logarithmic scale. The dashed line corresponds to $\tau_c = 83.5$ days (95% CI: (10.3, 157), $R^2 = 0.949$), the critical reporting lag, a value which defines a phase transition between two regimes: (i) a logarithmic decay with a characteristic critical value τ_c; (ii) an almost zero distribution.

$a = 0.0096$ (95% CI: (0.0062, 0.0130], $R^2 = 0.949$) is a nondimensional constant associated to a normalization and to a decay rate. The second regime in Fig. 5 (when $\tau > \tau_c$) is an almost zero distribution that resembles a power-law.

A mathematical property of logarithmic curves is a fast decay until a critical time. This suggests that observers have urgency in reporting sightings. After the critical time, the news was no longer news, and the case was rarely reported. Many physics applications in the literature involve logarithmic decays. Examples include relaxation in a colloidal

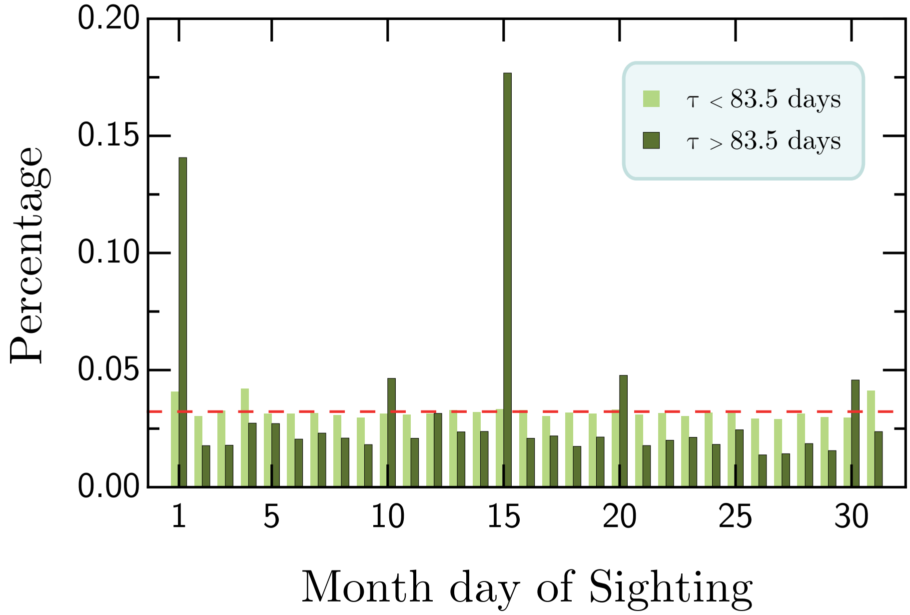

Fig. 6. Influence of the reporting lag time on the month day reported for each sighting event. The dashed line represents the expected value for reporting days randomly. For several days under the threshold of 83.5days (estimated from Eq. (1)), the month day of the sighting follows an approximately constant distribution as the left lighter bars tend to the dashed line. Nonetheless, a rounding pattern becomes evident after the threshold, sharpening the days first and fifteenth. In this figure, the frequency of each month day was normalized accordingly to its frequency throughout a year.

system [27], in models with hierarchically constrained dynamics [28] as well as in the force relaxation in the magnetic levitation of superconductor systems [29].

3.5. Month day of sightings

The next point investigated was how the reporting lag influences the month day when the sighting "happened". The percentage of reports for each month day is available in Fig. 6. To remove bias, the frequency of each month day was normalized accordingly to its frequency throughout four years.

We found that observers tend to report the sighting date reasonably for a few days after the contact (take $\tau < 83.5$ days). The percentage of sightings reported each day was well distributed around the dashed line, corresponding to a uniform probability distribution. However, for a more significant number of days after the sighting ($\tau > 83.5$ days), a distinguished pattern emerges: a sharp rounding pattern in the 1st and the 15th day of the month, as depicted in Fig. 6 by the darker bars. Throughout the database, 14,4% of the sightings are reported to have happened on one of these two month days. It is more than twice the expected value for a uniform probability distribution. A conclusion that can be obtained from this result is that observers forget the actual date of the contact after some time and then start rounding the dates. This result may also be connected with the reporting lag obtained from Eq. (1) and Fig. 5. It may be connected to the irrational form in which human minds convert numerical information for decision-making, as evidenced in price cognition [30].

It is also possible to identify some vestiges of the '0 or 5' rounding pattern in Fig. 6. Nevertheless, they are meaningless for statistical analysis.

4. Conclusion

This study analyzed data from more than 80,000 UFO sighting report cases arousing since the beginning of the last century. We searched for patterns in human behavior when reporting events. From Fig. 1 it was identified that new reports were sensitive to media broadcasting. It was evident that reported contacts happened mainly out of sunlight and that the contrast between daytime and night plays a role in the rank distribution of sighting hours, as identified in Fig. 2. In particular, two different exponential regimes were found for daytime and night hours, respectively. In Fig. 3 we identified a rounding mechanism prioritizing minutes ending in 0 and 5. This result was also visible in the inset of Fig. 4, where almost half of the sightings supposedly happened during an o'clock hour. In addition, such a rounding pattern was connected to a phase transition splitting the rank distribution of minutes ending in 0 or 5 from all the other ending minutes, as evident in Fig. 4. The first group exhibited a power-law decay, while the second group followed a much slower exponential decay.

We identified that observers have an urgency in reporting the sightings, and they tend to round numbers when reporting dates they do not remember, as depicted in Fig. 5. A critical reporting lag of 83.5 days after the contact was obtained. After this period of time, the news was rarely reported, pointing to a lack of excitement in reporting. In particular, after this critical value, month days were rounded for the first or the fifteenth day, that result is apparent in Fig. 6. We hope this work could provide valuable insights into human reporting behavior investigations. Also, we hope other databases could be used further to investigate the rounding pattern and its relationship to memory. In particular, the investigation of the human rounding pattern could lead to concerning results in experimental data collected in both mechanical and analog measuring instruments.

CRediT authorship contribution statement

Fernando J. Antonio: Conceived the idea, Designed the draft, Performed the data analysis, Participated in the production of the text and the analysis. **Andreia S. Itami:** Participated in the production of the text and the analysis. **Jônatas F. Dalmedico:** Participated in the production of the text and the analysis. **Renio S. Mendes:** Participated in the production of the text and the analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors are thankful for the financial support of the Brazilian agencies CNPq, CAPES, Brazil, and Fundação Araucária.

References

[1] D. Pope, U. Simonsohn, Round numbers as goals: Evidence from baseball, SAT takers, and the lab, Psychol. Sci. 22 (1) (2011) 71–79, http://dx.doi.org/10.1177/0956797610391098.
[2] T. Chen, Round-number biases and informed trading in global markets, J. Bus. Res. 92 (2018) 105–117, http://dx.doi.org/10.1016/j.jbusres.2018.07.027.
[3] S. Kandel, O. Sarig, A. Wohl, Do investors prefer round stock prices? Evidence from Israeli IPO auctions, J. Bank. Financ. 25 (8) (2001) 1543–1551, http://dx.doi.org/10.1016/S0378-4266(00)00131-X.
[4] L. Harris, Stock price clustering and discreteness, Rev. Financ. Stud. 4 (3) (2015) 389–415, http://dx.doi.org/10.1093/rfs/4.3.389.
[5] R. Shette, S. Kuntluru, Rounding-up in reported income numbers: Evidence from Indian companies, Rev. Accounti. Finance 13 (2) (2014) 156–170, http://dx.doi.org/10.1108/RAF-04-2013-0042.
[6] J.D. Mason, A.F. Healy, W.R. Marmie, The effects of rounding on memory for numbers in addition problems, Canad. J. Exp. Psychol./Rev. Canad. de Psychol. Exp. 50 (3) (1996) 320–323, http://dx.doi.org/10.1037/1196-1961.50.3.320.
[7] F.J. Antonio, R.S. Mendes, S.M. Thomaz, Identifying and modeling patterns of tetrapod vertebrate mortality rates in the Gulf of Mexico oil spill, Aquatic Toxicol. 105 (1–2) (2011) 177–179, http://dx.doi.org/10.1016/j.aquatox.2011.05.022.
[8] J. Knape, P. de Valpine, Effects of weather and climate on the dynamics of animal population time series, Proc. R. Soc. B: Biol. Sci. 278 (1708) (2011) 985–992, http://dx.doi.org/10.1098/rspb.2010.1333.
[9] A. Alessa, M. Faezipour, Flu outbreak prediction using Twitter posts classification and linear regression with historical centers for disease control and prevention reports: Prediction framework study, JMIR Public Health Surveill. 5 (2) (2019) e12383, http://dx.doi.org/10.2196/12383.
[10] T. Yasseri, R. Sumi, A. Rung, A. Kornai, J. Kertész, Dynamics of conflicts in wikipedia, PLoS One 7 (6) (2012) 1–12, http://dx.doi.org/10.1371/journal.pone.0038869.
[11] H.-H. Jo, M. Karsai, J. Kertész, K. Kaski, Circadian pattern and burstiness in mobile phone communication, New J. Phys. 14 (1) (2012) 013055, http://dx.doi.org/10.1088/1367-2630/14/1/013055.
[12] A. Paraschiv-Ionescu, E. Buchser, K. Aminian, Unraveling dynamics of human physical activity patterns in chronic pain conditions, Sci. Rep. 3 (2013) http://dx.doi.org/10.1038/srep02019.
[13] A.S. Itami, F.J. Antonio, R.S. Mendes, Very prolonged practice in block of trials: Scaling of fitness, universality and persistence, Physica A 424 (2015) 82–89, http://dx.doi.org/10.1016/j.physa.2014.12.036.
[14] UFO Sightings around the world | Kaggle, 2017, Website: https://www.kaggle.com/camnugent/ufo-sightings-around-the-world. Accessed on October 2017.
[15] National UFO reporting center, 2017, Website: http://www.nuforc.org. Accessed on October 2017.
[16] K.D. Randle, CRASH: When UFOs Fall from the Sky: A History of Famous Incidents, Conspiracies, and Cover-Ups, Career Press, 2010.
[17] RAAF Captures flying saucer on ranch in roswell region, 1947, Newspaper: Roswell Daily Record, July 8. P. Front.
[18] K.S. Olmsted, Real Enemies: Conspiracy Theories and American Democracy, World War I to 9/11, Oxford University Press, 2009.
[19] C. Carter (Producer), The X-Files, 1993, TV Series.
[20] Tale of stinky extraterrestrials stirs up UFO crowd in Brazil, 1996, Newspaper: https://www.wsj.com/articles/SB835915673862027500. Accessed on October 2017.
[21] D. Devlin (Producer), Independence day, 1996.
[22] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (5439) (1999) 509–512, http://dx.doi.org/10.1126/science.286.5439.509.
[23] F. Wu, B.A. Huberman, Novelty and collective attention, Proc. Natl. Acad. Sci. 104 (45) (2007) 17599–17601, http://dx.doi.org/10.1073/pnas.0704916104.
[24] O. Kounadi, T.J. Lampoltshammer, E. Groff, I. Sitko, M. Leitner, Exploring Twitter to analyze the public's reaction patterns to recently reported homicides in London, PLoS One 10 (3) (2015) 1–17, http://dx.doi.org/10.1371/journal.pone.0121848.

[25] P. Lorenz-Spreen, B.M. Mønsted, P. Hövel, S. Lehmann, Accelerating dynamics of collective attention, Nature Commun. 10 (1) (2019) 1759, http://dx.doi.org/10.1038/s41467-019-09311-w.
[26] J.-B. Van der Henst, L. Carles, D. Sperber, Truthfulness and relevance in telling the time, Mind Lang. 17 (2002) 457–466, http://dx.doi.org/10.1111/1468-0017.00207.
[27] M. Sperl, Logarithmic relaxation in a colloidal system, Phys. Rev. E 68 (2003) 031405, http://dx.doi.org/10.1103/PhysRevE.68.031405.
[28] J.J. Brey, A. Prados, Slow logarithmic relaxation in models with hierarchically constrained dynamics, Phys. Rev. E 63 (2001) 021108, http://dx.doi.org/10.1103/PhysRevE.63.021108.
[29] B. Smolyak, M. Zakharov, On the force relaxation in the magnetic levitation system with a high-Tc superconductor, Supercond. Sci. Technol. 27 (2014) http://dx.doi.org/10.1088/0953-2048/27/5/055018.
[30] M. Thomas, V. Morwitz, Penny wise and pound foolish: The left-digit effect in price cognition, J. Consum. Res. 32 (1) (2005) 54–64, http://dx.doi.org/10.1086/429600.