Are Skeptics Wrong about the Value of Anecdotes?

Anecdotes can have an important role in science if the correct population for conclusions is identified and special care is taken about bias. They can even be utilized in formal testing of some skeptical hypotheses about anomalous phenomena.

CHARLES G.M. PAXTON

It is therefore not just to consider anecdotes merely as a source of entertainment, because they amuse; if it shall be found that they serve also for the purposes of utility, they will deserve to be classed higher in the scale of study than hitherto they have been.

—Isaac D'Israeli, *A Dissertation on Anecdotes*, 1793

It is generally assumed among many scientific skeptics that anecdotes are not part of science (e.g., Shermer 2003; Hall 2013). The maxim "the plural of anecdote is not data" is well established in skeptical literature, as are variants such as "the plural of anecdote is not evidence" (e.g., Polidoro 2017) and "the plural of anecdote is misinformation" (e.g., Stea 2020). Online skeptical articles that utterly dismiss anecdotal data proliferate (e.g., Fallacy Man 2016; Krej 2013; Novella 2007). I would argue that such dismissal is unjustified and that anecdotal information is rather more used and useful in science than many skeptics imagine.

Causal Stories

Ignoring the point that the original maxim from political scientist Raymond Wolfinger was that "the plural of anecdote is data" (see Nguyen 2015 for a history), the phrase has now become a warning to those who collect anecdotes (here defined per the *Oxford English Dictionary* as "a short account of an amusing, interesting, or telling incident or experience") and potentially think of them as data—or at least as evidence for a particular conclusion. One problem with personal anecdotes, as many skeptics have rightfully discussed, is incorrectly inferring a causal story. "Dancing naked under a full moon cured my COVID-19" may be a compelling personal narrative for a single layperson, but it does not pass muster for a scientist. Assuming the account is subjectively true (the patient did have COVID-19, did dance naked, and did get better), the risk of a causal correlation fallacy is high because essentially it is a statistical claim with a sample size

of one. There is often neither an obvious causal mechanism that would allow a deterministic qualitative argument nor is there a significant statistical argument. Dancing naked may have cured the COVID-19, but the patient might have gotten better anyway; there is no baseline for comparison. The only way to even begin to infer causation would be an experiment or (weaker) an observational study with samples of people dancing or not dancing. If such samples existed, then they could be used to infer an effect. As this set of circumstances seldom exists for complex claims of disease remission, they are rightly dismissed by scientists. The claim of causality is even weaker if the anecdote is at some distance from the original source (e.g., pop singer Nicki Minaj's September 13, 2021, tweet that a COVID-19 vaccination caused her cousin's friend to develop impotence).

However, such tales might still be useful as a source of hypotheses even if they can be manipulated by charlatans and snake-oil salesmen. Anecdotes can be part of the hypothesis formation part of the classic scientific method, when the investigator has some casual observation they want to investigate formally that might indicate a particular cause. Some journals in medicine even allow case reports to be reported (e.g., *Neurocase, Epilepsy & Behavior Case Reports*, *BMJ Case Reports*). This is especially valuable if the syndrome is extremely rare.

However, causal stories are but one type of anecdote. Patients could report symptoms without an initially obvious putative cause, or people might simply recount incidents from their lives that involve some unusual or interesting claim. No generalization or causal argument need be implied. And it's not only medics. Zoologists, for example, may report one-off interesting cases of animal behavior (e.g., Moeliker 2001; Fulling et al. 2017) or sightings of out-of-place animals (e.g., Scheinin et al. 2011). Obviously, an anecdote supported by a photograph or other additional information makes a stronger case than a recollection alone—especially at a distance in time.

Anecdotes as Data

Could these accounts of singular events collectively be considered data? The *Oxford English Dictionary* defines *data* as "related items of (chiefly numerical) information considered collectively." Confusingly, the term can be singular, meaning a dataset or even archaically meaning what we now call a "datum," a single item of information.[1] It stands to reason that any anecdote could be considered a "datum" (they are information, after all) and a collection of anecdotes "data," but that is not the real test of the utility of anecdotes as data in a scientific context. The dictionary also says that data is "typically obtained by scientific work and used for reference, analysis, or calculation." So the real question is whether anecdotes can be used for reference, analysis, or calculation. This means we need to consider the value of anecdotes as quantitative data.[2]

There are three reasons these sorts of anecdote might not be considered data when collected together for analysis, calculation, and reaching conclusions:

1) The anecdotes simply may be untrue.

2) They may be imprecise, in the technical statistical sense that an anecdote may be a relatively noisy recollection of an actual event, perhaps to misperception or memory or exaggeration.

3) Anecdotes could be inaccurate in the statistical sense of biased relative to the truth, which may not actually be known. This could be because of confirmation bias, misperception, lack of controls, or selective memory but also in another important way discussed in the biased anecdotes section below.

Of course, the first reason is also a feature of normal data, so that cannot be grounds for dismissing anecdotes in entirety even if anecdotes are at far greater risk of being untrue. No, it is really the additional reasons that are grounds for rejecting anecdotes. In the case of the second reason, however, if there were many anecdotes of the same event, then the average account might be representative of the data under consideration, because *imprecision* to a statistician means the noise on average is unbiased. So, for example, if there are ten witnesses to a murder, all ten could give an estimate of the height of the murderer. They might all be wrong in their individual estimates, but their mean estimate would be closer to the truth. The more witnesses, the closer the average of the estimates would probably be to the truth *assuming the witnesses are unbiased in their estimates*. This property is extremely useful because it means collected anecdotes can be used as data if they are merely imprecise.

For example, when a drug has passed through clinical trials and gone to market, possibly thousands of patients will take the drug over longer time periods than could be tested in a clinical trial. Low-frequency side effects of drugs can be reported systematically by the patients (this is known as a Phase IV clinical trial) and compared to reports of problems in a control group. Now some of those reported side effects will have nothing to do with the novel drug or treatment, but some of them might. The reported anecdotes about those side effects will occur at a greater frequency in the treatment group compared to the control group. There are also vaccine reporting systems. The individual anecdotes of side effects are imprecise, but collectively they might represent a useful sample. Such unbiased imprecision may well be rare, but if present, anecdotes can be usable data. Here I give two further examples from epizoology and astronomy.

Dogslife is a web-based longitudinal study of Labrador retriever health in the United Kingdom (Clements et al. 2013; Pugh et al. 2015). Dog owners supply their reports of their dogs' health to a central database, which is then used to undertake a longitudinal (in time) study of their health. Owners were recruited and asked to supply quantitative information on the weight and other characteristics of their pets but also any illnesses that did not necessarily result in a visit to the vet. This meant the reports of illnesses were not verified. They were anecdotes, but the results of such a survey can be useful in understanding the spread of diseases, such as an acute vomiting syndrome that appeared in dogs in late 2019 and early 2020 (Woolley et al. 2021). A single story of illness in the life of a Labrador would have little evidential value, but together

such accounts have considerable usefulness.

Similarly, Tatum and Stumpf (2000) used eyewitness estimates of the position of fireballs in the sky to estimate the tracks of meteors over Vancouver Island in 1998. A single account of a meteor might be imprecise, but with multiple witnesses the meteor's progress can be estimated with high precision.

Biased Anecdotes

This leaves the question of biased anecdotes. Anecdotes may simply be inaccurate (as might normal data). Indeed, the first situation above might be looked on as the most extreme form of the third situation. Thus, it is only inaccuracy that *could* render anecdotes as not amenable to science. I say *could* rather than *must* because the bias is a function both of the data *and* the population the conclusions are being drawn about. If you are feeling ill, you might well go to your doctor and report a pain in your abdomen, "a telling experience" for you, as per the definition of anecdote above. If the doctor responds, "This reported experience has no interest to me; it is a mere anecdote," you would rightly feel affronted that relevant evidence was not being considered. Your anecdote is actually representative of your (admittedly subjective) well-being. This addresses the central issue in the consideration of anecdotes: they can be biased but that bias depends on what population is under consideration. By population, I do not mean population in the sense of "the population of the United Kingdom" but in the sense of the *statistical population*—the larger group that conclusions are being made about. Your anecdote about pain is representative of the you at the moment of visiting the doctor. It is not representative for the well-being of neighbors, siblings, the collective inhabitants of your town, or even you at another point in time. But it is still useful to the doctor. It is an unbiased sample of the statistical population of reports of your pain at a particular point in time. Of course, it may be referred pain, meaning the source of the trouble is not where you report the pain, but nevertheless it is information potentially usable by the doctor. Medical diagnosis thus revolves around the use of anecdotes in addition to formal observations such as chemical tests and X-rays. Doctors may not like to think of reported internal symptoms as anecdotes, but surely they are if specific and initially not corroborated.

Likewise, witness statements of a crime, for example, are not dismissed by the judge as irrelevant anecdotes to the proceedings of a court case. The observations are admitted as evidence but with always the concern that the witness might be inaccurate in the formal sense or lying.

The above are singular cases, but what happens if the question is "What proportion of the population reported abdominal pain in a given year in town X?" The population of interest is the whole town. Now if a sample is collected of reports of abdominal pain from patients in the local hospital, then this might well be biased; there could be disproportionately more people with abdominal pain in the hospital than in the wider community (or vice versa if mild stomach complaints are considered). Then these anecdotes would be inappropriate for drawing conclusions about the population of the town—but would be appropriate for drawing a conclusion about the proportion of people in the town's hospital who feel a particular pain.

So anecdotes that are biased cannot be used as data (unless somehow the bias can be corrected for), but whether they are biased can depend on the statistical population under consideration.

Anecdotes about Anomalies

At this point, hopefully the value of anecdotes is clear. But you might still be reluctant to accept the reports of anomalies as data. People report giant ape-like entities often called Bigfoot across North America. The witnesses typically report them as taller than humans. If I were to make conclusions about Bigfoot height from a random sample of anecdotal reports, say from the Bigfoot Field Researchers Organization (BFRO), my conclusions about Bigfoot height (if they actually existed and most especially if they do not exist) may well be biased. However, if I were making conclusions about how tall Bigfoot are *reported to be*, then I can proceed with my analysis (assuming the BFRO database is an unbiased sample of Bigfoot reports; it might not be). Thus, the anecdotes that would be an entirely inappropriate sample for making conclusions about Bigfoot are an entirely appropriate sample for making conclusions about the population of *reports* about the creature. This is useful because while Bigfoot may or may not exist, reports about them undoubtedly do, and this allows us to investigate the sociological and cultural aspects of anomaly reports in a rigorous manner and escape what Harriet Hall (2012) has referred to as "tooth fairy science," performing science on things that may not exist.

For example, a colleague and I (Paxton and Naish 2019) investigated whether there had been changes in how sea serpents were described over the past two centuries by undertaking a statistical analysis of the collected anecdotes. This was a test of a statement asserted, in fact without any evidence, by science fiction writer and skeptic L. Sprague de Camp (1968; 1983) that "After Mesozoic reptiles became well-known, reports of sea serpents, which until then had tended towards the serpentine, began to describe the monster as more and more resembling a Mesozoic marine reptile like a plesiosaur or a mosasaur." We found that de Camp was partially right. There was evidence that sea serpent reports became more like the plesiosaurs (but not mosasaurs) of the Mesozoic Era over time with the presence of necks being more emphasized by reporters (see Figure 1). Of course, we cannot be sure about causation; perhaps something other than increased paleontological knowledge caused sea serpent reports to become more "necky," but the point is, we were not making claims about sea serpents; we were making claims about sea serpent *reports*. Likewise, Paxton and Shine (2016) established the consistency (between and within witnesses) in quantitative measures (length, distance, etc.) from accounts of aquatic

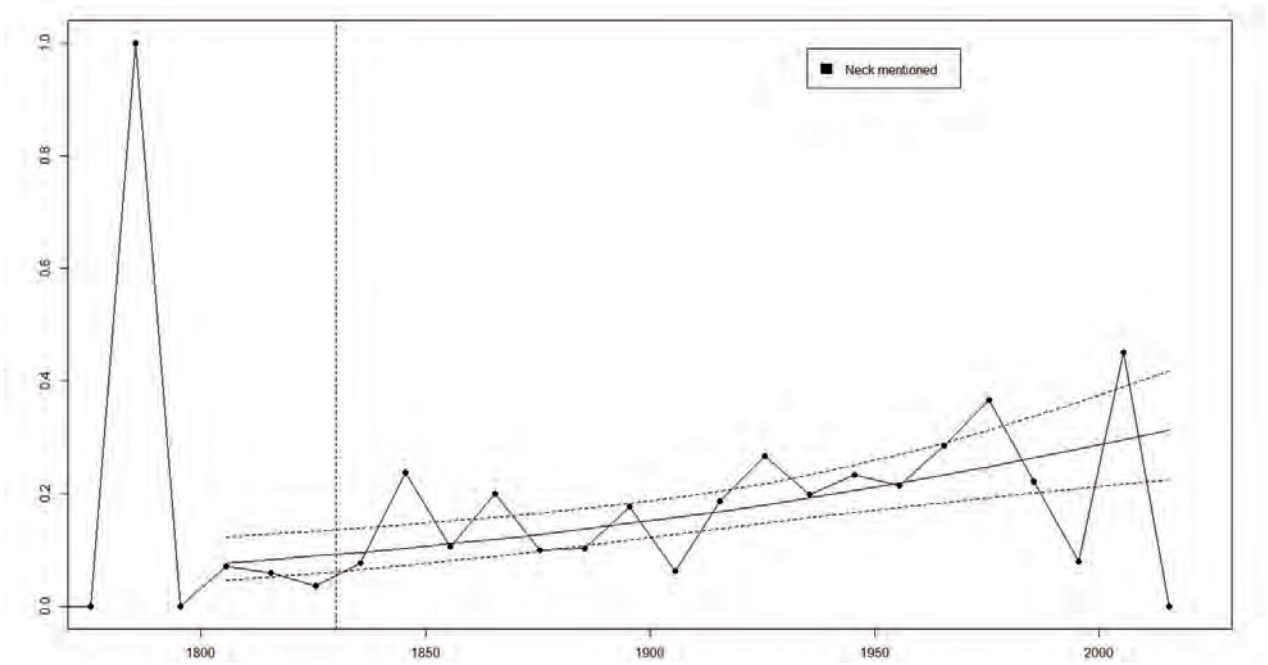

Figure 1: Anecdotes as data. The proportion of sea serpent reports (binned into five years) that mention a neck through time (from Paxton and Naish 2019). Solid lines with dots: actual data. Solid line without dots: fit from a model. Dashed lines: 99 percent confidence intervals on model fits.

monsters. We were not saying sea monsters were consistent but that reports of sea monsters by witnesses in the same encounter were consistent. In these cases, no baseline is required; we can infer things by simply looking at the patterns in the collected anecdotes.

So as long as reports of anomalies are recognized as part of a chain from eyewitness experience to collected analyzed reports, appropriate conclusions may be extractable from the anecdotes. And because many anomaly reports may be creations of human psychology and sociology rather than some actual perturbation of nature, it seems entirely appropriate to consider them in this way and identify changes in the signal in the chain from observation to collated report.

However, there is another point about anecdotes of anomalies. If phenomena are real but very rare, they will only be reported as anecdotes. As we have seen, these may be rare behaviors that could provide insight into animal or human cognition (see Byrne 1997 or Bates and Byrne 2007 for discussions) or phenomena from the natural world. Many low-frequency, albeit important, natural phenomena, such as meteors (Burke 1986), sprites (luminous disturbances above thunderclouds, Boeck et al. 1998), and rogue waves (Draper 1964) were initially reported only as anecdotes. Now, of course, we have corroborative evidence for these phenomena, but originally we did not. One of the few possible approaches to understanding such phenomena would be statistical analysis of the collected anecdotes for identifiable patterns in occurrence. I am currently considering hypotheses of aquatic monster, ghost, and UFO generation in this light. Such an approach allows skeptical hypotheses not to be merely asserted, as alas they so often are, but *tested*. Skeptics can complete the scientific process by testing their hypotheses. The plural of anecdote can be testable data.

Anecdotes can be used to consider other things too. Eyewitness accounts have been used to date the probable extinction of the dodo (Roberts and Solow 2003; Hume et al. 2004), the Danubian population of the Atlantic sturgeon (Jarić et al. 2009), and the Atlantic population of the gray whale (Mead and Mitchell 1984), as well as the possible persistence of the Caribbean monk seal (Boyd and Stanfield 1998). However, these methods assume the testimony is correct, which may not be the case.

Conclusions

So, contrary to the common opinion of many skeptics, anecdotes are far from being solely the material of pseudoscience peddlers and charlatans. Anecdotes are actually vital to the scientific process in generating hypotheses, confirming hypotheses—and, yes, even as data.

But the point about anecdotes is wider. Anecdotal and other forms of what might be called nonconventional information can be used by scientists to investigate and solve real-world problems. Statisticians and scientists can use and even analyze all sorts of information that does not fit the classical definition of data: this can include anecdotes but also indigenous knowledge (e.g., Giles et al. 2016; Ban et al. 2017) and expert opinion (Cooke 1991; O'Hagan et al. 2006), which is useful if a system cannot be readily investigated—for example in estimating if underground pipes need to be replaced. Such "data" may be far from ideal (systematically collected information is always to be preferred) but unconventional information may be a viable choice if no other information is available. Skeptic Robert Carroll (2014) changed his view of the value of anecdotes when he was diagnosed with pancreatic cancer and had to make decisions about treatments not neces-

sarily as yet supported by the complete outcomes of randomized controlled trials. The only information then available was more or less anecdotal supported by expert opinion.

The "plural of anecdote" maxim has been accepted by many in the skeptical movement because it fits their prejudices rather than reflecting actual scientific practice. Yes, we know there are massive flaws in human perception, memory, and reasoning that lead to spurious causal correlation errors and claims of paranormal experiences. Misuse of anecdotes to make dubious causal claims—especially in what is called "alternative medicine"—is a huge problem, but this does not make anecdotes useless. The distinction between anecdote and data is not a dichotomy but a blurry spectrum. The distinction depends on how the anecdotes were collected, the analyses, the population under consideration, and the conclusions being drawn. Just because scientific skeptics lack the ingenuity to think of novel ways in which strange forms of information can be used in an unbiased manner, that does not mean that statisticians and others don't! The question of what constitutes "good" evidence can be a nontrivial problem. The time has come to drop the overly simplistic "the plural of anecdote is not data" slogan. ■

Acknowledgments

My thanks to Valentin Popov and four anonymous reviewers for comments on the manuscript.

Notes

1. Two reasons it is unwise to insist the right form is always "data are"!
2. I consider the use of anecdotes in the qualitative social science a different issue.

References

Ban, N.C., L. Eckert, M. McGreer, et al. 2017. Indigenous knowledge as data for modern fishery management: A case study of Dungeness crab in Pacific Canada. *Ecosystem Health and Sustainability* 3. doi: 10.1080/20964129.2017.1379887.

Bates, L.A., and R.W. Byrne. 2007. Creative or created: Using anecdotes to investigate animal cognition. *Methods* 42(1): 12–21. doi: 10.1016/j.ymeth.2006.11.006.

Boeck, W.L., O.H. Vaughan, R.J., Blakesleeb, et al. 1998. The role of the space shuttle videotapes in the discovery of sprites, jets and elves. *Journal of Atmospheric and Solar-Terrestrial Physics* 60: 669–677.

Boyd, I.J., and M.P. Stanfield. 1998. Circumstantial evidence for the presence of monk seals in the West Indies. *Oryx* 32: 310–316.

Burke, J.G. 1986. *Cosmic Debris: Meteorites in History*. Berkeley, CA: University of California Press.

Byrne, R.W. 1997. What's the use of anecdotes? Distinguishing psychological mechanisms in primate tactical deception. In R.W. Mitchell, N.S. Thompson and L. Miles (eds.), *Anthropomorphism, Anecdotes, and Animals: The Emperor's New Clothes?* Albany, NY: State University of New York Press, 134–150.

Carroll, R. 2014. The skeptical dictionary: Anecdotal (testimonial) evidence. Online at http://skepdic.com/testimon.html.

Clements, D.N., I.G. Handel, E. Rose, et al. 2013. Dogslife: A web-based longitudinal study of Labrador Retriever health in the UK. *BMC Veterinary Research* 9: 13. Online at https://doi.org/10.1186/1746-6148-9-13.

Cooke, R.M. 1991. *Experts in Uncertainty: Opinion and Subjective Probability in Science.* Oxford, UK: Oxford University Press.

De Camp, L.S. 1968. Dinosaurs in today's world. *The Magazine of Fantasy and Science Fiction* 34(3): 68–80.

———. 1983. *The Fringes of the Unknown.* Buffalo, NY: Prometheus Books.

Draper, L. 1964. 'Freak' ocean waves. *Oceanus* 10: 13–15.

Fallacy Man. 2016. 5 reasons why anecdotes are totally worthless. Online at https://thelogicofscience.com/2016/02/10/5-reasons-why-anecdotes-are-totally-worthless/.

Fulling, G.L., T.A. Jefferson, D. Fertl, et al. 2017. Sperm whale (*Physeter macrocephalus*) collision with a research vessel: Accidental collision or deliberate ramming? *Aquatic Mammals* 43(4): 421.

Giles, A., L. Fanning, S. Denny, et al. 2016. Improving the American eel fishery through the incorporation of indigenous knowledge into policy level decision making in Canada. *Human Ecology* 44: 167–183. Online at https://doi.org/10.1007/s10745-016-9814-0.

Hall, H. 2012. Tooth Fairy science and other pitfalls: Applying rigorous science to messy medicine, part 1. *Skeptical Inquirer*. Online at https://skepticalinquirer.org/exclusive/tooth-fairy-science-part-1/.

———. 2013 Down the garden path: Faulty thinking and self-delusion. Skeptical Inquirer 37(4): 57.

Hume, J., D.M. Martill, and C. Dewdney. 2004. Dutch diaries and the demise of the dodo. *Nature* 429: 1. doi:10.1038/nature02688.

Jarić, I., M. Lenhardt, G. Cvijanović, et al. 2009. *Acipenser sturio* and *Acipenser nudiventris* in the Danube—extant or extinct? *Journal of Applied Ichthyology* 25: 137–141.

Krej, 2013. The futility of anecdotal evidence. Online at http://rationalblogs.org/rationalwiki/2013/08/10/the-futility-of-anecdotal-evidence/.

Mead, J.G., and E.D. Mitchell. 1984. Atlantic gray whales. In S. Schwartz, M-L. Jones, and S. Leatherwood (eds.), *The Gray Whale*. San Diego, CA: Academic Press, 33–53.

Moeliker, C.W. 2001. The first case of homosexual necrophilia in the mallard *Anas platyrhynchos* (Aves: Anatidae) *Deinsea* 8: 243–247.

Nguyan, D. 2015. Don't forget: The plural of anecdote is data. *Dan Nguyen's Blog | Thoughts, Data and Computational Journalism*. Online at http://blog.danwin.com/don-t-forget-the-plural-of-anecdote-is-data/.

Novella, S. 2007. No love for anecdotal evidence. Online at https://theness.com/neurologicablog/index.php/no-love-for-anecdotal-evidence/.

O'Hagan, A., C.E. Buck, A. Daneshkhah, et al. 2006. *Uncertain Judgements: Eliciting Expert Probabilities*. Chichester, UK: John Wiley and Sons.

Paxton, C.G.M., and D. Naish. 2019. Did nineteenth century marine vertebrate palaeontological discoveries influence sea serpent reports? *Earth Sciences History* 38(1): 16–27. Online at https://doi.org/10.17704/1944-6178-38.1.16.

Paxton C.G.M., and A.J. Shine. 2016. Consistency in eyewitness reports of anomalies: Examples from accounts of aquatic "monsters." *Journal of Scientific Exploration* 30: 10–17.

Polidoro, M. 2017. Ten practical tactics to unravel the uncanny. Skeptical Inquirer 41(1): 25–26.

Pugh, C.A., B.M. Bronsvoort, I.G. Handel, et al. 2015. Dogslife: A cohort study of Labrador Retrievers in the UK. *Preventive Veterinary Medicine* 122(4): 426–435. Online at https://doi.org/10.1016/j.prevetmed.2015.06.020.

Roberts, D.L., and A.R. Solow. 2003. When did the dodo become extinct? *Nature* 426: 245.

Scheinin, A., D. Kerem, C. MacLeod, et al. 2011. Gray whale (*Eschrichtius robustus*) in the Mediterranean Sea: Anomalous event or early sign of climate-driven distribution change? *Marine Biodiversity Records* 4: E28. doi:10.1017/S1755267211000042.

Shermer, M. 2003. Show me the body. *Scientific American* 288: 27.

Stea, J.N. 2020. The plural of data is misinformation. Online at https://www.psychologytoday.com/gb/blog/writing-integrity/202011/the-plural-anecdote-is-misinformation.

Tatum, J.B., and L.L. Stumpf. 2000. The Vancouver Island fireballs of spring 1998. *Publications of the Astronomical Society of the Pacific* 112(777): 1487–1495. Online at https://doi.org/10.1086/317695.

Woolley, C.S.C., I.G. Handel, B.M. Bronsvoort, et al. 2021. Surveillance of a vomiting outbreak in dogs in the UK using owner-derived and internet search data. *Vet Record* 2021: e308. Online at https://doi.org/10.1002/vetr.308.

Charles G.M. Paxton is a statistical ecologist at the University of St. Andrews who teaches scientific thinking and delves into the history of sea monsters. He tweets @charlespaxton4.