Journal of Scientific Exploration, Vol. 30, No. 1, pp. 16–26, 2016 0892-3310/16

RESEARCH ARTICLE

Consistency in Eyewitness Reports of Aquatic "Monsters"

Charles G. M. Paxton

Centre for Research into Ecological and Environmental Modelling, University of St Andrews,
The Observatory, Buchanan Gardens, St Andrews, Fife, United Kingdom

Adrian J. Shine

Loch Ness Project, Drumnadrochit, Inverness-shire, United Kingdom

Submitted September 27, 2015; Accepted January 26, 2016; Published March 15, 2016

Abstract—Little work has been undertaken on the consistency/repeatability of reports of natural historical anomalies. Such information is useful in understanding the reporting process associated with such accounts and distinguishing any underlying biological signal. Here we used intraclass correlation as a measure of consistency in descriptions of a variety of quantitative features from a large collection of firsthand accounts of apparently unknown aquatic animals (hereafter "monsters") in each of two different cases. In the first case, same observer, same encounter (*sose*), the correlation was estimated from two different accounts of the same event from the same witness. In the second case, the correlation was between two different observers of the same event (*dose*). Overall, levels of consistency were surprisingly high, with length of monster, distance of monster to the witness, and duration of encounter varying between 0.63 and 1. Interestingly, there was no evidence that *sose* accounts generally had higher consistency than *dose* accounts.

Keywords: cryptozoology—eyewitness testimony—memory conformity—anecdotes

Introduction

Anecdotal written accounts of undiscovered species of animal are often considered inadmissible as evidence (e.g., Loxton & Prothero 2013, Shermer 2003, Shermer 1997). Yet anecdotal information has been and continues to be used in organismic biology. There have been efforts to combine conventional survey data with more anecdotal information (e.g., Service et al. 2014, Huntington, Suydam, & Rosenberg 2004, but see McKelvey, Aubry, & Schwartz 2008), and the collation and analysis of anecdotal accounts of phenological events is now commonplace (Birchenough et al. 2015, Fitchett, Grab, & Thompson 2015). In a similar way, insights may be obtained from the collection and analysis of cryptozoological reports (Paxton 2010, Paxton 2009).

It is reasonable to assume that eyewitness reports of unusual natural events might bear comparison with eyewitness reports of traumatic events such as crimes or accidents in terms of the processes operating on eyewitnesses. Just as in forensic analysis of eyewitness testimony, a major interest of the analyser of accounts of natural historical anomalies is determining the underlying reality of what was seen, with the major caveat that such reports like those of other witnesses are the outcome of a process of perception, memory, recollection, transmission, and recording (Loftus 1996), with the subsequent danger of reinterpretation by the analyst. Indeed, the report may not actually be true. Thus any statistical investigation of reported anomalies will explore both the actual phenomena seen *and* the reporting process associated with it.

Understanding the natural historical anomaly reporting process could allow determination of whether future reports of anomalies represent real phenomena as yet not understood or recognized by the scientific community. For example, some genuine, important natural phenomena were originally reported only as anecdotes and dismissed by contemporary commentators, e.g., meteorites (Burke 1986), sprites (Boeck et al. 1998), and rogue waves (Draper 1964). Analysis of reports could also allow an understanding of misperception, why people report false events as true, and the distortions that occur in the report transmission process. Statistical analysis also allows estimation of the consistency that might occur in the anomalies reporting process, which might also be relevant for assessing reliability and understanding noise more generally in the natural historical bibliographic record.

In the case of unknown aquatic animals, the cumulative description curve of giant (>2 m) marine animals through time suggests there are still animals to discover (Paxton 1998), therefore it is just possible, albeit unlikely, that such animals might be seen by non-specialist observers prior to discovery. A large database of reports of unknown aquatic animals, hereafter "monsters," is now available (Paxton 2009), allowing a statistical exploration of the reported phenomena. Understanding the consistency of such reports is a vital precursor to any further analysis because inconsistent reports may be less likely to be reliable. Note that here we use repeatability/consistency as synonyms.

Unlike more general investigations of the consistency/repeatability of memory (Baugerud, Magnussen, & Melinder 2014, Odinot, Wolters, & van Giezen 2013, Krähenbühl, Blades, & Eiser 2009, Bramsen et al. 2001, Smeets, Candel, & Merckelbach 2004), the consistency/repeatability of eyewitness accounts of anomalous events in general, let alone of cryptozoological reports, has, to our knowledge, never been estimated. Here we explore consistency of eyewitness testimony of cryptozoological anomalies by looking at the special cases in the database of aquatic monster accounts where there were quantitative estimates of reported length of the monster, reported distance of the witnesses to the monster, and reported duration of encounter. Consistency was estimated in two different situations. Firstly, where there were repeated accounts of the same event from the same observer (same observer same encounter, *sose* cases), and, secondly, where there were multiple witnesses of the same event (different observer same encounter, *dose* cases). Both cases are of interest, *dose* because it allows insights into (mis)perception between individuals, *sose* because it allows an estimation of how consistent individuals are and provides a standard for comparison with the *dose* cases. All other things equal, *dose* cases might be expected to be less consistent than *sose* cases, and any differences between them in repeatability should be due to differences in estimation between witnesses. Also, *sose* accounts provide a measure of the drift in testimony that may occur over time.

It should be stressed that here we were considering only consistency, which is related to precision (i.e. the amount of "noise" in the reports), as consistent estimates will have high precision. Consistency is not the same as accuracy (i.e. a lack of bias). We could not determine the accuracy of the reports from the analyses undertaken here as we did not know the underlying truth. Estimates may be consistent and precise but still inaccurate relative to the unknown reality.

Materials and Methods

The Available Data

In each analysis, the data (see Supplementary Material file "dose data", and Supplementary Material file "sose data") were based on subsets of a compilation of sea and freshwater monster accounts from various primary and secondary sources, including books, newspaper accounts, and firsthand testimony personally collected by the authors (Paxton 2009) and by the Loch Ness Phenomena Investigation Bureau from 1962 to 1972 (Witchell 1979). Reports had to be of animals seen at the surface of the water. For example, the famous Grant and Spicer reports (Gould 1934) of the Loch Ness Monster on land were not included, nor were the famous underwater images taken in 1972 and 1975 (Scott & Rines 1975). To be included as a report, the body of the putative animal must actually have been seen. For example, monster reports where only a splash and/or wash were seen were not included (e.g., Burton 1961:118, Whyte 1957:30). It is not known if all the reports are truthful or indeed necessarily of living things or animals. Nor can it be known if the witnesses have interpreted anatomy correctly (e.g., Paxton, Knatterud, & Hedley 2005). Exposed and admitted hoaxes or absolutely known misidentifications were omitted from the dataset. Suspected hoaxes or misidentifications were not. Therefore, we make no overall claims as to the truthfulness of the reports under consideration. The data are non-randomly distributed in space and time and are clearly biased in favor of English-language sources: predominantly the British Isles, the United States of America, and Canada. One particular locality dominates: Loch Ness. Only firsthand accounts were considered (i.e. where there are direct quotes of the witnesses) as there is evidence of bias in secondhand accounts (Paxton 2009). Of course, even these direct quotes may not actually be direct quotes of the witnesses but embellishments by reporters.

Response Variables

The data of interest were the easily quantifiable aspects of the reports: length of the monster seen, reported distance of witness from the monster, and the duration of the encounter. Reported length in aquatic monster accounts can represent at least three actual lengths: the estimated total length of the whole animal, or the estimated seen length of the animal above the surface of the water, or sometimes the nature of the length being estimated is unspecified. Therefore, total, seen, and unspecified lengths were considered separately. Reported distances came in several forms. Sometimes the reported distance was the estimated distance on initial sighting, sometimes it was the nearest reported distance, and sometimes it was unspecified (often when the object did not move relative to the observer). In this case, initial and unspecified distances were amalgamated, but the nearest reported distances were considered separately. If a description of a distance was given as "less than *r*," the distance was taken as *r*. Likewise, "at least *r*" was taken as *r*.

Multiple accounts of the same encounter by the same witness. In the case of same observer same encounter (*sose*) cases, from each available encounter, random pairs of reported distance, length, and duration were drawn from the available distances, lengths, and durations from the witness accounts. Because these are quantitative measures, an index of consistency for each sighting characteristic can be made. Repeatability within witnesses was then calculated using a Type 1 intraclass correlation coefficient (ρ, Shrout & Fleiss 1979, see also Zar 1996:398–401) in the statistical programming environment *R* (*R* Developmental Core Team 2014) using the *R* library psych (Revelle 2014) to create an index between zero and one, with one representing perfect consistency, and zero no consistency at all.

These *sose* cases are interesting as they provide an insight into individual witness reliability.

Multiple witness accounts of the same encounter. Repeated reports from different observers of the same encounter (*dose* cases) were treated in a similar manner to the multiple accounts from the same witness above, except in this case the pairs of lengths, distances, and durations were drawn from different randomly chosen witnesses of the same event. Multiple witnesses in accounts of aquatic monsters are apparently common (i.e. witnesses often state that others were present), although written accounts from witnesses are quite rare. Often there is one primary witness who refers to others and/or others append their name to a single written account or the encounter is reported by one witness only. Such reports would not be considered as *dose* reports. Wholly independent accounts of the same incident (i.e. accounts where it can be reasonably concluded that the witnesses are completely unaware of other accounts) are extremely rare, only one putative case of wholly independent reports is known to us (from two witnesses who witnessed a sea monster from the passenger boat *Taiyuan* in 1907, (Heuvelmans 1968:382–383). More often there are multiple quotes from different witnesses who were together at the same time and who would have had ample time to talk to each other prior to giving a statement. Both the latter types of report were considered as *dose* here.

There are accounts where people saw the same monster from different localities, but only distances from observers from the same location were compared. Similarly, there are encounters where one or more witnesses arrived after the event had commenced. These were disregarded from the duration analysis.

Results

Multiple Accounts from the Same Witness of the Same Encounter (sose Cases)

In the database there are 171 encounters with distinct repeat firsthand accounts. Not all accounts from the same encounter have estimates of distance, length, or duration. So the sample size for each analysis is much lower than 171. Multiple accounts by the same witness of the same encounter are not always identical, witnesses strangely do not seem to consult their earlier accounts. For example, the naturalist E. G. B. Meade-Waldo stated the initial distance at which he saw a sea serpent from the yacht *Valhalla* in 1905 was 100 yards (Meade-Waldo & Nicoll 1906). A few years later when writing a letter to the author Rupert Gould (1930:129), he stated the initial distance to be 200 yards. Table 1 gives an indication of the spatial and temporal range of the *sose* data.

TABLE 1
Summary of Events with Either Multiple Observers (dose) or Single Observers with Repeated Accounts (sose)

Class of Cases	Time Range	Location
sose (overall)	1817–2011	Worldwide, freshwater, marine
sose (distance)	1852–2000	Worldwide, freshwater, marine
sose (nearest distance)	1890–2007	Worldwide, freshwater marine
sose (length total)	1852–1975	Atlantic, Mediterranean, freshwater, marine
sose (length seen)	1852–2007	Atlantic, Mediterranean, freshwater, marine
sose (length unspecified)	1933–1998	Loch Ness, freshwater
sose (duration)	1875–2007	Atlantic, freshwater, marine
dose (overall)	1817–2009	Worldwide, freshwater, marine
dose (distance)	1817–2009	Atlantic, Indian, freshwater, marine
dose (nearest distance)	1819–1996	Worldwide, freshwater, marine
dose (length total)	1817–1975	Worldwide, marine
dose (length seen)	1819–1996	Worldwide, freshwater, marine
dose (length unspecified)	1907–2009	Worldwide, freshwater
dose (duration)	1817–2009	Atlantic, Mediterranean, freshwater, marine

In the case of initial or unspecified distance, $\rho = 0.93$ (95% confidence interval: 0.85 – 0.96), n = 32, $P < 0.001$; for nearest reported distance, $\rho = 0.95$ (0.85 – 0.99), n = 12, $P < 0.001$; for estimated total length, $\rho = 1$ (1 – 1), n = 5, $P < 0.001$; for seen length, $\rho = 0.97$ (0.89 – 0.99), n = 11, $P < 0.001$; for unspecified length, $\rho = 0.96$ (0.86 – 0.99), n = 10, $P < 0.001$; and for the case of duration of encounter, $\rho = 0.63$ (0.32 – 0.82), n = 25, $P < 0.001$.

Multiple Witnesses of the Same Encounter (dose Cases)

In the database, there are 190 encounters with distinct multiple firsthand accounts. As in the previous case, not all different witnesses have estimates of distance, length, or duration, so the sample size for each analysis is much lower than 190. Because of the lack of independence between witnesses of the same event, the correlation calculated below should be considered upper bounds, as presumably awareness of other accounts would lead to greater similarities. Table 1 gives an indication of the spatial and temporal range of the data.

In the case of initial and unspecified distance estimates combined, $\rho = 0.94$ (95% confidence interval: 0.89 – 0.97), n = 54, $P < 0.001$; in the case of nearest approach distance, $\rho = 0.68$ (0.25 – 0.89), n = 17, $P = 0.001$; in the case of estimated total length, $\rho = 0.95$ (0.71 – 0.99), n = 6, $P < 0.001$;

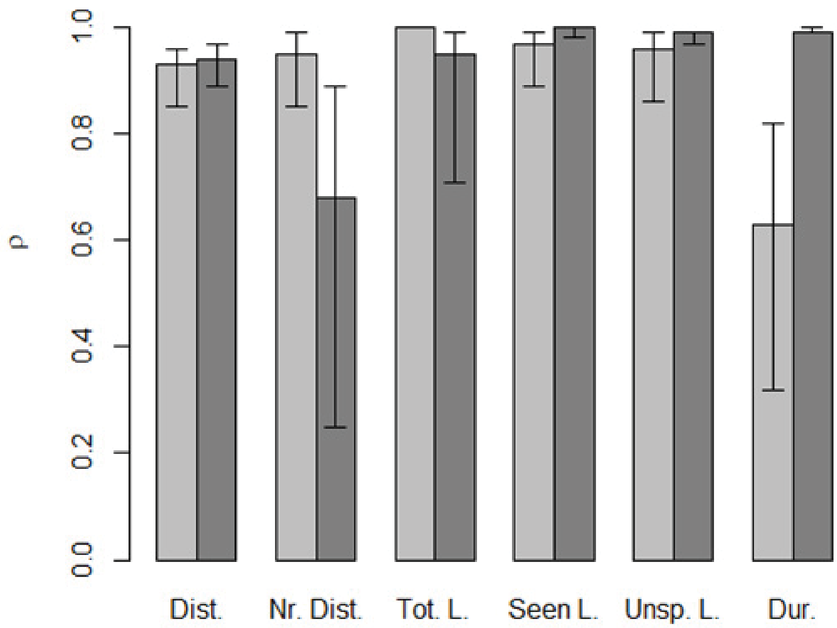

Figure 1. Single witness (sose) consistency (light grey) and multiple witness (dose) consistency (dark grey) from aquatic monster accounts.

in the case of seen length, ρ = 1 (0.98 – 1), n = 8, P < 0.001; in the case of unspecified length, ρ = 0.99 (0.97 – 0.99), n = 28, P < 0.001; and in the case of duration of encounter, ρ = 0.99, (0.99 – 1), n = 61, P < 0.001.

Multiple Witness Accounts Versus Same Witness Accounts

Figure 1 compares the estimated *r* for the different variables in the *sose* and *dose* cases. There is no evidence from this study that *dose* cases are less consistent than *sose* cases.

Error bars indicate 95% confidence intervals. Where error bars cannot be seen, the 95% confidence interval is (1,1).

Discussion

Where feasible, the estimation of repeatability should be a fundamental feature in any assessment of the reliability of anecdotally reported phenomena, assuming reasonably that consistency within and between witnesses is one possible indicator of reliability. To our knowledge, this is the first time consistency has been estimated in reports of anomalies. Witnesses describe seeing aquatic "monsters" as an extraordinary moment in their lives, even "unforgettable" (Dinsdale 1973:69, Dinsdale 1972:93), and so presumably such events are emotionally arousing and should lead to relatively good memory of the encounter compared with more prosaic events. The events are not normally noted as being unpleasant, but instead as rather exciting (e.g., Holiday 1968:111), although there are accounts where the witnesses appear to be perturbed by the event (e.g., the account of Badger in *Daily Star* March 8, 1999). Other investigations of consistency tend to consider consistency in terms of the qualitative features remembered (i.e. are the same features remembered) rather than the consistency of the quantitative estimates. For example, Dutch serviceperson peacekeepers remembered 0.72 of events recorded in an initial survey on resurveying (Bramsen et al. 2001). Thus the consistency scores in themselves do not suggest that the witnesses are generally lying, although it would be useful to formally test the relationship between consistency and truth in reports of anomalies. Where tested in other contexts there is no relationship (e.g., Smeets, Candel, & Merckelbach 2004, Fisher & Cutler 1996). As might be expected, witnesses are not wholly internally consistent in cryptozoological reports, yet the *sose* correlations are quite high, suggesting that individual reports can be looked upon as quite precise. *dose* results are similar. However, given the possibility of reference back to original witness statements in the *sose* cases and discussion between witnesses in the *dose* cases, these results should be looked on as best-case scenarios for eyewitness consistency.

It should be stressed that any inconsistencies seen are not necessarily solely a result of imperfections of memory but also possibly imprecision in the way the report is interpreted and printed by others. Gould (1934) provides examples of witnesses who claimed they were misquoted by the press, although as only firsthand accounts (i.e. witnesses are directly quoted) are considered here this effect should hopefully be mitigated. The results suggest that just as in other forms of memory recollection (Loftus 1996), imprecision does exist in witness accounts of anomalous phenomena. There is no current significant evidence that *sose* reports are more consistent than *dose* reports, although it might be assumed they should be, given different perceptions of different witnesses. The reason for the difference between the *sose* and *dose* duration results is unclear, but, if real, perhaps the duration of the encounter for single witnesses is a somewhat unimportant feature of their experience so it is recalled with little consistency.

The consistency in reports by the same individual is presumably primarily driven by memory and post-witness publication and collection effects (e.g., reporter misquotes, typographical errors, etc). Whereas consistency in reports by different individuals will be driven by all the above and differences between witnesses in the reported dimensions. That *dose* and *sose* accounts have similar repeatability (except in the case of duration) implies that variation due to differences in the reported dimensions is negligible relative to the other factors. This could be because of broad agreement in the dimensions under consideration but also because of memory conformity, that is, that witnesses who are in contact with each other may converge on a common memory of what took place. Alternatively it could be for both the sose and dose situations the post-witness noise in the reporting process overwhelms the variation due to other effects. We cannot distinguish between these mechanisms here. Formal estimates of memory conformity as distinct from repeatability suggest that about 70% of witnesses can be induced into reporting information they did not themselves witness (Gabbert, Memon, & Allan 2003, Wilson & French 2004), so presumably witnesses could influence each other's estimates of dimension.

Precision is not the same as accuracy, but if there was very low precision (hence low repeatability) associated with the quantifiable aspects of reports this would indicate that these features of any single report would be likely to be unsystematically inaccurate. Based on the current albeit best-case evidence here, such low precision is surprisingly not the case.

Acknowledgments

June O'Neill and Roland Watson for supplying much source material on the New England Sea Serpent and the Loch Ness Monster, respectively. Richard Wiseman for comments on the manuscript.

Ethical Statement. There were no ethical issues.

Funding Statement. No sources of funding.

Data Accessibility. The datasets supporting this article have been uploaded as the Supplementary Material files "dose data" and "sose data".

Competing Interests. We have no competing interests.

Authors' Contributions. CGMP suggested the analysis, undertook the statistical analysis, and wrote the paper. AJS supplied some of the source material, and suggested improvements in the manuscript.

References Cited

Baugerud, G. A., Magnussen, S., & Melinder, A. (2014). High accuracy but low consistency in children's long-term recall of a real-life stressful event. *Journal of Experimental Child Psychology, 126:*357–368. doi:10.1016/j.jecp.2014.05.009

Birchenough, S. N. R., Reiss, H., Degraer, S., Mieszkowska, N., Borja, Á., Buhl-Mortensen, L., et al. (2015). Climate change and marine benthos: A review of existing research and future directions in the North Atlantic. *Wiley Interdisciplinary Reviews—Climate Change, 6*(2):203–223. doi:10.1002/wcc.330

Boeck, W. L., Vaughan, O. H., Blakesleeb, R. J., Vonnegut, B., & Brook, M. (1998). The role of the space shuttle videotapes in the discovery of sprites, jets and elves. *Journal of Atmospheric and Terrestrial Physics, 60:*669–677. doi:10.1016/S1364-6826(98)00025-X

Bramsen, I., Dirkzwager, A. J. E., van Esch, S. C. M., & van der Ploeg, H. M. (2001). Consistency of self-reports of traumatic events in a population of Dutch peacekeepers: Reason for optimism? *Journal of Traumatic Stress, 14*(4):733–740.

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

Burton, M. (1961). *The Elusive Monster*. London: Hart-Davis.

Dinsdale, T. (1972). *The Loch Ness Monster*. London: Routledge & Kegan Paul.

Dinsdale, T. (1973). *The Story of the Loch Ness Monster*. London: Target.

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

Fisher, R. P., & Cutler, B. L. (1996). The relation between consistency and accuracy of eyewitness testimony. In *Psychology and Law: Advances in Research* edited by G. M. Davies, S. Lloyd-Bostock, M. McMurran, & C. Wilson, pp. 21–28, Berlin: De Gruyte.

Fitchett, J. M., Grab, S. W., & Thompson, D. I. (2015). Plant phenology and climate change: Progress in methodological approaches and application. *Progress in Physical Geography, 39*(4):460–482. doi:10.1177/0309133315578940

Gabbert, F., Memon, A., & Allan, K. (2003). Memory conformity: Can eyewitnesses influence each other's memories for an event? *Applied Cognitive Psychology, 17*(5):533–543.

Gould, R. T. (1930). *The Case for the Sea Serpent* (first edition). London: Philip Allan.

Gould, R. T. (1934). *The Loch Ness Monster*. London: Geoffrey Bles.

Heuvelmans, B. (1968). *In the Wake of the Sea-Serpents*. New York: Hart-Davis.

Holiday, F. W. (1968). *The Great Orm of Loch Ness*. London: Faber and Faber.

Huntington, H. P., Suydam, R. S., & Rosenberg, D. H. (2004). Traditional knowledge and satellite tracking as complementary approaches to ecological understanding. *Environmental Conservation, 31:*177–180. doi:10.1017/s0376892904001559

Krähenbühl, S., Blades, M., & Eiser, C. (2009). The effect of repeated questioning on children's accuracy and consistency in eyewitness testimony. *Legal and Criminological Psychology, 14*(2):263–278. doi:10.1348/135532508X398549

Loftus, E. (1996). *Eyewitness Testimony.* Cambridge: Harvard.

Loxton, D., & Prothero, D. R. (2013). *Abominable Science*. New York: Columbia University Press.

McKelvey, K. S., Aubry, K. B., & Schwartz, M. K. (2008). Using anecdotal occurrence data for rare or elusive species: The illusion of reality and a call for evidentiary standards. *Bioscience, 58*(6):549–555. doi:10.1641/b580611

Meade-Waldo, E. G. B., & Nicoll, M. J. (1906). Description of an unknown animal seen at sea off the coast of Brazil. *Proceedings of the Zoological Society of London 2:*719–721.

Odinot, G., Wolters, G., & van Giezen, A. (2013). Accuracy, confidence and consistency in repeated recall of events. *Psychology Crime & Law, 19*(7):629–642. doi:1080/1068316x.2012.660152

Paxton, C. G. M. (1998). A cumulative species distribution curve for large open water marine animals. *Journal of the Marine Biological Association of the United Kingdom, 78:*1389–1391. doi:10.1111/j.1469–7998.2009.00630

Paxton, C. G. M., Knatterud, K., & Hedley, S. L. (2005). Cetaceans, sex and sea serpents: An analysis of the Egede accounts of a "most dreadful monster" seen off the coast of Greenland in 1734. *Archives of Natural History, 32*(1):1–9. doi:10.3366/anh.2005.32.1.1

Paxton, C. G. M. (2009). The plural of 'anecdote' can be 'data': Statistical analysis of viewing distances in reports of unidentified large marine animals 1758–2000. *Journal of Zoology, 279:*381–387. doi:10.1111/j.1469-7998.2009.00630.x

Paxton, C. G. M. (2010). The Monster Manual. *Fortean Times, 265*:54.

R Developmental Core Team (2014). *R: A language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing.

Revelle, W. (2014). Library 'psych'.

Scott, P., & Rines, R. (1975). Naming the Loch Ness monster. *Nature, 258*:466–468. doi: 10.1038/258466a0

Service, C. N., Adams, M. S., Artelle, K. A., Paquet, P., Grant, L. V., & Darimont, C. T. (2014). Indigenous knowledge and science unite to reveal spatial and temporal dimensions of distributional shift in wildlife of conservation concern. *PLOS ONE, 9*(7):e101595. doi:10.1371/journal.pone.0101595

Shermer, M. (1997). *Why People Believe Weird Things*. New York: Freeman.

Shermer, M. (2003). Show Me the Body. *Scientific American, 288*:27.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlation: Uses in assessing rater reliability. *Psychological Bulletin, 86*:420–428.

Smeets, T., Candel, I., & Merckelbach, H. (2004). Accuracy, completeness, and consistency of emotional memories. *American Journal of Psychology, 117*(4):595–609. doi:10.2307/4148994

Whyte, C. (1957). *More Than a Legend*. London: Hamish Hamilton.

Wilson, K., & French, C. C. (2004). Memory conformity and paranormal belief. *The Parapsychological Association Convention*, 469–471.

Witchell, N. (1979). *The Loch Ness Story*. London: Book Club.

Zar, J. H. (1996). *Biostatistical Analysis*. London: Prentice-Hall.