

***Authors' response:***

M. A. WOODLEY OF MENIE, M. A. SARRAF, J. E. KING, J. R. GREENWELL (†)

**The Anomaly Remains**

We would like to thank the commenters for their time and effort in replying to our article. We organize our response to the comments around four core themes: 1) the role of psychosocial and cultural factors in influencing eyewitness reports, 2) the reliability of eyewitness memory, 3) the quality of Green's dataset, and 4) the adequacy of the statistical approach employed in the target article.

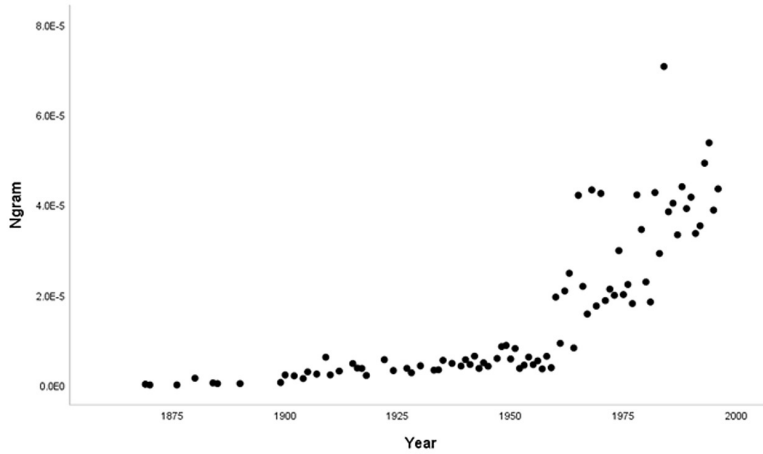
***The Role of Psychosocial and Cultural Factors in Influencing Eyewitness Reports***

Meurger (this issue) proposes a “group-induced perceptive model,” which seems to posit that cultural factors have a role in determining the content of individuals' experiences, or at least their interpretations of those experiences, and that this might explain why the sorts of “cryptids” reported over time change, why the cryptids reported at any given time tend to “have” certain characteristics, and so on. Magin (this issue) favors a model of this sort too. In our remarks on Paxton and Naish's (2019) paper concerning how cultural factors may have influenced patterns among reports about “sea serpents” in the 19th century (the key factor in that case being knowledge of Mesozoic marine reptiles), we note that there is a very valuable database through which temporal trends in culture can be tracked – specifically the Google Ngram Viewer (Michel et al., 2011), which samples words from millions of texts and tracks the utilization frequencies of those words over time. This database has been used extensively in what has come to be termed “culturomic” research, such as in the long-term tracking of human emotions, behavior, and cognition (see, Del Giudice, 2012; Figueredo et al., 2019; Michel et al., 2011; Woodley of Menie, Fernandes, Figueredo & Meisenberg, 2015). An obvious application of this database is to track the usage of words, or strings of words, bearing most directly on the topic at hand, e. g. “Bigfoot,” “Sasquatch,” etc., to ascertain whether temporal trends in these terms might explain changes in the frequency of sighting reports (for which we have data spanning the years 1869 to 1996). This will allow us to test Meurger's “group-induced perceptive model,” which seemingly predicts that culture drives “cryptid” experiences, reports, etc.

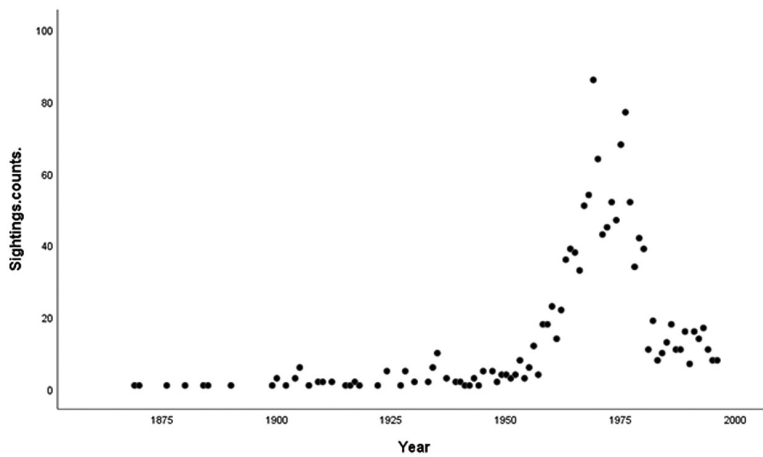
Two hypotheses can be considered. First, if cultural representations of Sasquatch are influencing reports (e. g. by providing background bias that might lead people to erroneously believe that they saw an anomalous hominoid when in fact they saw, e. g., a bear), then we might reasonably expect the two trends to be correlated. For instance, if there is a sudden uptick in Sasquatch reports in the 1960s and 1970s (which our data indicate), then proponents of culture-based models of “cryptid” encounters might reasonably expect signs of increased interest in “Sasquatch,” etc. to co-occur with such an uptick, and might treat this increased interest as the driver of the uptick – higher usage frequencies of “Sasquatch”-related terms in print media would be one sensible choice of an indicator of culture-level interest in “Sasquatch.” Second, for those who are more “literalistic,” to use Meurger’s terminology, in their approach to cryptozoology, it might be expected that sightings are influencing cultural representations, in which case we might predict a lagged relationship between the two, e. g. sightings clusters from the 1960s might contribute to an uptick in interest in Sasquatch a decade later, in the 1970s, etc. If hominoid sightings precede the relevant cultural trends, this would count against the hypothesis that culture is the crucial factor behind sighting reports.

Using Google Ngram viewer (specifically using the English 2012 corpus) we searched the database using the following search string: Sasquatch+sasquatch+Bigfoot+bigfoot+BigFoot+ape man+apeman. We generated a plot of Ngrams from the years 1869 to 1996 – sampling data that matched the years for which we had overlapping Sasquatch sightings. The use of mixed terms allowed us to avoid capturing trends reflecting shifts in utilization frequencies of words due only to changes in the popularity of the words to refer to anomalous hominoids. The term “Sasquatch” was first used in 1929. The term “bigfoot” was first used in relation to the Sasquatch in the 1950s. The term “ape man,” which seems to have been used historically as an occasional synonym for Sasquatch, had its first recorded usage in the 1860s. Another term that is sometimes used as a synonym for Sasquatch is “wood ape”; but there are no instances of this term recorded in the Ngram viewer, indicating that its utilization frequency is too low to have had an effect on the overall trend. It should be noted that Ngram viewer outputs are per capita estimates, weighted against the total volume of literature being generated at any given point. This avoids inflation of counts due to the mere fact that there are more written texts being generated over time. The plot of the Ngram search string is presented in Figure 1.

Figure 2 presents the temporal distribution of Sasquatch sightings from the Green database.



**Fig. 1:** The utilization frequencies of an Ngram search string (Sasquatch+sasquatch+Bigfoot+bigfoot+BigFoot+ape man+apeman) from the years 1869 to 1996. N=86 years for which Ngrams overlap hominoid sightings.



**Fig. 2:** The sightings frequencies (i. e. counts per year) from the Green database. N=86 years.

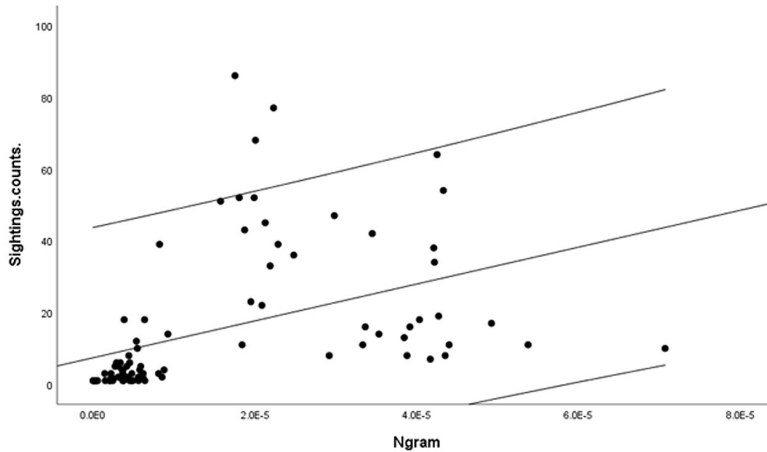
Table 1 presents the results of testing the two hypotheses, i. e. 1) change in sighting frequencies = change in Ngrams, and 2) change in sighting frequencies → change in Ngrams. This was achieved by conducting two temporal bivariate regressions. The first regression used the data points aligned in time, whereas in the second the Ngrams were forward lagged by ten years relative to the sighting counts, so that sightings in e. g. 1960 were used to predict Ngrams in 1970, etc. Poisson regression was used as the independent variable was count data.

Hypothesis	Variables	$\beta$ (SE)	Z	$pr(> z )$	AIC
Sightings = Ngrams	Intercept	2.70 (.03)	95.06	$2 \times 10^{-16}$	1970.7
	Ln Ngram	.27 (.02)	12.44	$2 \times 10^{-16}$	
Sightings → Ngrams	Intercept	2.63 (.03)	87.29	$2 \times 10^{-16}$	1763.9
	Ln Ngram (forward lagged)	.44 (.02)	19.33	$2 \times 10^{-16}$	

**Table 1:** The results of a Poisson regression testing two hypotheses: 1) that the change in Sasquatch sightings tracks the change in Ngrams, and 2) that the change in sightings precedes the change in Ngrams. Model degrees of freedom = 85 in both cases.

Both models produce statistically significant results (i.e. the natural logarithm of Ngrams positively predicts the change in sighting counts over time). The  $\beta$  coefficient for the second model, i. e. the one in which the relationship is lagged by ten years, is larger in magnitude than for the first model ( $\beta = 0.44$  vs.  $0.27$ ). Furthermore, the Akaike Information Criterion (AIC) value is smaller for the second model, meaning that this model is more parsimonious and is therefore a better fit to the data. Alternative model specifications in which the natural logarithm of time was used as an additional independent variable in order to capture unmodeled temporal covariances were also run – but these suffered from unacceptably high levels of multicollinearity (specifically the Variance Inflation Factors were substantially higher than 10; Kutner, Nachtsheim, Neter & Li, 2005). The best fitting model is graphed in figure 3.

Based on these results it can be suggested that the “group-induced perceptive model” is not well supported, and that a stronger relationship exists when the sighting count data are used to predict cultural trends in a forward-lagged manner (i. e. when sightings predict the relevant cultural trends 10 years hence). That the “group-induced perceptive model” seems to fail is consistent with the various lines of evidence that we reviewed in the target article indicating that human perception is not all that malleable, such that it is unlikely that cultural factors are seriously biasing interpretations of perceptions that are the basis of hominoid sighting reports; it seems more likely that the anomaly gives rise to the cultural representation at issue. We do not of course rule out some role for the sort of psychological, cultural, etc. phenomena that Meurger’s “group-induced perceptive model” posits. As Jussim (2012) notes, expectation biases can influence relevant mental processes – but apparently typically to a very small degree.



**Fig. 3:** The association between Ngrams (forward lagged by 10 years) and Sasquatch sighting counts for 86 overlapping years. Line of best fit along with 95% confidence intervals are presented.

There is then the possibility, along the lines proposed by Meurger (2006) in the case of the remains of cave bears being mistaken for the remains of dragons, that the people represented in our database are observing something mundane, but are misinterpreting these experiences as encounters with an anomalous hominoid (rather than fabricating them wholesale, or hallucinating them). It has been suggested for example that black bears might be behind a great many hominoid sightings on the basis of apparent ecological niche overlap, coupled with some superficial morphological similarities (Lozier, Aniello & Hickerson, 2009). But the patterns in the eyewitness data do not support this mistaken identity thesis (consider the presence of ape-like features, the total absence of a tail, the use of arms for manipulating objects, the presence of tracks that evidently are not ursine in origin, etc.). Therefore, based on the results of our culturomic analysis, and all of the previous analyses presented in the target article, we believe that the case for the view that “Sasquatch” represents a genuine anomaly, rather than a product of group-induced perceptual misinterpretation, is strengthened, while remaining very far from definitive.

Finally, Meurger’s observation that eyewitness observations of “dragons” and “mermen” appear to have apparently been quite common in the past, ties in with a point raised by Magin (this issue), specifically: “If you also take into account that Bigfoot is actually seen and allegedly filmed in all regions of Europe, even though the zoological probability of its existence in

Scotland or Ireland, not to mention Germany, is depressingly low, the verdict that eyewitnesses are right about what they allegedly saw is even less credible in itself . . ." (p. 55).

On the matter of this low probability, we agree with Magin. But the question becomes whether all substantive anomalous phenomena under the umbrella of "cryptozoology" are zoological in the ordinary sense, or whether there are "paranormal" possibilities as well. One Fortean researcher, Jerome Clark, has speculated that there might be at least two broad kinds of anomalous occurrences: "event anomalies" and "experience anomalies" (Clark, 2018). The former are thought to be objectively real as that term is normally understood. The latter are thought to have a less certain ontological status – they could be something akin to "visionary experiences" due to an unknown, perhaps paranormal, cause. Crucially, Clark (2018) suggests that there may be a relationship between event and experience anomalies, with the latter in some way parasitic upon the former by adopting aspects of the phenomenology of the former in particular instances. On Clark's model, then, anomalous hominoid sightings in improbable places might constitute such "parasitic" experience anomalies.<sup>1</sup> It is these "parasitic" experi-

---

1 Whether the changing image of hominoid or hominoid-like creatures described by Shoalts (2019) – whose work Meurger (this issue) mentions – might be explained in terms of Clark's (2018) ideas is worth considering, but we do not pursue this possibility in detail. On the other hand, one could argue that Meurger might be overstating the power that Shoalts' (2019) findings have to undermine cryptozoological "literalism." A couple of passages from Shoalts are worth reproducing here:

These early accounts written by explorers, travel writers, and missionaries in western North America, all of which describe "mountain monster" legends that can loosely be styled "proto-sasquatch" tales and which are supported by third-party Indigenous sources, seem to allude to at least three different types of creature. There are the legends of mountain-dwelling anthropomorphic giants noted by Walker, Kane, Macfie and Dawson, the fearsome, fur-covered, long-clawed and mysterious "matlox" recorded by Mociño, which is similar to the beast featured in Roosevelt's Bauman story, and finally the vaguely described, massive creatures noted by Thompson and Cox that apparently left grizzly-like tracks. (Pike's account is too vague to classify, but seems most similar to Thompson's and Cox's.) On their own, none of them quite match the later, familiar image of the "bigfoot" or sasquatch. Yet, when mingled together, one gets a pretty close approximation of a modern sasquatch: a seven to eight foot tall hairy bipedal creature that lives in the mountains and makes giant, grizzly-like prints. Such "mingling," a staple of folklore, must have taken place at countless shared campfires in the Rockies and Pacific Northwest over generations of frontier encounters between Indigenous guides, European explorers, trappers, and other travellers. (2019: 81)

[M]any of the low-status, often illiterate voyageurs and white hunters or trappers (i.e. "mountain men") who lived on the frontier, retained beliefs about monsters closer to their Indigenous companions than to their social and economic elites. . . . In the post-1860s . . . new waves of settlement moved into the North American West . . . . These new settlers also introduced explicitly ape or gorilla-like traits into existing legends, many of which had featured scary anthropomorphic giants but not necessarily (at least in all cases) very ape-like. (2019: 92)

It does not seem to us that any of this excludes the possibility that encounters with an anomalous

ences that are likely more significant in the case of these historical reports, since what they concern has a highly uncertain ontological status. People could be reporting something akin to “visionary experiences” due to some unknown paranormal cause. Also relevant to this may be the *decline effect*, or the tendency for effect sizes of one sort or another to decline over time. There is evidence that the causes of decline effects in parapsychology are not merely methodological – and in the case of dramatic paranormal effects, very pronounced decline effects have been suggested. Ian Stevenson (1990) has gone so far as to conjecture that there may have been a long-term decline in the prevalence of major paranormal phenomena, at least since the 19th century. One Fortean researcher, Paul Devereux (2007), notes a number of apparently sober anthropological field observations of profoundly unusual paranormal phenomena in non-Western populations. Perhaps “high-strangeness” *zooforn* phenomena fell victim to the shift in the “reality set” of Western peoples that accompanied the rise of secularism and a broadly

---

hominoid formed the basis of the legends and folklore on which Shoalts focuses. Every reasonable person can agree that legend and folkloric traditions, if based on some actual experience(s), generally should not be expected to represent that experience(s) with high accuracy, but will instead typically have substantial accretions and embellishments, with the resulting narratives usually or at least often being rather fantastical. The fact, which Shoalts implies, that some stories of rather ape-like creatures could be found in North America even before the post-1860s settlement waves, and that from a sort of composite of “proto-sasquatch” narratives “one gets a pretty close approximation of a modern sasquatch,” seems to leave open the possibility that encounters with an anomalous hominoid were parlayed into various “monster” tales that make up those “proto-sasquatch” narratives. After all, as King et al. (2020) observe, “the two gorilla species . . . were ethnoknown for several years prior to formal description, having even taken on something of a mythical status among naturalists at the time” (p. 37). Wylie (1980) observes that “[f]or a long time gorillas were [erroneously] assumed to be violent and dangerous to man . . . ancient traditional lore that [gorillas] enjoyed carrying human females off for sexual purposes [had to be thrown out]” (p. 170). Why could distortions of that kind, and even more dramatic ones, not apply in the case of a currently unconfirmed hominoid? In considering these possibilities, one should keep in mind the variety of distinctions between genuine legend and folkloric narratives and serious eyewitness accounts.

The folklorist Thomas Bullard (2010), discussing the work of David Hufford (1982), notes that there is evidence that belief in anomalous phenomena does not merely originate from narrative traditions, but rather that it exists and persists in part because of people’s encounters with the anomalous:

Hufford generalized supernatural assault accounts under the term of “Mara attack” and recognized that the experience, with its consistency of phenomenology and independence from tradition, posed a head-on challenge to the prevailing academic consensus on anomalistic encounters. The evidence made a compelling case that tradition does not always give rise to experience; rather experience sometimes returns the favor and gives rise to tradition. For Hufford, “a fundamental reason that spiritual beliefs [note from authors: “spiritual beliefs” could here be “belief in anomalous phenomena”] have been able to resist the enormous social pressures toward secularization [note from authors: “secularization” could here be “skepticism”] is that they are, in part, rationally founded on experience (that is, empirically grounded).” (p. 301)

“scientific” worldview? Needless to say, this interpretation is extremely controversial, and is not considered plausible by all authors on this response piece.

An obvious objection to this sort of speculative theorizing is that it is unparsimonious and that applying “Occam’s razor” should lead us to reject such ideas out of hand. Unfortunately, parsimony is a frequently misunderstood concept. The principle of parsimony, properly understood, only indicates that among explanations that are equal on *all other* criteria of good explanations, the simplest of those explanations is to be preferred. If, for example, explanations are unequal with respect to explanatory power, considerations of simplicity are not necessarily decisive. Further, if all explanations available have serious deficiencies, one need not and in fact *should not* accept any of them – one should rather realize that an *anomaly* remains to be explained, that *more research* is needed. That is precisely the stance that we took in our target article.

### ***Reliability of Eyewitness Memory***

The comments reveal a range of perspectives on the evidential value of eyewitness memory, a subject that we addressed at length in our target article. Meldrum (this issue) sides with us on the question of eyewitness reliability, adding that claims about the inherent unreliability of eyewitness testimony are often misused by skeptics to *a priori* exclude eyewitness reports from scientific consideration. Bauer (this issue) suggests that the value of eyewitness testimony is potentially highly variable, and that a small number of high-credibility witnesses are to be preferred to a large number of less credible ones. There are certainly individual differences in eyewitness reliability, just as there are individual differences in seemingly all human traits; it should be noted that aggregates of data can give quite reliable estimates, given the tendency, assuming a normal distribution of reporting error, for error at the extremes to simply cancel out. Thus we maintain that consideration of large numbers of eyewitness reports may have considerable scientific utility.

On the topic of eyewitness memory reliability, Magin (this issue) clearly strongly disagrees with our position. We contend that he seems to misunderstand the relevant research items that he does discuss (he does not mention most of the germane research cited in our target article), which he (incorrectly) maintains we misrepresent – the papers of Wixted (2018) and of Wixted, Mickes, and Fisher (2018). We are gratified to note that a paper by leading psychologists studying memory reliability, recently published in the prestigious journal *Current Directions in Psychological Science*, which assesses the state of the evidence on human memory reliability and which came to our attention only after our piece was submitted – adduces the Wixted et al. (2018) paper (and the Koriat, Goldsmith, and Pansky article, also cited in our target article) for the same purpose we do (they cite those papers in support of the following): “Memory is clearly



malleable *but not unreliable under normal circumstances* and in the absence of contamination or prolonged suggestion by psychologists, therapists, or anybody else” (Brewin, Andrews & Mickes, 2020: 124; emphasis added). The authors also make apparent their view that “memory [is] malleable but essentially reliable” (Brewin et al., 2020: 121). In claiming that we misunderstood Wixted et al. (2018), Magin’s objections would surely apply equally to experts on memory reliability whose understanding of Wixted et al. (2018) parallels our own – does he wish to argue that they are also mistaken?

Note that we *did not* maintain that eyewitness memory is infallible, perfectly reliable, or non-malleable. Indeed, several statements that appear in our target article make that clear (see p. 37–38 and p. 43 of King et al., this issue), but appear to have been overlooked in Magin’s reply. We also present reasons that Wixted et al. (2018) offer for thinking that eyewitness memory is not so frequently mishandled in the “real world” to render it of no evidential value (see p. 38 of King et al., this issue), against what skeptics of eyewitness memory (in the police investigation context and elsewhere) have repeatedly argued.

A key issue concerns the use of the term “inherent.” In response to our (accurate) use of Wixted (2018) and Wixted et al. (2018) in support of our argument against the blanket claim that “eyewitness memory is *highly* and *inherently* unreliable” (emphasis added), Magin observes that the Wixted research supports the idea that eyewitness memory can be *compromised* by “test[ing] the memory of the eyewitness *again and again*” (Magin, this issue: 52, citing Wixted, 2018: e14). Clearly, we are now well beyond consideration of the “inherent” reliability of human memory and are discussing what happens when investigators *interfere with* eyewitness memory by repeated probing, a phenomenon that we mentioned in our target article (see King et al., this issue: 38).

Magin would appear to be mistaken in his interpretation of Wixted et al.’s research when he writes that “[m]emories only attain a high degree of reliability when certain protocols are closely followed which have *just recently been established*” (this issue: 52; emphasis added). The implication here is that high eyewitness reliability is *rare*, and that police have only recently attempted to establish the conditions needed to obtain it. This is contrary to what Wixted et al. (2018: 331) actually write, which makes plain that the *strong tendency* has been for police to receive accurate information from eyewitnesses,<sup>2</sup> with improvements of interviewing tech-

---

2 “In all, laboratory studies of eyewitness memory that use generally accepted interviewing protocols and do not intentionally provide misleading information or entice witnesses to guess find that accuracy is quite high (~85% – 90%). Field studies of police interviews with victims and witnesses of real crime show, if anything, even higher rates of accuracy” (Wixted et al., 2018: 331). This indicates that eyewitness data gathered through interviews tend to be highly accurate, even if perhaps less accurate than such data gathered by having eyewitnesses freely recall observations.

niques expected to increase the *amount* of information received and not the *accuracy* of the information received, “because accuracy is already high” (Wixted et al., 2018: 331).

Part of the basis for Wixted et al.'s (2018) claim about high memory accuracy on the part of witnesses to and victims of crime is in research that is decades old (see Wixted et al., 2018: 330–331), which is inconsistent with Magin's statements concerning the need to follow protocols that have “just recently been established” for recalled information from witnesses to be trustworthy. Wixted et al. (2018) include an example of a poor investigation leading to extraordinarily bad eyewitness performance – and yet, even here, the *majority* of reported information is still accurate: “In this case, only 58% of the reported attributes were correct, as corroborated by CCTV” (p. 330).

Magin (this issue) claims that the view that “eyewitness memory is highly and inherently unreliable” “was strong, and it remains strong” (seemingly he means to say that the evidence for the view is strong, which is different from the view itself being strong), and he makes this claim in the light of work from Wixted and colleagues. And yet Wixted et al. (2018) explicitly contradict that claim. If what we have already provided is not yet sufficient, consider this further statement from Wixted et al. (2018): “Despite [...] positive developments, we submit that the once surprising revelation about the malleability of eyewitness memory has led to a severe overcorrection such that the field now regards eyewitness memory not only as potentially unreliable but also as *inherently* unreliable. *In our view, the evidence does not support this idea and instead clearly refutes it*” (p. 326; emphasis added).

Magin is critical of “free-form” reports, yet we have noted the voluminous evidence offered that *freely recalled* memories tend to be highly accurate or even very highly accurate, a point that we stressed (see King et al., 2020: 29) (for evidence, see Diamond, Armson & Levine, 2019; Koriat et al., 2000). One reason that this might be the case is that when eyewitnesses are simply prompted to freely recall some event, there is little to no opportunity for investigators to contaminate the memories of the former with false information, opportunities that are more abundant in interviews where investigators ask eyewitnesses many questions. Bear in mind here that, to repeat, the evidence *overwhelmingly indicates*, as made manifest in our original piece, that the threat to eyewitness reliability primarily stems from *incompetent investigations*, not some “inherent” unreliability of eyewitness memory. Wixted et al. (2018) are very clearly in agreement here: “eyewitness memory has been wrongfully convicted of mistakes that are better construed as having been committed by other actors in the legal system, *not by the eyewitnesses themselves*” (p. 324; emphasis added). Again, it seems to us that much turns on Magin's apparent misconception of the word “inherent,” leading him to challenge our correct claims that the relevant evidence does not support the unqualified assertion that eyewitness memory is inherently and highly unreliable.

Magin goes on to raise another misconceived point: “talking about memory and recall does not even take into account how enormously people can misperceive what they see” (this issue: 54).<sup>3</sup> In the context of controlled and properly conducted experiments testing memory reliability, for eyewitness *memory* of perceived external events to be accurate, it of course must be true (setting aside odd “thought experiment” scenarios) that the *perceptions* recalled were accurate in the first place. Once more, the relevant literature shows clearly that *freely recalled* memories of observations that could be *objectively verified* (and therefore involved sense perception on the part of the witnesses to form the memories), tend to be highly accurate or very highly accurate (Diamond et al., 2019; Koriat et al., 2000), that eyewitness memories probed through interviewing tend to be highly accurate (Wixted et al., 2018), and that findings evidencing the inaccuracy and malleability of memory have been exaggerated with the effect of leading researchers and lay people mistakenly believe that memory is highly and inherently unreliable (Brewin et al., 2020; Diamond et al., 2019; Koriat et al., 2000). So Magin’s critical attempt to artificially distinguish perception and memory in this context seems to stem from confusion.

Magin discusses purported instances of poor eyewitness performance – specifically involving inconsistency among eyewitness reports on the same event and alleged inaccurate interpretations of observations on the part of eyewitnesses, culled from newspapers and academic grey literature (e.g. the non-peer-reviewed *Anomaly – Journal of Research into the Paranormal*) – in an effort to cast doubt on the evidential value of eyewitness testimony. The reader should understand that one could just as easily go about critiquing all of mainstream science with this approach, choosing to report only those cases in which experiments and other tests of hypotheses failed or failed to replicate. Moreover, one could just as easily (we would suggest even more easily) pick through the literature to find many cases of multiply attested anomalies where the reports of eyewitnesses to the same event are highly convergent (see, e.g., Nahm, 2019 for one such case).

We end this section by noting that not all of the data that we used in our analysis were extracted from sightings freely recalled by eyewitnesses. We can say that at minimum, ~13% of sightings were freely recalled, given that they were sent in the form of letters. Moreover, the

---

3 In anomalistics debates, the term “misperception” is almost always a misnomer:

“Misperception” is a debunker-minted word that resists rational definition. A person cannot possibly “perceive” something “wrongly” or “misperceive” something unless he or she has a rare eye or brain defect such as color blindness, which can be calibrated and compensated for. Either witnesses perceive something, or they don’t. They can be mistaken in their interpretation of their sensory perceptions, but perceptions by their senses are themselves incapable of cognition or thought and cannot make mistakes. The incident light rays either entered the eyes or did not do so; there is no fuzzy in-between here. Human observers may make a *misjudgment* or misinterpretation of what they *accurately* perceived, but as (typically) non-Ph. Ds, they cannot scientifically investigate their own observations . . . . (Sparks, 2018: 380)

fraction of freely recalled sightings could be quite large (~56% or even more), but without a doubt, data for some of the sightings were extracted via interviews. While interviews may typically yield eyewitness testimony that is somewhat less accurate than free recall of eyewitness memories, the preponderance of data indicates that interviews of eyewitnesses tend to yield highly accurate information, even when ideal procedures are not followed (Wixted et al., 2018: 331).

### *Quality of Green's data*

Three of the commenters (Magin, Meurger, and Paxton) suggest that Green may have been selective in his collection of data, essentially building the apparent internal consistencies that we observe into the dataset. Meurger (this issue) attempts to substantiate this thesis with reference to the work of Wylie (1980). We have taken the time to consult this work for ourselves. Wylie (1980) suggests, in a footnote on page 27 of his book, that his basis for concluding that Green was so biased is in his interview with Green, provided in summary form in the sixth chapter of his book – the following is the footnoted sentence: “Green’s analysis of the data that he so carefully presents is arbitrary, apparently based on his personal choice as to what is good or bad evidence” (p. 27).

Having read through this interview summary, we note that it offers no actual evidence of any such bias on Green’s part. If anything, we got the distinct impression that Wylie went into his interview of Green with prior commitment to the idea that Green was biased in his selection of data, insisting on this conclusion even when nothing in his interview of Green warranted it (or at least nothing that Wylie presents from that interview). In fact, some statements on Green’s part suggest that he took the approach of including anything “Sasquatch” related that he could find in his database, that there was a relative *lack* of selectivity in his data choice – for example:

I’m not claiming to have any opinion as to the validity of most of the material that I’ve accumulated,” Green says. “However, I’m assuming two things which go to two sides of the thing. One is that there’s a lot more either outright fakery or wishful thinking today than there was when the subject was virtually unknown – anytime up to the Patterson movie. In the early years of the investigation it was almost certain that if you got a lead, there would be something solid to it, because if there weren’t, there wouldn’t be a lead. Now it’s reached the point where, with people like Rene and myself, you have to say to yourself once in a while, ‘I damned well better go ... check on this thing.’ But you’ve been on so many wild-goose chases that you don’t have the same enthusiasm anymore. (Wylie, 1980: 166)

At one point, Wylie (1980) evidently accused Green outright of selecting data to fit a favored hypothesis – but nothing in Green’s response actually suggests that he in fact did any such thing, and there is nothing preceding or following Green’s response to suggest it either:

Admiring Green's logic, if not his premise, I inquire about the supposed internal consistency, the alleged coherence within the hundreds of reports, that has led him to his conclusions. I point out that if one does not arbitrarily select the data to eliminate reports that do not support a consistent description of the Sasquatch, there is much less internal coherence than most Bigfoot experts like to admit.

"I'm not saying that every speculation of mine will be proved correct. Not by any means," Green says. (ibid.: 168)

Near the end of the section of his book that presents the Green interview, Wylie (1980) writes the following: "No reason is provided regarding the selection of 'acceptable' evidence beyond the rather lame explanation that it 'makes sense.' Green uses a biased sample, relying on his definition of common sense" (pp. 169–170). Yet where Wylie (1980) offers direct excerpts from Green's half of the interview, nowhere does Green even use the word "acceptable" or the phrase "makes sense" (or anything with "sense" in it). Are these supposed to be interpretations of the thrust of Green's relevant statements, or did Green say "acceptable" and "makes sense" at some points? If he did, and Wylie had a damning admission from Green about his alleged biased data selection, or even statements strongly suggesting it, why did Wylie not include that in full in his book?<sup>4</sup>

Our impression of Wylie (1980) is that he seems to misleadingly frame Green's remarks so that he (Wylie) could have spurious evidence consistent with *his* favored explanation of Green's data.<sup>5</sup> More generally, this illustrates why it is that great caution ought to be exercised when

- 
- 4 In support of Wylie's (1980) view, Meurger claims that Green was "unfaithful to his own source material" because the latter maintained that "Sasquatch" was basically nocturnal, whereas our analysis of Green's data has indicated that the putative hominoid is diurnal. But, as far as we can tell, this does *not* support Wylie's critique of Green. Wylie did not claim that Green was *misrepresenting* his data – he argued that Green assembled a dataset in such a way that it conformed to his presupposed notions about the nature of "Sasquatch." That Green was apparently wrong on the nocturnal/diurnal point could be due to nothing more than an innocent mistake on his part, e. g. poor or incomplete analysis of his data. It does not in any way demonstrate that he was engaged in unethical questionable research practices.
  - 5 Wylie (1980) suggests, without any clear supporting evidence, that Green excluded data inconsistent with the idea that "Sasquatch" is pacific. But this is not at all apparent from the Green database that we analyzed – it *does* include reports of alleged hominoids acting aggressively, indeed one of our items dealt explicitly with the matter of aggressive behavior on the part of the hominoid. Wylie (1980) also argues that mythic tales about violent gorillas, prior to true scientific discovery of gorillas, should undercut confidence in eyewitness accounts of anomalous hominoids. But this is a weak argument. First, this sword cuts both ways – skeptics of the time insisted gorilla accounts were *nothing but* myth, and obviously they have been shown to be fantastically wrong. Second, and relatedly, no matter how mythically distorted, those gorilla reports were shown to accurately correspond in many respects to an actual undiscovered animal despite their inaccuracies. Third, gorillas *sometimes are* frighteningly

making claims about possible scientific misconduct, or questionable research practices, in the absence of sufficient evidence. Science is fundamentally an exercise in trust and can only flourish when that condition is sustained. There is a history of high-profile allegations of questionable research practices against specific researchers that spectacularly backfired, often revealing instead what were clearly biases on the part of those alleging the misconduct (see, e.g. Lewis et al., 2011).

We acknowledge that Green's data collection *could* have been biased, and indeed it would be irresponsible for us to suggest that we have excluded the possibility of bias; but we simply have nothing in the way of sufficient evidence for the view that Green's data collection was in fact biased in some problematical way. Paxton (this issue) argues that we are in a position of having to "demonstrate that the reports considered, are an unbiased sample of whatever unstated statistical population the authors are seeking to make conclusions about," following up with the question of how we are to know that Green or another relevant researcher was not biased. The trouble, of course, is that this objection could be put to an enormous amount of mainstream retrospective scientific research. Even if the authors of a paper clearly specify their search procedure, inclusion criteria, and so on, other researchers often can only trust that the data collection was done as stated and that the authors are not outright lying, especially when data searches are hard to replicate. The simple and central fact of the situation is that the possibility of bias in Green's data collection cannot be excluded unless another systematic attempt to collect anomalous hominoid reports from the same time periods and regions is conducted. Even then, great care would have to be taken to ensure that whatever differences might emerge between the database resulting from such an effort to "replicate" Green's data collection and Green's actual database were truly indicative of bias on Green's part, rather than merely suggestive of innocent mistakes or legitimate differences in how the data gathering was approached in each study.

### *Statistical Considerations*

Bauer and Paxton both comment on our use of statistics. Bauer (this issue) suggests that we are excessively precise in terms of reporting (certain) results to multiple decimal places. This "excessive" precision simply stems from an attempt to report what was found. These numbers could just as easily be rounded down to the nearest whole if "excessive precision" is an impediment to taking the work seriously – whereas for those who, for whatever reason, want more precision, ascertaining the more precise values would not be possible had we reported the values to, say, the nearest whole number.

---

violent, meaning that even the mythic tales of violent gorillas may well have originated from accurate observations that were distorted over time.

Paxton (this issue) makes three major points regarding our “idiosyncratic” use of statistics. These are dealt with in turn.

The first concerns the context in which the Green data used in the modal analysis of eyewitness reports are treated. Paxton notes that “[i]t is important to identify the population under consideration when undertaking a statistical analysis and here it is essential. Are the authors making conclusions about bigfoots, bigfoot reports, John Green’s reports or a subset of John Green’s reports?” This distinction is argued to be relevant on the basis that different conclusions potentially obtain when different assumptions are made about the nature of what is being analyzed. Critically, for these data to be germane to the matter of eyewitness consistency in reports of a putative anomalous North American hominoid, it would be necessary to show that “there are no systematic biases in their data analyses or collection that would frustrate their conclusions. Or more realistically/practically, justify an assumption that there are probably no such biases.”

Paxton goes on to state that “[t]wo thirds of the data has been removed using vague criteria. It is hardly surprising that the remainder is consistent. The authors now say that the rejected reports were not coded but an earlier draft stated that the rejected reports were ‘irrelevant and indirect’ implying there was content based criteria for rejecting the reports.” The reason for this change was simple. We were in error when we claimed that sightings were rejected on the basis of being irrelevant and indirect; indeed, it was thought that this was the case for a number of such excluded sightings. A more thorough examination of the original coding and selection protocol, which was occasioned by Paxton in his capacity as a peer reviewer on our target article, allowed us to offer a much more accurate description of which sightings were excluded and on what basis. In the final version of the target article, we state that the fully coded Green dataset includes a total of 1,386 reports for which data on at least one relevant characteristic could be obtained, and that only a very small number of sightings were excluded (one, although a second report may have also been excluded) – specifically on the basis of absurdity (the excluded report alleged that a hominoid was walking around Hollywood, California). Doubtlessly, retaining what was excluded would have made no difference to the results. At any rate, this is very far from the impression of haphazard or unsystematic handling of the data that one would get from reading Paxton’s comments, although we freely acknowledge the inadequacy of our initial claim concerning exclusion protocols, and are grateful to Paxton for encouraging us to investigate this matter in more detail.

A more substantive concern is in the fact that these data do form part of a much larger dataset, which contains many more sighting reports, a point mentioned in our target article. Specifically, we state that “[t]he online database of reports compiled by Short (Short & Green, n. d.) contains a larger number of sightings ( $N=4,724$ ), covering a wider range of years (up to



2001). These additional reports could not be employed in the current analysis because they have not been coded" (King et al., 2020: 9). The data employed in our target article's analysis extends to 1996, and were electronically coded by John Green's daughter. Coding of the dataset presumably would have begun in or near the year 1996, hence subsequent sightings were excluded. Bobbie Short did not start her electronic collation of Green's full "Sasquatch" sightings database until 2001, which is when the earliest version of her website *Bigfoot Encounters* appears on the Wayback Machine, and it would appear that it was her own data that greatly inflated the numbers of sightings. As a historical note, this project was originally intended for publication as a book to be entitled: *Observing the Sasquatch: Statistical results from an analysis of 1388 bigfoot sighting reports*. Work on this came to a stop in 2005 when JRG died, and the analyses "laid fallow" until 2019, when JK informed MAW of the existence of these data and of his desire to work them up into a publication (JK and MAW are both connected through the evolutionary psychology unit at the University of Arizona.) MAW re-conducted the comparison of modes using the data files and "cuts" that had originally been employed by JK and JRG in their "first pass" at the data.

It is entirely possible that additional relevant sightings (falling between the years covered in the target article analysis) were omitted by Green, but we would contend that this would have been due to practical issues related to working with (mostly) paper records, rather than due to some selectivity on the part of Green (or his daughter). We also note the following: "In any case, whether this subsample is representative of the full sample can be directly tested in future research by comparing the results of our analysis to those obtained from the use of the full dataset, once this is coded" (King et al., this issue: 20). Thus, we are clear about the possibility that our subsample might not be representative of the whole and reframe this as an empirical matter (specifically that coding the full dataset would allow for meaningful comparisons to be made). Moreover, we acknowledge that the empirical picture might evolve further when the full dataset is taken into consideration. Indeed, this possibility certainly cannot be excluded at this stage. When the target article analysis is compared with the earlier work of Beeson (1979, 1988), who employed a much smaller subsample of the same (Green) data (about a couple of hundred reports), we note certain convergences with our own analysis, such as the high percentage of singleton reports, but also some discrepancies such as "inordinate nocturnality"; so the patterns identified in our target article may change upon analysis of the larger dataset.

The second matter raised by Paxton (this issue) concerns the "cutting" of the eyewitness data for the purposes of establishing the modal classes in our comparison analysis. More precisely, Paxton (this issue) notes that "[t]he 'mode test' employed here ... is rather a test of whether there are more reports from day (from the unknown population of interest) compared to night (from the unknown population of interest). So the authors' odd methodological approach blinds them to testing the most interesting hypotheses about their data" (p. 65). Paxton suggests instead that "the obvious thing to do with regard to time of day of reports for example, is to



compare the counts of bigfoot (reports/sightings) to the expected hours of daylight (for a given latitude, at a particular time of year) as surely the value of an analysis here is to find out if the reports are *disproportionately* nocturnal/crepuscular/diurnal” (ibid.). This does not seem to be a logical approach to analyzing these data, since concepts such as “night” would seem to have a meaning that is conserved with respect to location (i. e. it is the period in a day characterized by the absence of sunlight), thus there is no obvious reason to think that our modal comparisons need to account for specific sample or population characteristics for these basic classificatory variables to generalize. The same argument could be made for the object of the descriptions, i. e. the putative hominoid. Indeed, the very basis of our analysis involved attempting to determine whether there was in fact clustering around certain observations, which means that agnosticism concerning the presence of these clusters, as reflected in the assumption of a null hypothesis characterized by no modal differences, was well justified.

The final issue raised by Paxton concerns the cumulative description curve analysis, and the reproducibility and interpretability of the results of this analysis. Paxton notes that we are not explicit with respect to how we handle the intervals between years in which descriptions are made. We concede that this is a shortcoming of our methods section. But, to be clear, we *did not* code a year unless a description event was made. This approach was used for two reasons. First, this is more conservative as it reduces the potential for temporal autocorrelation, which necessarily increases with density of sampling in time. Therefore, by spacing out the data in time this confound can be reduced. Second, the relevant event for the purposes of analysis is surely the instance in which a new description is made. The years in which the count number remains unchanged are not really relevant to the analysis. Moreover, it would seem to be this difference between Paxton’s attempted replication of our curve (in which he adds data for “missing” years) and our own (in which we only code for years in which there is a change in the number of taxa described) that accounts for the (slight) differences in the model outputs. It should also be noted that we obtained the quadratic form functions from Excel, and that there is a possibility of slight variation in outputs when these parameters are estimated using different software packages (it is known well that this happens, certainly in our native discipline of psychology and definitely among statisticians).

Paxton mentions that he is able to get an asymptotic curve to fit the data. But this is to miss the point of using a quadratic function. The main advantage of a quadratic function is that it has an inflection point, which can potentially be utilized in making predictions about the point in time when a given description record might become saturated. Moreover, in generating our estimates of the numbers of new species awaiting description we did not focus on the confidence intervals (CIs) for a specific point in time (i. e. when the main trend-line inflects, as Paxton does), as these values are not so relevant. What matters more is when the upper and lower 95% CIs inflect independently – to ascertain this, we plotted the predicted values for these out

separately so as to establish the different inflection points. This not only allowed us to determine whether new species might be added to the description record, given the assumption that the primary trend merely adheres to the *growth parameters* associated with either the upper or lower CI limits (but obviously not the predicted numbers, which at any given point in time will be higher or lower than the real number); it also gave us the ability to use the differences in the inflection points to generate a forecasting range of years – something that, to our knowledge, has been lacking in cumulative species description curve analyses thus far. It could be that there are better approaches to take in getting at these parameters, but it was with all of this in mind that we set about analyzing these trend data.

### ***Conclusions***

We conclude with the observation that the purpose of the target article seems to have been misunderstood by some though certainly not all of the commenters. Paxton (this issue) even goes so far as to suggest that we give the impression of engaging in “advocacy.” The target article argues at *no point* that any anomalous hominoid actually exists, a point that only Krall (this issue) seems to have explicitly mentioned in his commentary. We are and were fully aware that the evidence that we present is not close to sufficient to support any such claim.

Moreover, we agree with Meldrum (this issue) that the value of eyewitness data is certainly limited, and such data should not be treated uncritically. In our target article, we affirmed these points, and also made clear that our data do *not* demonstrate that there is any anomalous hominoid in North America. Rather, our data, and the considerations from contemporary scientific work on eyewitness memory brought to bear on them, led us to the cautiously phrased conclusion that there is “a basis for recommending that greater effort be expended to definitively investigate the possibility of an unconfirmed primate inhabiting certain woodlands of North America” (ibid.: 43).

### **References**

- Beeson, R. W. (1979). The improbable primate and modern myth. *Northwest Anthropological Research Notes*, 13, 91–110.
- Beeson, R. W. (1988). Improbable species, deceit, and social control in the context of behavioral ecology. *Northwest Anthropological Research Notes*, 22, 239–251.
- Brewin, C.R., Andrews, B., & Micks, L. (2020). Regaining consensus on the reliability of memory. *Current Directions in Psychological Science*, 29, 121–125.
- Bullard, T.E. (2010). *The myth and mystery of UFOs*. Lawrence, KS: University Press of Kansas.

- Clark, J. (2018). Experience anomalies. In J. Clark (Ed.), *The UFO encyclopedia* (3rd ed., pp. 444–446). Detroit, MI: Omnigraphics.
- Del Giudice, M. (2012). The twentieth century reversal of pink-blue gender coding: a scientific urban legend? *Archives of Sexual Behavior*, *41*, 1321–1323.
- Devereux, P. (2007). A moveable feast. In T. Pfeiffer, J.E. Mack, & P. Devereux (Eds.), *Mind before matter: Visions of a new science of consciousness*. London: O Books.
- Diamond, N., Armson, M., & Levine, B. (2019). The truth is out there: Accuracy and detail in recall of verifiable real-world events. *PsyArXiv*, 1–37. doi:10.31234/osf.io/ud63x
- Figueredo, A. J., Peñaherrera-Aguirre, M., Fernandes, H. B. F., Lomayesva, S. L., Woodley of Menie, M. A., Hertler, S. C., & Sarraf, M. A. (2019). The ecology of empire: The dynamics of strategic differentiation-integration in two competing Western European biocultural groups. *Politics and the Life Sciences*, *38*, 210–225.
- Hufford, D. J. (1982). *The terror that comes in the night: An experience-centered study of supernatural assault traditions*. Philadelphia, PA: University of Pennsylvania Press.
- Jussim, L. (2012). *Social perception and social reality: Why accuracy dominates bias and self-fulfilling prophecy*. New York, NY: Oxford University Press.
- King, J., Woodley of Menie, M. A., & Sarraf, M. A., & Greenwell, J. R. (2020). Eyewitness reports concerning a putative North American hominoid: anomaly or artefact? *Zeitschrift für Anomalistik*, *20*, 14–48. doi:10.23793/zfa.2020.14
- Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy. *Annual Review of Psychology*, *51*, 481–537.
- Kutner, M., Nachtsheim, C., Neter, J., & Li, W. (2005). *Applied linear statistical models* (5th ed.). Irwin, CA: McGraw-Hill.
- Lewis, J. E., Degusta, D., Meyer, M. R., Monge, J. M., Mann, A. E., & Holloway, R. L. (2011). The Mismeasure of science: Stephen Jay Gould versus Samuel George Morton on skulls and bias. *PLOS Biology*, *9*, e1001071.
- Lozier, J. D., Aniello, P., & Hickerson, M. J. (2009). Predicting the distribution of Sasquatch in western North America: anything goes with ecological niche modelling. *Journal of Biogeography*, *36*, 1623–1627.
- Meurger, M. (2006). *Histoire naturelle des dragons*. Rennes: Terre de Brume.
- Michel, J., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., ... & Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, *331*, 176–182.
- Nahm, M. (2019). Historical perspective: Justinus Kerner's case study into the "prison spook" in Weinsberg and spooky actions at a distance in 1835–1836. In G. Mayer (Ed.), *N equals 1: Single case studies in anomalistics* (pp. 153–200). Zürich: LIT.
- Paxton, C. G. M., & Naish, D. (2019). Did nineteenth century marine vertebrate fossil discoveries influence sea serpent reports? *Earth Sciences History*, *38*, 16–27.

- Shoalts, A. (2019). *The evolution of "Monsters" in North American exploration and travel literature 1607–1930*. Doctoral dissertation. McMaster University: Hamilton/Canada.
- Short, B., & Green, J. (n. d.). Big Foot Encounters: state by state sightings list. Retrieved from <https://web.archive.org/web/20180413195002/http://bigfootencounters.com/>
- Sparks, B. (2018). Debunking and debunkery: A skeptical history. In J. Clark (Ed.), *The UFO encyclopedia* (3rd ed., pp. 379–400). Detroit, MI: Omnigraphics.
- Stevenson, I. (1990). Thoughts on the decline of major paranormal phenomena. *Proceedings of the Society for Psychical Research*, 57, 149–162.
- Wixted, J. T. (2018). Time to exonerate eyewitness memory. *Forensic Science International*, 292, e13-e15.
- Wixted, J. T., Mickes, L., & Fisher, R. P. (2018). Rethinking the reliability of eyewitness memory. *Perspectives on Psychological Science*, 13, 324–335.
- Woodley of Menie, M. A., Fernandes, H. B. F., Figueredo, A. J., & Meisenberg, G. (2015). By their words ye shall know them: Evidence of genetic selection against general intelligence and concurrent environmental enrichment in vocabulary usage since the mid 19th century. *Frontiers in Psychology*, 6, 361.
- Wylie, K. (1980). *Bigfoot: A personal inquiry into a phenomenon*. New York, NY: The Viking Press.