Irreproducible research and a typology of replication efforts

Adrian Treves [1*]

Sedona Chinn [2]

Miha Krofel [3]

[1] Carnivore Coexistence Lab, Nelson Institute for Environmental Studies, University of Wisconsin, Madison, WI 53706, U.S. atreves@wisc.edu ORCID: 0000-0002-3052-4708

[2] Community & Environmental Sociology, University of Wisconsin-Madison, Madison, WI 53706, U.S. schinn@wisc.edu ORCID: 0000-0002-6135-6743

[3] University of Ljubljana, Biotechnical Faculty, Address: Jamnikarjeva 101, 1000 Ljubljana, Slovenia. miha.krofel@gmail.com ORCID: 0000-0002-2010-5219

*correspondence: atreves@wisc.edu

Abstract

The scholarly and scientific literature does not automatically correct itself. Erroneous findings may persist without correction or retraction. Following prior definitions of zombie articles that are retracted but continue to be cited affirmatively, we add another category of 'undead' articles. We define 'vampire' articles by two jointly necessary conditions. The first condition is irreproducibility, demonstrated by one or more failed efforts at replication or conditions that make replication impossible. We offer a novel typology of four categories of replication efforts. We propose a rule of thumb for how many failed efforts at replication might be required for each type of replication effort before the original finding is deemed irreproducible. The second necessary condition for a putative vampire article is that it is cited affirmatively in public policy, the scholarly literature, or private (non-governmental, not published in the scientific literature)

communications, after the first condition is met. We discuss rules of thumb for how many such affirmative citations might lead qualified researchers in that subfield to propose correction, retraction, or editorial note of concern for the article in question. Our first case concerns aerial shooting at coyotes and the second case predicted over-fishing. We discuss the damaging effects of vampire articles and why the metaphor has heuristic value and utility. Perhaps the current retraction guidelines could be sharpened to include considerations around vampire articles. We also discuss lessons from the communication sciences about how to remedy misinformation, offering recommendations to researchers, publishers and editors, concerned with correcting their literatures and public trust in science generally.

Keywords: correction, post-publication review, replication, reproducibility crisis, retraction, scientific publishing, trust in research, unreliable findings

Background

Ideally, science and other research progress by replacing weaker evidence with stronger. In so doing, previous inference can be updated, and – occasionally – transformed into strikingly different knowledge about how the world works. But the weaker research is not necessarily expunged after replacement. Troublingly, inferences from weaker research can persist by different actors intentionally or unintentionally citing it affirmatively in subsequent research, policy, or private sharing. Such scenarios are particularly troubling when an article has been retracted for falsification, fabrication, major breaches of integrity, or uncorrectable flaws. Some in the scientific community call these articles, ‘zombie ’articles’ (1-3). Here, we propose to add to this taxonomy of the ‘undead’ research finding’, by describing ‘vampire articles’. Unlike zombie articles, so named by analogy with the undead (retracted article) still moving and threatening the living (being cited affirmatively), here we consider broader harms to knowledge

systems in our concept of vampire articles, which are not corrected or retracted, despite failure to replicate or impossibility of replication, and yet continue to be cited affirmatively in policy, research, or public discourse. The irreproducible findings of vampire articles can mislead the public and can distort otherwise sound policy and waste research resources investigating unsound claims The affirmative citation of erroneous findings drains the vitality from public policy and discourse or future research. Therefore, to build trust in more reliable research findings, e.g., (4), we must all work together to find a way to identify (and then to purge) the unreliable articles by transparent standards. We acknowledge the metaphor of undead articles makes the topic more memorable (Figure 1), but we find heuristic value in the metaphor (see Discussion). We strive to treat vampire articles as a criticism related to the methods and findings, and not to the original authors, editors, or publishers.

Why our topic is important and two cases that illustrate the importance

To understand the broader harms of undead articles, we consider the pathway to zombie articles which illuminates the two added steps that lead to vampire articles (Figure 1). Many scholarly papers contain errors, as do zombie articles. Zombie articles pass peer review and become published but are caught by later screening and retracted, often by the authors themselves, editors, or post-publication reviewers. Although the delay may be long and the process tortuous, zombie articles are ultimately retracted, a case of the ideal of the scientific literature being corrected, despite continued citation. On the other hand, some articles with errors elude correction or retraction despite repeated evidence that the findings are not reproducible. That is our first condition for the genesis of a vampire article, described further below. Eluding retraction or correction might prove fairly innocuous, if researchers, policy-makers, and the public ignore the erroneous findings. Indeed, the history of science contains hundreds of

examples of erroneous articles that are never corrected or retracted but, instead, are superseded and forgotten. We are not concerned with this large historical bank of unreliable findings, unless an article fulfills the second criterion. The second condition is that the erroneous article garners attention and spreads in one or more ways (research, policy, public). Only the combination of both conditions (unretracted irreproducible findings that are spread privately or publicly) can make a vampire article.

Figure 1. The metaphor of undead articles juxtaposed for heuristic purposes. The scholarly literature is full of unreliable findings laid to rest (graveyard), but some articles continue to be cited despite their retraction (zombie), and yet others resist correction or retraction and cause harm (vampire). We add our own text to an AI-generated image by Siri-assisted AI in Mac OS 15.7.3 using the prompt "Use ChatGPT to draw a wide stone archway in the foreground with a graveyard in the background. Add a headstone with no inscription; add a silhouette of an unearthed grave with a zombie crawling out of it on the right side; add a shadowy silhouette of Nosferatu in the center foreground. Draw in a gloomy grayscale but realistic style."

Irreproducibility

The first condition we propose for vampire articles is irreproducibility, which we define as repeated failures to replicate the finding or impossibility of replication. Replication is an attempt to repeat a finding and is therefore a benchmark of reproducibility (5, 6). Replication depends on the original authors clearly and thoroughly sharing methods and data (7). Replication of a study refers to subsequent good faith, qualified efforts at repeating a finding. Replication efforts differ in type of approach and their requirements.

Typology of replication

We offer a typology of replication in Table 1 followi9ng and adding to prior work (6, 8-10). Table 1 begins from the premise that replication efforts must either use the same methods as the original finding or try to repeat the same finding by different methods. Table 1 describes replication efforts by the characteristics of the replication effort in relation to the original finding; collection of new data or scrutiny of the original data; its purported cause-and-effect relationships (mechanisms); the conditions under which replication might be possible or impossible; and the number of failed replication efforts we consider meaningful enough to conclude the 'finding failed to replicate'. The typology provides recommendations not strict rules for when the balance of views in the research community should shift from general skepticism to considering a particular finding to be irreproducible. We offer a threshold with a mix of quantitative and qualitative criteria for failures to replicate, but suggest that in many cases, statistical expertise will be needed to conclude that a study is irreproducible based on the strength of inference derived from all replication efforts, rather than a numerical threshold based on the number of studies.

Table 1. Typology of replication efforts.

	Type of replication published in peer-review literature *			
Attributes of replication efforts	**Exact**	**Technical same data**	**Technical new data**	**Conceptual**
Characteristics of the replication effort	All methods and data are identical	Same data different analytical or statistical methods	Different data, same or similar methods	Data and methods differ
New data must be collected	Not necessarily	No	Yes	Yes
Cause-and-effect relationship	Same	Same	Same	Different causal mechanism, similar effect
Conditions that make replication possible (and rarity)	Often only possible for studies with simulations that are code-based or computer-assisted	Common because replicators need not collect new data.	Uncommon because replicators must collect new data and replications can fail for many different reasons.	Rare, requires two pathways to the same outcome.
Conditions that make replication impossible	Original data are not shared, or steps in methods are missing or contradictory ; or one step was unique and cannot be simulated or repeated	Original data are not shared	steps in methods are missing or contradictory ; or one step was unique and cannot be simulated or repeated	Only one causal mechanism or pathway exists

Guidelines for the approximate number of replication efforts that must fail to treat a finding as irreproducible	Usually, 1 but more if the replication efforts do not produce strong inference	Usually, 1 but more if the replication efforts do not produce strong inference	2 or more (variability described in main text)	Usually, 1 but more if the replication efforts do not produce strong inference

* Efforts to replicate must also be published in the peer-reviewed literature defined as the process undertaken by independent qualified researchers overseen by a journal that has subscribed to the COPE guidelines (25) .

A major dichotomy in replication efforts hinges on whether new data must be collected (Table 1). Exact replications may collect new data but need not, as for example, when random-number generation and simulations are used as input data in both the original and the replication. Technical replication using new data and conceptual replication both require new data collection.

A prerequisite for all replications is that original methods and data were shared, and that the replication teams use these conscientiously and meticulously. If the former condition is not met, see below for impossibility of replication. If the latter condition is not met, a subsequent effort cannot be considered a replication. Regarding transparent methods, replication should be feasible without necessarily consulting the original authors, although such consultation can be helpful for interpreting complex, non-standardized methods.

Exact replications: Because these replications demand exactingly identical methods it seems to us to be restricted to artificial standardized conditions such as computer models in which software versions, scripts, code, variable specifications, etc. are all precisely identical between efforts. Because of the demanding requirements of exact replication, we propose 1-2

failed replications should be sufficient to label a finding irreducible. Two replication efforts would be needed if the first effort misinterpreted methods or included subjectivity in methods, for which reasonable alternative approaches exist. We acknowledge there will always be some subjectivity. Anyone may find a flaw in the published attempt and propose that a second replication effort is needed.

Technical replications: Technical replications do not demand exactly identical methods. Technical replications with the same data consist usually of statistical re-analysis or applying a different statistical approach, e.g., (11). Such replication efforts seem more common than other types in Table 1, in part because it is less expensive to repeat statistical analyses than collect new data. Because the choice of which statistical approach to use sometimes reflects differences in training and subjective opinions, we recommend statistical experts be involved, or better yet lead, the assessment of whether a technical replication effort with the same data has failed on statistical grounds (12). Because technical replications with the same data may identify a flaw in the original finding based on improved statistical analysis, a single failed replication may be sufficient. But some efforts at technical replications with the same data may fall short of repeating the methods precisely enough, so we suggest case by case consideration of alternative explanations for a failed replication effort, such as dissimilar methods, should be examined. That leads to our range of recommended replication attempts of 1 or more. The second attempt should meticulously repeat the methods in such cases or the uncertainty will persist.

Technical replications with new data present unique challenges because methods are neither exactingly identical nor are the input data the same. Inevitably, subjects, sites, even the researchers may be subtly different, e.g., (10). Therefore, we propose 2 or more failed replications would be necessary for this category (Table 1). Indeed, it could take many efforts at

replication with consideration of effect sizes and strength of inference before this category of replication leads the community to judge the first finding irreproducible. At some point, the tools of meta-analysis might be more appropriate than a qualitative judgment, and certainly, opinion within the community of qualified researchers will differ in such cases. It seems prudent to raise the threshold for technical replications with new data even when the original finding was surprising by other criteria (14). Technical replications with new data are likely to require more failed replications. An effort at technical replication with new data often falls short of confirming or rejecting reproducibility because the subsequent study teams might find it difficult to repeat every step precisely as done by the original study. Exact replication with new data is impossible for obvious reasons, as the replicators contend with randomness introduced by different sites, times, subjects, etc. The difficulty of controlling every variable — from subject animals in labs (10) to wild animals in the field, different investigators, and the many other variables outside lab settings — can be considerable. Uncontrollable variables in technical replications with new data make us cautious about labelling an original finding irreproducible. For example, five field studies of wild predator attacks on domestic animals, under diverse human husbandry systems, evaluated the effectiveness of deterrent lights called Foxlights®; the five studies found everything from the desired treatment effect (deterring predators away from domestic animals), no treatment effect, to a perverse undesirable effect (attracting predators to domestic animals) (15-18). Because the sites, domestic animals, predators, researchers, deployments, and surrounding conditions all differed across the five studies, it is unclear if these technical replications with new data should be considered a basis for doubting the first finding of a desired treatment effect (18), the perverse undesirable attractant effect also reported for a different predator in the same original finding (18), or some third option. In fact, the mixed findings

across different conditions may lend valuable insight into the conditions under which similar treatments may have different effects. Similar issues arise in social scientific surveys when wording of questions vary, order effects between questions arise, identity of investigators or respondents differ between studies, or mode of data collection varies. We recommend a higher bar of more failed replication efforts and caution by investigators to specify the possible differences in methods that confound interpretations (Table 1). See the section below on more complex cases.

However, in some cases a single failed technical replication with new data might be sufficient. For example, the well-funded Open Science Collaboration (OSC), "…attempted replications of 100 experimental and correlational studies published in three psychology journals using high-powered designs and original materials when available." Abstract, (13). Despite OSC reporting that they worked closely with original authors, they reported replicated treatment effects that were half the magnitude of original treatment effects. Only "36% of replications had significant results; 47% of original effect sizes were in the 95% confidence interval of the replication effect size; …, if no bias in original results is assumed, combining original and replication results left 68% with significant effects." Abstract, (13). Under the highly controlled lab conditions enjoyed by the OSC replication teams, one might argue that a single failed replication is sufficient to demand scrutiny of the original finding. Those authors did not label any finding irreproducible based on one failed replication but rather restricted themselves to discussing confidence intervals and overlap in effect sizes when discussing discrepancies between original findings and replication findings.

Conceptual replication: Finally, this category requires that a given finding can be obtained by two different pathways. We are unsure if such replications are possible in more than

196 a handful of specialized examples because this type of replication presupposes distinct causal
197 mechanisms producing the same effect. Moreover, the studies must be powerful enough to
198 elucidate causal mechanisms, which is not always the case. However, we see a common thread
199 between the following examples. In the first example of a cancer study, a gene mutation (the
200 inferred causal mechanism) was correlated with the origins of cancerous cells in the original
201 study, whereas an alternative mutation (alternative causal mechanism) leading to the same cancer
202 was discovered by a subsequent study (8). That seems to be a conceptual replication, which
203 aimed for the same outcome by a different mutation pathway. The conceptual replication seems
204 to have disproven the first causal mechanism in one replication effort. In the second case,
205 investigators ran the same treatment in two experimental designs. The experimental designs
206 produced different outcomes for the same treatment (21). The latter study showed that
207 inconsistent findings resulted from the decision to pre-test survey outcomes (21). In this case,
208 replication revealed a new causal mechanism triggered by variation in experimental design.
209 Although our two examples seem to suggest one replication effort might be enough if the design
210 is robust enough to reveal cause-and-effect, we acknowledge that maybe replications of both
211 causal mechanisms would be required. We also acknowledge most replication efforts are more
212 difficult to interpret as we discuss next.

213 **More complex cases of mixed success and failure of replication**

214 We anticipate cases in which the original article is subject to both successful replication
215 and failed replication and the replication efforts may be in different categories (Table 1 and see
216 our case 2 below). For example, findings on wolf mortality show conceptual replications and
217 technical replications with and without new data. First, an unobserved phenomenon (illegal wolf-
218 killing by humans) was inferred to be caused by a policy change (reviewed in 19,20). A

conceptual replication of the same system investigated a different pathway to test the correlation between the policy and the human behavior (poaching) (19,20). The first pathway (population-level observed undetected population-level dynamics and inferred wolf mortality whereas the second pathway observed individual survival and fates of marked wolves. Although that conceptual replication supported the original inferred causal mechanism with some refinements (19), a series of other studies have been proposed as technical replications with the same data (19, 20). On top of that confusing mix of findings, teams have published critiques and commentary on each, cf. reviewed and cited in (19, 20). In such protracted debates, one hopes for new data or advances in methods because the tangle of authentic replication efforts is obscured by flawed efforts, reviewed in (19, 20). There may also be cases in which teams of investigators fail to replicate, demand more information and the original finding is clarified or explained better, potentially triggering a new round of replication efforts (20). The case of wolf-poaching above is also characterized by independent authors' repeated concerns about the putative impossibility of replicating a finding because something essential is deemed to be missing ((19, 20). We explore impossibility of replication next.

Impossibility of replication

Impossibility is defined as "that cannot be, in existing or specific circumstances" Oxford English Dictionary online accessed 15 October 2025. The definition helps point the way because a finding that is impossible to replicate is one whose methods or data cannot exist as specified or did not exist at the time of the study. Impossibility of replication might arise in research papers if the methods are incomplete, self-contradictory, or stated in such a way they cannot be followed for replication by a qualified investigator. In other cases, data are not shared in a way that they can be re-analyzed, or under any circumstances, collected again (Table 1). Here, we do not refer

to research into the past, whose historicity' (unique past event) makes it almost impossible to repeat observations (22, 23). Rather, we refer to studies that fail to share data or report events that cannot be simulated or repeated under any circumstances.

In sum, when the original finding is either subject to one or more failed efforts at replication (Table 1) or deemed impossible to replicate in peer-reviewed studies, we might consider labelling that article irreproducible until further clarification or the original conditions return. We acknowledge labels on articles are problematic. Editorial boards may put labels on articles they have published (24, 25); an investigator may publicize their own loss of confidence in their own finding or publicize their own corrections or retractions (26, 27), but there are few trusted mechanisms for third parties to label an article as 'irreproducible' (28). To fill that gap, associations of researchers and reporters have sprung up to describe research misconduct (e.g., 25, 71), e.g. PubPeer, and a body of scholarship describes ways to transparently debate evidence and reproducibility (12, 29, 30). Therefore, the watchdogs and accumulating scholarly expertise support more careful and more discerning conduct by original authors; second-party editors, peer reviewers, publishers of scholarly articles; and third-party replicators. We turn to the issue of consensus and skepticism in the research community next.

Guarding against misuse of the label 'irreproducible'

Although replication efforts are currently rare and specialists in the field of a suspect article may be required to confirm whether an article might be irreproducible, nevertheless the threshold for claiming irreproducibility should be high. The threshold should be high to prevent the misuse of this label to disparage a rival team. As stated above, there are many kinds of replication which may not only replicate the original finding but also offer further insight into conditioning factors; discovering new causal pathways or conditioning circumstances do not

make a finding irreproducible. Our proposal in Table 1 for a number of failed replication efforts and clarity about types of replication efforts should motivate researchers whose findings are challenged by peer investigators to repeat their own findings. Therefore, a label of irreproducible or even vampire article would not be irreversible.

Remedies for irreproducibility

Even when all involved agree collegially that a published finding is irreproducible, the remedy may not be obvious. Publishers or editors may resist correction or retraction if no clear error can be identified other than failure to reproduce or impossible to reproduce (30), which we explore in case 1 below, or authors may not pursue labels or warnings to readers even when they agree that their own original finding might be in error (see our case 2 below).

Retraction could of course transform a vampire into a zombie article and, in ideal situations, can be a salutary process (e.g., (26). Correction may cure irreproducibility, for example, with the addition of omitted information on methods or the presentation of data that were not shared in the original. However, continued affirmative citation of zombie and vampire articles affects public discourse and policy as we have noted briefly above. We turn to this issue next.

Affirmative citation

The second condition of our proposal for a vampire article is that it elicits affirmative citations in public policy, the scholarly literature, or private communications (hereafter affirmative citation), after the first condition of irreproducibility is met. We follow a prior definition of affirmative citation, "…confirms, is supported by, depends on, agrees with, or is strongly influenced by the cited work", p.2 in (31). It is beyond our scope to examine in detail the diffusion of ill effects from erroneous research and how these may distort public policy or

investments in further research when cited affirmatively. But we devote considerable discussion to lessons learned in the communication sciences about misinformation and how to combat it.

We admit that a small number of affirmative citations in the scholarly literature or private communications do not worry us necessarily. Yet we are concerned when a government cites an erroneous study affirmatively even once, if that single citation leads to cascading real-world effects. Clearly, the judgment about whether affirmative citations warrant concern, and how many such citations, will be a subjective value judgment. We leave it to the researchers qualified to judge when the literature in their subfield needs correction. Nevertheless, pending meta-research on this question and pending a subfield's own introspection on the criteria, we offer some guidelines based on what we perceive as common sense from our case studies.

If a single affirmative citation by government leads to a shift in priorities for funding, staffing, or effort, such shifts can produce cascading effects on human and nonhuman safety and health, opportunity costs, etc. (see Case 1). Accordingly, when the affirmative citation is to the erroneous finding, a flawed method, or tries to build on the findings with entirely new projects, our concerns grow. When affirmative citations in the scholarly literature rise (or private communications such as online influencers start to get hundreds of mentions), we recommend the qualified researchers in a subfield ought to act. By contrast, when the original authors are the only ones citing themselves affirmatively concern should lessen. Likewise, when the affirmative citations accumulate slowly over years, concern may not be warranted. Finally, when affirmative citations are not to the erroneous finding but to some other aspect of the putative vampire article (e.g., a step in the methods or an insight in the discussion), our concerns diminish.

To avoid misleading the public or policy-makers, speed is essential. Scholarly research is not known for its speed relative to public uptake and the speed of misinformation (28, 32-34);

furthermore, replication efforts, corrections, or retractions may face additional delays due to constraints on the scholarly publication process (12, 29, 30). Given the relatively slower pace of strengthening or correcting the research literature and purging errors from it, it behooves us as authors of potentially erroneous work to respond with alacrity when problems with our published work is brought to light. Some of those concerns may be mistaken of course. But re-checking and verifying that our work is unimpeachable and doing so collaboratively without entrenching positions or creating animosities seem important for public policy and for fostering greater trust in science and research broadly (35).

The research community nonetheless faces an up-hill battle as errors may spread faster than corrections or retractions; indeed, affirmative citations to erroneous results often outpace affirmative citation to the corrections of those erroneous results (36), as we explore in case 2. Our two cases illustrate the time it takes to correct irreproducible findings cited affirmatively and the ultimate failures when the original authors or publishers do not take appropriate action. Our first case presents two technical replications with the same data, followed by evidence of omission of information and conflicting statements of methods by the original first author. Policy-makers perpetuated the irreproducible findings. Third parties impeded efforts at correction or retraction (Case 1). In the second case, the original authors themselves attempted correction by publishing an improved, new article, but appeared unable to stop the spread of misinformation from their earlier erroneous findings. Also, the second case describes a large number of replication efforts or challenges to the reproducibility of the original findings in >10 subsequent rebuttals (Case 2).

Case 1: Aerial shooting at coyotes to protect free-ranging sheep

Wagner & Conover (37), hereafter W&C1999, reported the outcomes of a treatment described as aerial shooting at coyotes during winter in mountain pastures used by private sheep owners on public grazing allotments in the western U.S. They studied mountain pastures during summer months, in which government agents undertook mechanical and explosive trapping, snaring, and ground-based shooting (defined as "summer pasture management", SPM). Those were labelled 'untreated' pastures (a pseudo-control because killing by SPM occurred on both treated and control pastures). Thus., control pastures experienced only SPM, whereas treated pastures experienced summer SPM and aerial shooting during winter. In 2004 and again in 2016, the findings were challenged on several grounds by two independent teams conducting technical replication with the same data and alleging irreproducibility. Later, the lead author of the original paper made comments in sworn testimony that suggested the impossibility of reconciling the published methods. Therefore, this case presents evidence for irreproducibility by the criteria in Table 1.

First, (38) raised several problems in a technical replication using the same data and published in a sister journal of the same publishing society. The publisher becomes relevant later as we discuss below. (38) noted pseudo-replication of subject pastures, biased selection of treated pastures in favor of a treatment effect, and multiple comparisons without adjusting the significance threshold, see p.1213 (38). W&C19999 did not correct, retract, or explain methods. Second, (39) added to the technical replication with the same data but also to the evidence of irreproducibility by identifying instances of sampling bias by post hoc, non-random selection of the pseudo-control pastures, treatment bias, and omission of methods (detailed in Additional file 1 Table 1, discussed below). Third, the irreproducibility of W&C1999, was further emphasized by events in a 2018 U.S. federal court case (40). In 2016 and 2018, W&C1999 became the

356 subject of a sworn affidavit co-led by the lead author in W&C1999 as “primary author” in (41)
357 (full document presented in Additional file 2). The documents that (41) entered into the court
358 record contained novel information on methods, some of which contradicted the published
359 methods in W&C1999.

360 In Additional file 1 (SI 1, Table 1A-C), we present irreconcilable descriptions of methods
361 make an article irreproducible. Concerned by the appearance of different methods, omissions of
362 material data, and logical inconsistencies about W&C1999 (Additional file 1 SI 1 Table 1A-C),
363 14 scholars sent a letter in July 2018 (Additional file 2) expressing their concerns to the editor in
364 chief of the journal that published W&C1999. The editor-in-chief and the publisher, The
365 Wildlife Society, responded that in their opinions a retraction was not warranted and that
366 concerns should be best addressed by a peer-reviewed comment submitted to their journal
367 (Additional file 2). Note that at the time, the publisher did not subscribe to the Committee on
368 Publication Ethics (COPE, 25), which might have affected their responses to concerns. Only
369 later, in 2022, did the publisher sign on to COPE under pressure from critics addressing
370 unrelated, other articles by the same publisher (42). This case illustrates an important issue about
371 the role of editors and publishers in the perpetuation of vampire articles.

372 Third party replicators cannot confirm or reject which of conflicting methods should be
373 ignored or corrected (Additional file 1 SI 1 Table 1A,B). Third parties can only raise questions in
374 such cases. Therefore, the onus is on publishers and editors to confirm that the review process
375 was untainted and that authors disclosed all necessary information. Second, the editor and
376 publisher dismissed three concerns about irreproducibility: a published technical replication in
377 the sister journal Wildlife Society Bulletin (38); a second such effort pointing to new problems in
378 analysis an finding incomplete methods precluding replication, published in a different family of

journals (39); and evidence from the lead author of W&C1999 swearing under oath to methods that were not in the original publication (Additional file 2). To this day, the editors have not corrected, retracted or posted editorial concerns over W&C1999. The invitation to submit to peer review in the original journal seems unfair because no new methods or data could be offered and peer reviewers are not generally qualified to judge if sworn affidavits in court bear on an editorial decision about a prior article (W&C1999), which they are not reviewing. In short, the editor-in-chief had the necessary evidence in hand (Additional file 2) to make a judgment about the need for a notice of concern, correction, or retraction. Instead, nothing changed when the 14 authors of the letter of concern declined to submit to peer review in the same journal. Therefore, we view this case as one in which interested parties (publisher and the journal editor in chief) protected a vampire article. Perhaps the current retraction guidelines could be sharpened to avoid reduce the likelihood that publishers willbury their heads in the sand to avoid confronting an irreproducible article.

Finally, W&C1999 has been cited 25 times since 2004 (first rebuttal published) and 19 times since 2016 (second rebuttal published; citation counts from Google Scholar® accessed October 2025). Therefore, this case fulfils both conditions for a vampire article. Also, the 2018 lawsuit was in part over the U.S. government agency (USDA-APHIS-WS) affirmatively citing W&C1999. That agency continues to do so (72). Therefore, this case fulfils both conditions for a vampire article. Fortunately, iit is never too late for a journal (or agency) to correct the record.

Case 2: Projection of future fish stock declines

A high-profile 2006 paper by Worm et al. published in Science predicted that many important fishery stocks would be almost completely depleted by over-fishing within 48 years (43). Broadcast media and prominent presses picked it up then and continued to broadcast the erroneous conclusion, even after multiple rebuttals and failures to replicate, including by the

original author. The particulars of this case have been well summarized by previous authors, and we quote them at length below. Nevertheless, we present it in summary fashion again to illustrate how the original article arises as a vampire, when it is neither corrected nor retracted (43). Indeed, in this case, Science offers the utility tool of an eLetter comment on articles, so the authors had a third recourse of expressing no confidence in their findings, especially after ten rebuttals (36) and their own revision was published in 2009. Also, we present the case to explore the recommendations on how to correct misinformation in the research literature before it does additional harm.

Banobi et al. (36) reviewed the aftermath of Worm et al. 2006 (43) and its 11 rebuttals. The original unreliable study remains published and continues to be cited more (2.3 per year) than the rebuttals (1.5 per year per rebuttal) in the scholarly literature (36). For Worm et al. 2006, "97% of citations to the original challenged article were uncritical and did not cite the rebuttals"." p.9, Banobi et al. Indeed, they found rebuttals of 7 fisheries studies were cited <6% as often as the original rebutted article. Therefore, this case illustrates the waste of time and resources to challenge an irreproducible finding that is widely disseminated in public discourse. This case also illustrates the resistance of broadcast media to report on failed replications. Banobi et al. summarized, "…11 July 2010 headline in the prestigious London newspaper, The Sunday Times, trumpeting 'Fish stocks eaten to extinction by 2050' …," p.9, (omitting an internal citation to the London Times reporter). That inaccurate claim was based on Worm et al. 2006 (43). Not only does the article fail to mention any of the 11 rebuttals that question this projection, but it misses the later consensus paper by the same author and many of his critics that reverses the earlier projection of collapse and instead expects rebuilding to occur in 5 of 10 well studied ecosystems (Worm et al. 2009).". p.9, (36).

This case illustrates why articles with errors should be corrected or retracted, not simply addressed by the authors with a subsequent publication that might be missed by media, policy-makers, and lay readers. Indeed, we agree with (36) on the perils of misinformation in science, when they wrote, “High-profile articles such as those discussed here receive wide public attention outside the biological research community; they form the basis for headlines and sound bites, and help to shape public opinion on issues such as marine conservation, and voters in turn influence the decisions of policy-makers. Thus, high-profile research findings have a compounded impact, making it even more crucial that public policy is based on balanced science reflecting all viewpoints, and not just on the science as it is first reported.” p.9, (36). They suggested what to do about irreproducible research. Among their several prudent recommendations, we echo two.

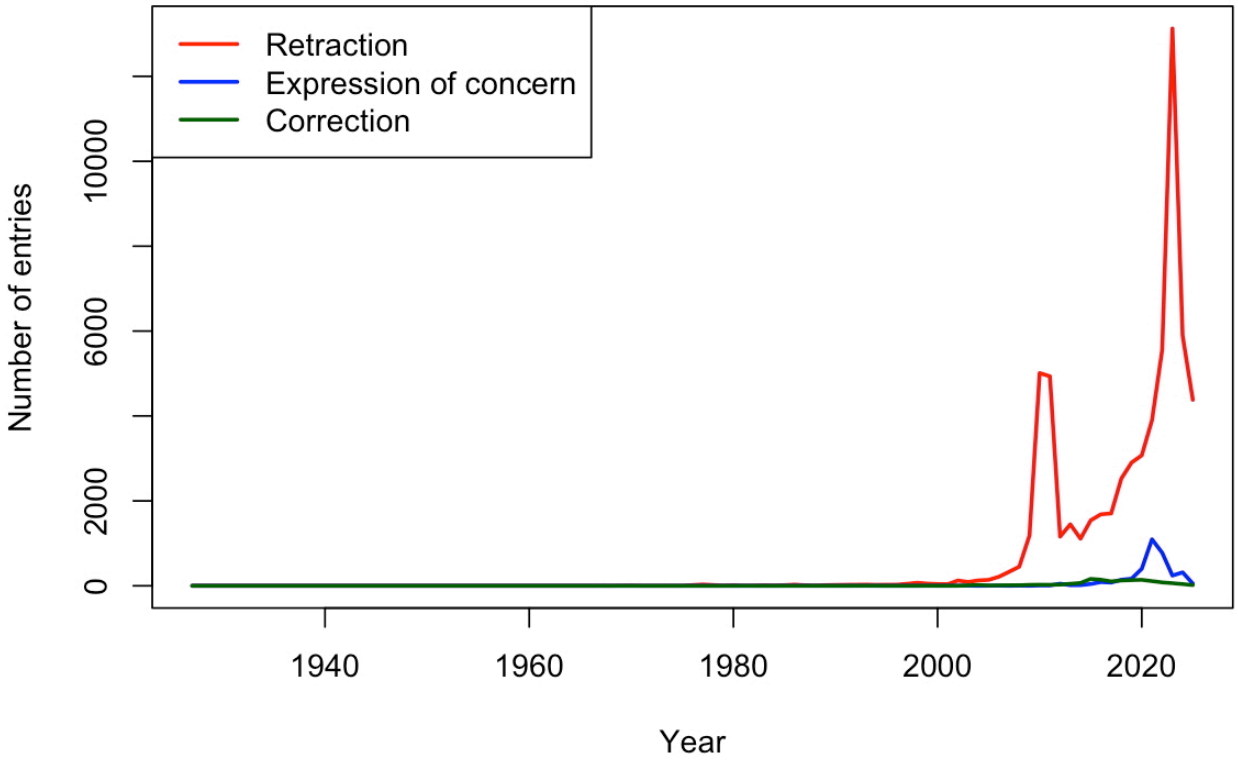
First, editors should link rebuttals permanently to the original article. We would add that rebuttals published in the same journal might make that linkage stronger, even if the rebuttal exceeds word limits or other stylistic issues in the journal. Second, (36) concluded, “Our results indicate that rebuttal authors may to a large extent be wasting their breath.” p.9. Their analysis shows that citations to challenged articles did not diminish over time, no matter how many rebuttal articles were published. Therefore, we propose the remedy is retraction, correction, or bold editorial notices on the title page of online, irreproducible articles

Banobi et al. (36) showed that affirmative citations by broadcast media to the replication efforts were low or zero whereas the ‘splash’ of affirmative citations to the first erroneous article. Our label ‘vampire article’ will be helpful in that regard by making clear to editors and publishers that replications have failed, and public harms are underway.

Discussion

449 The international scientific community is grappling with the substantial number of
450 retractions and corrections in the literature (Figure 2; 71). A category of published findings, that
451 were retracted but continue to be cited was recently named zombie articles (1-3). Here we add to
452 the class of 'undead articles' by defining 'vampire' articles.

453 **Figure 2.** The annual number of retractions, corrections, and editorial notes of concern (y
454 axis) per year (x axis) drawn from the Retraction Watch database accessed 14 January 2026 (71).



455

456 Vampire articles fail independent, peer-reviewed efforts at replication or are found to be
457 irreproducible because of the impossibility of replication, yet remain in the literature and cited
458 affirmatively. Both conditions, irreproducibility and affirmative citation, must be met before one
459 can identify a vampire article. Identifying vampire articles will be more subjective than
460 identifying zombie articles because journals have not yet identified, (flagged, corrected, or
461 retracted) the former.

462 To raise a high threshold for claiming an article is a vampire and reduce the subjectivity
463 of identifying them, we offer what we believe is the first typology of replications and criteria for
464 labelling a finding irreproducible (Table 1). We add that the category of technical replication
465 using new data may require a higher threshold of failed replication efforts before being labelled
466 irreproducible because studies in some fields inevitably generate numerous potentially
467 confounding variables when new data are collected for replication purposes. Thus, we presented
468 two cases in which high thresholds were met to label articles irreproducible (Cases 1, 2). We
469 welcome additions or modifications to the typology in Table 1.

470 Remedies for our cases and for other vampire articles may require third parties, not just
471 the research community stopping citing irreproducible articles. A single government agency
472 might achieve this for Case 1, as the US Department of Agriculture has promoted the vampire
473 article despite evidence for its irreproducibility. Remedying Case 2 may require broadcast media
474 to collaborate with researchers to write headlines about a correction or retraction of the original.
475 Of course, news outlets like sensational or arresting headlines; in this case, some good news (fish
476 stocks not as imperiled) might attract viewers. Moreover, in Case 2 at least, the subsequent
477 publication of a revision by the same lead author in 2009, also in Science, and 10 other rebuttals
478 and reanalyses, did not prompt a new wave of more correct media coverage. But scientists can
479 encourage dedicated, qualified reporters to correct the public record also.

480 Would authors be motivated to participate and expose their original error? Although
481 some have published lessons they learned from their retractions, e.g., (26), we suspect many
482 would have fears to do so. Perhaps researchers who acknowledge their own fallibility and
483 advertise their doubts or corrections of their own findings will someday gain in reputation rather
484 than the converse.

Rebuttals and failed replication efforts seen through a lens of misinformation

Although we deem rebuttals essential to identifying vampire articles, the way those rebuttals are written should also be considered carefully to avoid the spread of misinformation. Repeating misinformation even in a rebuttal may be counter-productive (44). For one, repetition of misinformation decreases the efficacy of future corrections by making the misinformation more fluent—easy to remember and retrieve in one's mind—and therefore more likely that the misinformation will subconsciously influence attitude formation and judgment (34). Repeated misinformation can come to feel familiar, which increases its credibility (45) and may contribute to keeping an inaccurate study alive in public debate. For this reason, researchers writing rebuttals should tread carefully. A summary of the original article presumed wrong will be necessary, but the conclusions of the rebuttal should be repeated, strictly limiting references to demonstrations of exactly why and how the article is wrong (33). Indeed, there is some research showing that a focus on explaining flawed logic can be as or more effective at correction than fact-focused approaches (46).

Also, rebuttals should aim for catchy shorthand and pithy titles, in hopes of achieving memorability, particularly in conversations with non-researchers (33). Complex explanations are more difficult to process than simple ones, which can disrupt engagement with the findings (47). Beyond rebuttals, the research community and its allies will require an array of different interventions (33). Stating that rebuttals should repeat accurate conclusions delivered in simple, catchy language is not so easily done. The typology of undead articles may be a step in that direction (e.g., zombie articles and vampire articles are memorable ways to criticize poor research). But the actual findings may be impossible to communicate quickly and simply to lay audiences.

Terminology and the metaphor of vampires

We acknowledge the metaphor of undead articles makes the topic more memorable as needed to fight misinformation (see above), not necessarily more scholarly or precise (Figure 1). However, we list several advantages of using the metaphor of undead for such articles.

Vampires suck the life blood. Just as mythical vampires drink blood or suck the vitality from their victims to remain active, erroneous research that is cited affirmatively cam harm many audiences. Researchers may waste time and resources pursuing fruitless avenues suggested by the vampire article. The public and policy-makers can be misled, producing unsound policy and individual behavior. The life-blood of research is reliable findings fed by public and private resources that contribute because they trust research to accomplish clear goals. That relationship is threatened by vampire articles that drain the life-blood of research and public policy deliberations.

Sunlight kills vampires. Sunlight is the best disinfectant, to paraphrase USA Supreme Court Justice Brandeis (48). This metaphor captures the importance of transparency in the modern open science and research integrity movements (12, 30, 49-52). The opposite of sunlight is secrecy. Non-disclosure of competing interests played a large role in the delays caused by researchers working for the tobacco, petroleum, firearms, plastics, and other industries hiding harms or exaggerating the benefits of their products for profit (32, 53-55). Secrecy is a tool for concealing weaknesses. Vampires in myth hide from sunlight.

Vampires are harder to kill when living humans protect them. As in the myth, ancient vampires recruit living protectors. Individuals and groups outside the research community may proliferate irreproducible findings that capture the popular imagination. Consider the resilience of pseudo-science ideas of alien abduction, intelligent design, climate change denial, and vaccine

sceptics, all of which have been subject to politicized campaigns of disinformation (54). Many different, individual motivations can help weak evidence overwhelm stronger evidence. The affirmative citation of vampire articles by the public or by policy-makers also makes them harder to correct or retract than simple unreliable findings buried in the scholarly literature. But blame should not rest solely or even largely with the public.

Editors, reviewers, and publishers can intentionally or unintentionally introduce personal, professional, or ideological bias into decisions about which evidence to publish or correct (Case 1). Some findings are advanced by the persuasiveness of a source, whether by individual force of personality, prose, or reputations, the wealth of organizations or individual donors, or government approval. All these factors have been shown to protect or proliferate some unreliable findings (32, 53-57).

Awareness of these issues has led the scientific publishing community to advance research integrity. For example, the influence of voluntary groups, such as COPE, the Committee on Publication Ethics (25), has grown because they discuss and promote ethical practices in research publications. For letters of concern, corrections, and retractions, COPE (25) provides practical guidelines for strengthening publication integrity. Perhaps the current retraction guidelines could be sharpened to include considerations around vampire articles. Likewise, a few individual publishers are taking the lead and putting pressure on their peers to improve (7, 58). Therefore, we make an explicit recommendation to editors and publishers, which is implicit in the literature on retractions. When letters of concern are received by editors and when failed replications are called to the attention of editors, those original articles cited affirmatively in public policy and subsequent research deserve more attention than articles with less influence on

553 the public. This echoes Ioannidis (14) who noted that attention-grabbing findings are more likely
554 to be false.

555 Letters of concern and efforts at replication should both interest editors when the article
556 being replicated was published in their journal. Letters of concern usually trigger introspection
557 by editors and publishers (1, 3, 12, 24, 25, 29, 55, 56). So too should efforts at replication,
558 successful or failed. The response of editors to both is of the utmost importance (Case 1).
559 Similarly, journals other than the one in which a putative vampire article was published also play
560 a more important role than has been explicit in the meta research literature. As both of our cases
561 illustrate, third-party replicators published in other journals than the original vampire article.
562 Extremely slow, disrespectful, or disinterested responses from editors can make authors of
563 rebuttals more cautious and unlikely to submit a critique (or any article) to that same journal.
564 Therefore, editors may face novel challenges with rebuttals. For one, they may find themselves
565 considering articles in other journals that failed to replicate findings published in their own
566 journal. Therefore, editors cannot only rely on their accustomed list of peer reviewers but should
567 consider peer reviewers from journals publishing commentaries, critiques, and replication efforts
568 of the original article. We suspect this is a novel recommendation and we welcome editorial
569 feedback. We are aware of an editorial justification for not retracting articles no matter how
570 flawed. One rationalization may be (as in Case 1) that debates should play out as multiple peer-
571 reviewed point and counterpoint response. We pointed out in Case 1 that such debates may be
572 inappropriate when fundamental problems in the original article come to light. Although
573 conspicuous labels of editorial concern are gaining currency for online publication, print versions
574 of the original article will lack such labels. Furthermore, secondary affirmative citations may
575 proliferate and leave the label behind, analogous to laundering ill-gotten income (31).

Moreover, we recommend editors of research articles do more than pay attention to rebuttals and expressions of concern; editors seem to us to have a duty to call readers' attention to ongoing debate even if it occurs in other journals. Therefore, we endorse the practice of affixing editorial cautions to articles that have accumulated peer-reviewed rebuttals or expressions of concern. Although we grant readers might ignore flags of concern and these can disappear into the clutter at the top of articles, we recommend that editors foster more of the scholarly debate that was historically envisioned when the institution of peer review spread and gained prominence as a means of improving research. A flag of editorial concern should point readers to free access to rebuttals of the flagged article. However, when editors become convinced that they have published a vampire article after reading the scholarly debate and seeing affirmative citation, flags of concern are inadequate in our view. By the analogy of vampires, beware the inclination to post warning signs around the coffin rather than a stake through the heart.

Lessons learned from research in the communication sciences support our recommendation that notices should be memorable and make it clear that the journal has lost confidence in the flagged article. Likewise, corrections should focus on the corrected findings loudly and clearly—not the original ones. Retractions in turn should not be covert and kept quiet but trumpeted and editors should throw their weight into correcting the broadcast media or policy-makers who repeat irreproducible findings. In short, the research community needs the support of the publishing community to combat misinformation. We acknowledge much of what we propose above will face enormous obstacles for under-staffed and over-worked editors and peer reviewers. But we emphasize again that irreproducibility MUST be combined with affirmative citations that have broad practical implications before the research community, and

599 their publishers take action. The two jointly necessary conditions should reduce the obstacles to
600 action.

601 *Vampire articles spawn more harmful research.* Ultimately, a vampire article should be
602 retracted. But retraction dynamics are slow (Cases 1 and 2). Therefore, articles that should be
603 corrected or retracted may cause proliferation of misleading findings. Weak or false findings can
604 raise opportunity costs for other researchers who might otherwise collect stronger evidence. If a
605 vampire article escapes detection and spawns additional studies that build upon it, we may find a
606 self-perpetuating and proliferating set of misleading findings. When they cite each other, the
607 resulting patterns have been called citation cartels (59); for another example, see (60) for
608 allegations of financial incentives for disinformation by advocates for trophy hunting. Also,
609 proliferation may thrive when similarly unreliable studies can be produced more rapidly than
610 efforts to falsify them. For example, subsequent articles stating the methods in the vampire
611 article "were followed" rather than specifying methods de novo. Thus, vampire articles may
612 spawn additional weak studies more quickly than they can be retracted.

613 When a foundational article is irreproducible and yet it spawns more articles that are also
614 not disproven, we may see research progress diverted or halted until a paradigm shift occurs,
615 e.g., the 'myth of balance of nature' has taken over a century of challenges and still persists (61).
616 Therefore, the metaphor of the vampire spawning further undead seems to us to be a helpful
617 heuristic for encouraging swift selfless action by qualified experts in the given subfield.

618 *Detecting vampires requires more than a mirror.* Detecting irreproducible findings is not
619 easy and the tools for doing it are not at hand in everyday life. Our topic may be unfamiliar to
620 some readers who might imagine that retractions succeed when the evidence is clear that a
621 finding is unreliable. Lay audiences can easily find themselves at sea (see the example of wolf-

poaching preceding our Case 1). We recommend policy-makers that are citing peer-reviewed research underpinning their policy interventions should ask trusted scientific advisors to confirm the following for the sources they cite: has it been retracted, corrected, or marked with a notice of concern? To answer this, the Retraction Watch database (71) may help (e.g., Figure 2). Finding commentaries and replication efforts may succeed simply by online text search of the original authors' names or the title of their original articles.

If rebuttals or failed replication efforts are found in such searches, close reading of the original, comments, and replication efforts would be a reasonable next step before citing the original for policy formulation. Much of that close review would require qualified researchers without competing interests to assess whether to cite the original article affirmatively. This sounds like an overwhelming task for most decision-makers with their skeletal or non-existent staff of trusted, qualified researchers without competing interests. For some cases, civil-sector or government watchdog agencies will be needed. This raises questions about the possibility of legal enforcement against persistent proliferating vampire articles.

Politicization of science and the weaponization of misinformation

When we refer to politicization of research, we are not discussing differences in policy preferences related to research topics. Rather, politicization of science happens when actors strategically exploit uncertainties inherent to scientific knowledge to serve a policy goal or agenda (62). This may involve exaggerating uncertainty in the research evidence to avoid taking action on an issue, e.g., (63), or more direct manipulation. For example, political actors challenging climate science obfuscated and misrepresented scientific research by highlighting contrarian researchers and funding claims that were not peer-reviewed, among others (64). Policymakers may be unaware of the weaknesses of the original study, but continued use of

vampire articles may also be part of a broader strategy to cherry pick research that can be interpreted to support predetermined policy goals. The uncertainty inherent to the process of scholarly inquiry presents vulnerability to politicization and misperceptions.

Scientific hypotheses are not proven. Rather, hypotheses are positions supported by a body of evidence which, particularly in the case of rapidly moving research areas, may be quickly found wanting (65). In this context, it can be difficult for non-specialists to weigh the literature and discern the quality of studies they find, or to trust apparently changing findings and recommendations. While disclosing some forms of uncertainty is well-received among the general public (e.g., technical estimates, study limitations), disagreement among researchers is a form of uncertainty that can provoke distrust (66). Contentious disagreement between scholars can cast doubt on the whole area of research, not only an irreproducible study (67). This may be exacerbated when refuting the findings of a study without a clear alternative explanation. Individuals form mental models of events in their heads that illustrate causal links between events; telling others that a core element of the model is incorrect, without replacing it, may backfire. For example, telling parents that vaccines do not cause autism, while the causes of autism remain uncertain evokes psychological discomfort (34). Rebuttals of vampire articles face these uphill challenges. Disagreement between research teams may be perceived as an indicator of poor competence in the field and failing to provide a clear alternative explanation or conclusion may leave readers confused and likely to continue to rely on the vampire article. Legal protections against politicized vampire articles might be needed.

For example, in the U.S., some government agencies and two federal statutes could in principle mitigate the problems posed by vampire articles. For one, the federal Data Quality Act (DQA) should encourage agencies to remove flawed research from policy documents. However,

668 the DQA is not subject to judicial review and requires agency self-policing. Most agencies'
669 research integrity policies require review by the agency following a petition by an outside group,
670 e.g., (68). Worse in our context, the National Environmental Policy Act is exempted from the
671 DQA, which would likely immunize Environmental Impact Statements (EIS) or assessments
672 from DQA challenges adding to the long-standing criticism that EIS typically make meaningless
673 claims (69). Moreover, even if an EIS were to be rescinded (e.g., by court order), the decision
674 would affect only that jurisdiction, so poor science in the EIS might live on in other jurisdictions
675 (Case 1). A different approach is seen in the U.S. Federal False Claims Act (FCA). Institutions
676 citing suspect research when seeking or reporting on federal funding may be subject to US
677 Department of Justice procedures under the FCA. (70). FCA penalties depend on finding
678 research misconduct of a deliberate, or at least reckless character. Research misconduct is
679 defined in federal law (42 CFR 93 Public Health Service Policies on Research Misconduct; Final
680 Rule p. 103), as fabrication, falsification, or plagiarism. FCA procedures may force an institution
681 to freeze the federal funds of the research team in question, during an inquiry. Whistle-blowers
682 have been responsible for 70% of successful FCA complaints (70). Successful FCA challenges
683 force institutions to return federal funding from past or current grants awarded to those convicted
684 of research misconduct. To our knowledge, federal agencies or state or tribal agencies receiving
685 federal funds are not subject to FCA enforcement. That loophole may facilitate politicization of
686 agency research. Although our examples above focus on US policy, we believe that the issues
687 highlighted here are relevant for the broader research community and most governments around
688 the globe. That brings us to other limitations of our work and the concept of vampire articles.

689 **Limitations**

We do not expect the scholarly literature ever to be free of errors. We are concerned here only with failures to replicate studies that continue to be cited affirmatively. Nor do we expect the research community to move with alacrity to expand efforts at replication. Although we value a scientific literature purged of confirmed errors, we acknowledge the ideal is unlikely to be achieved for most older findings. We anticipate more recent articles and those coming in the future will be subject to more scrutiny. We anticipate that fields advancing quickly will present corrections in the course of literature reviews, following collection of new data and reanalyses prove older results erroneous rather than efforts at correction or retraction of published work. Nevertheless, in our disciplines, and we predict in many others, findings that are difficult to repeat, because of lack of funds or inaccessible methods to replicate them, may persist in the literature for long periods along with their harms. In the latter situations, we still urge efforts to replicate but we acknowledge that we have no solution for findings that are contingent on unique events (historicity) or on unique events or subjects. Indeed, we have little to offer to improve purely qualitative research that reports on individual uniqueness. Finally, a major limitation of our approach is that our recommendations impose an added burden on editors and peer reviewers to scrutinize studies that garner affirmative citations outside the scholarly literature. This seems a nearly impossible task on their own. But aided by the qualified researchers specialized in the subfield of an article getting affirmative citations in public discourse and aided by post-publication reviewers and efforts at replication of major findings, it seems a more tractable challenge.

Conclusions

The ethics of the conduct of research and cleaning of the of its literature are important because the public pays for much research and is urgently in need of reliable and valid

information for policy. Without scholars vigilant for irreproducibility followed by swift corrections to the literature, we fear that public trust in science will continue to erode.

Authors, editors, publishers, and third-party replicators are essential partners in protecting public trust in science and perhaps other scholarly fields. That means they must view each other as allies in weakening the effects of errors in research. Public funding to fortify efforts at replication might go a long way to reducing the harms done by erroneous findings. Further, the second condition of vampire articles, that of affirmative citation, depends in part on better education about research uncertainty and strength of inference among decision-makers. Vampire articles make the already elusive progress of science and other research fields harder to achieve.

List of abbreviations

W&C 1999: Wagner & Conover 1999 (ref. 37)

Database = Retraction Watch database (ref. 71)

COPE = Committee on Publication Ethics (ref. 25)

USDA-APHIS-WS = United States Department of Agriculture Animal and Plant Health Indirection Services – Wildlife Services (ref. 73).

Declarations

Ethics: Not applicable

Consent: Not applicable

Data availability: We provided all the information on case 1 in supplementary information files.

Competing interests: The authors declare no competing interests but present their CVs for readers to judge for themselves (see http://faculty.nelson.wisc.edu/treves/CCC.php accessed 24 October 2025.

Funding: The authors declare that they perceive no potentially competing financial interests and present recent funding histories and CVs here (see http://faculty.nelson.wisc.edu/treves/CCC.php accessed 22 October 2025) for readers to see for themselves.

Author contributions: AT conceptualized, collected case study information, wrote, revised, and submitted. SC contributed communications science literature review, helped write and revise. MK helped conceptualize, write, and revise.

Acknowledgments: We thank C.T. Darimont and F. J. Santiago-Ávila for comments on early versions and critical feedback. We thank A. Fisher for preparing Figure 2.

Author information: AT teaches and conducts research on reproducibility, open science, and research integrity. SC teaches and conducts research on misinformation and disinformation.

Additional file 1: SI 1 Table 1. Irreconcilable methods for study of aerial shooting at coyotes, described in A W&C1999 (37); B: Wagner et al. 2016 (41) relating to the same method; and C: we explain why logical inconsistencies and omissions make W&C1999 impossible to replicate.

Additional file 2: SI 2 all combined.pdf relates to the W&C1999 case 1.

References

1. Binning SA, Jutfelt F, Sundin J. Exorcise citations to the 'living dead' from the literature. Nature. 2018;558:981.

2. Lab Times E. 'Zombie articles…'. Lab Times. 2012;7:3.

3. Brainard J. 'Zombie papers' just won't die. Retracted papers by notorious fraudster still cited years later. Science. 2022;377(6601):11-2.

4. Krauss A. Why all randomised controlled trials produce biased results. Ann Med. 2018;50(4):312-22.

5. Goodman S, Fanelli D, Ioannidis J. What does research reproducibility mean? Science Translational Medicine. 2016;8(341):ps12 DOI: 0.1126/scitranslmed.aaf5027.

6. Iqbal SA, Wallach JD, Khoury MJ, Schully SD, Ioannidis JPA. Reproducible Research Practices and Transparency across the Biomedical Literature. PLoS Biol. 2016;14(1):e1002333. 10.1371/journal.pbio.1002333.

7. Sanders J, Jon Blundy, Anne Donaldson, Steve Brown, Rob Ivison, Miles Padgett, et al. Transparency and openness in science. Royal Society Open Science. 2017;4:e4160979.

8. Base Pairs. The cancer answer that wasn't [Internet]. Cold Spring Harbor Laboratory; 2018 March 15, 2018. Podcast. Available from: https://www.cshl.edu/podcasts/base-pairs-episode-14-cancer-answer-wasnt/.

9. Gernsbacher MA. Three ways to make replication mainstream. Behavioral and Brain Sciences. 2018;41:e129. https://doi.org/10.1017/S0140525X1800064X.

10. Webster MM, Rutz C. How STRANGE are your study animals? Nature. 2020;582,:337-40. 10.1038/d41586-020-01751-5.

11. Oliynyk RT. Human-caused wolf mortality persists for years after discontinuation of hunting. Scientific Reports. 2023; 13: 11084. |https://doi.org/10.1038/s41598-023-38148-z.

12. Allison DB, Brown AW, George BJ, Kaiser KA. Reproducibility: A tragedy of errors. Nature. 2016;530:27-9. doi:10.1038/530027a.

13. Open Science Collaboration. Estimating the Reproducibility of Psychological Science. https://osf.io/ezcuj/2015. p. https://osf.io/ezcuj/. 10.17605/OSF.IO/EZCUJ.

14. Ioannidis JP. Why most published research findings are false. PLOS Medicine. 2005;2(8):e124.

779 15. Pineda Guerrero AA. Human-Carnivore Coexistence: The Functional and Perceived
780 Effectiveness of Solar Lights and Attitudes Toward Jaguars and Pumas in Colombia: University
781 of Wisconsin-Madison; 2023.
782 16. Treves A, Fergus AR, Hermanstorfer SJ, Louchouarn NX, Ohrens O, Pineda Guerrero
783 AA. Gold-standard experiments to deter predators from attacking farm animals. Animal
784 Frontiers. 2024;14(1):40-52. 10.1093/af/vfad072.
785 17. Hall K, Fleming PA. In the spotlight: can lights be used to mitigate fox predation in a
786 free-range piggery? Appl Anim Behav Sci. 2021;2:105420.
787 https://doi.org/10.1016/j.applanim.2021.105420.
788 18. Ohrens O, Bonacic C, Treves A. Non-lethal defense of livestock against predators:
789 Flashing lights deter puma attacks in Chile. Front Ecol Environ. 2019;17(1):32-8.
790 10.1002/fee.1952.
791 19. Santiago-Ávila FJ, Treves A. Poaching of protected wolves fluctuated seasonally and
792 with non-wolf hunting. Scientific Reports. 2022;12:e1738 10.1038/s41598-022-05679-w.
793 20. Louchouarn, N.X. Don’t judge the roar by its echo: Tests of assumptions, tools and
794 policies for human-carnivore coexistence in North America . PhD thesis, University of
795 Wisconsin-Madison, USA, 2023.
796 21. Chinn S, Hart PS. Effects of consensus messages and political ideology on climate change
797 attitudes: inconsistent findings and the effect of a pretest. Clim Change. 2021;167(3-4):47.
798 10.1007/s10584-021-03200-2.
799 22. Gould SJ. The promise of paleobiology as a nomothetic, evolutionary discipline.
800 Paleobiology 1980;6:96-118.

801 23. Biondi F. Paleoecology grand challenge. Frontiers in Ecology and Evolution. 2014; doi:
802 10.3389/fevo.2014.00050.
803 24. Collaborative Working Group. RePAIR consensus guidelines: Responsibilities of
804 Publishers, Agencies, Institutions, and Researchers in protecting the integrity of the research
805 record. Research Integrity and Peer Review. 2018;3:15. 10.1186/s41073-018-0055-1.
806 25. Committee on Publication Ethics, COPE Guidelines: Retraction Guidelines. Promoting
807 integrity in scholarly research and its publication2019. https://doi.org/10.24318/cope.2019.1.4
808 accessed 22 October 2025
809 26. de Haas B. What my retraction taught me. Nature. 2021;589:331.
810 27. Loss-of-Confidence Project [Internet]. 2025. Available from:
811 https://lossofconfidencecom.wordpress.com/.
812 28. Antonelli A, Perrigo A. The pitfalls of taking science to the public. Science.
813 2018;359(6373):283.
814 29. Marcus A, Oransky I. The data thugs. Science. 2018;359(6377):730-2.
815 30. Heathers J. The Protraction of Retraction: A criminology paper has problems. What
816 should be done? Medium [Internet]2019. Available from:
817 https://jamesheathers.medium.com/the-protraction-of-retraction-8a00ffc9b210.
818 31. Suelzer EM, Deal J, Hanus KL, Ruggeri B, Sieracki R, Witkowski E. Assessment of
819 Citations of the Retracted Article by Wakefield et al With Fraudulent Claims of an Association
820 Between Vaccination and Autism. JAMA Network Open. 2019;2(11).
821 10.1001/jamanetworkopen.2019.15552.
822 32. Oreskes N. Why Trust Science? Princeton, NJ: Princeton University Press; 2019.

823 33. Bode L, Vraga E. The Swiss cheese model for mitigating online misinformation. Bulletin
824 of the Atomic Scientists. 2021;77(3):129–33. https://doi.org/10.1080/00963402.2021.1912170.
825 34. Walter N, Tukachinsky R. A Meta-Analytic Examination of the Continued Influence of
826 Misinformation in the Face of Correction: How Powerful Is It, Why Does It Happen, and How to
827 Stop It? Communication Research. 2020;47(2): 155–77.
828 https://doi.org/10.1177/0093650219854600.
829 35. Chinn S, Hart PS. Can't You All Just Get Along? Effects of Scientific Disagreement and
830 Incivility on Attention to and Trust in Science. Science Communication. 2021;44(1):108-29.
831 10.1177/10755470211054446.
832 36. Banobi JA, Branch TA, Hilborn R. Do rebuttals affect future science? Ecosphere.
833 2011;2(3). 10.1890/es10-00142.1.
834 37. Wagner KK, Conover MR. Effect of preventive coyote hunting on sheep losses to coyote
835 predation. J Wildl Manage. 1999;63:600-12.
836 38. Mitchell BR, Jaeger MM, Barrett RH. Coyote depredation management: current methods
837 and research needs. Wildl Soc Bull. 2004;32(4):1209-18.
838 39. Treves A, Krofel M, McManus J. Predator control should not be a shot in the dark. Front
839 Ecol Environ. 2016;14:380-8.
840 40. Western Watersheds Project et al. v USDA Wildlife Services. U.S. District Court for the
841 District of Idaho 1:17-cv-00206-BLW Doc 22-3; 2018.
842 41. Wagner K, Graves G, Hall TW, Hebert SW, McDonald EW, Gustad K, et al. Final
843 environmental [sic] assessment predator damage management in Idaho. United States
844 Department of Agriculture Animal and Plant Health Inspection Service Wildlife Services, Idaho
845 Department of Fish and Game; 2016 November 2016.

846 42. Krausman PR. Improving the Journal of Wildlife Management: a response to the
847 perspectives of Johnson et al. (2021) and Gould et al. (2021). The Journal of Wildlife
848 Management. 2022;86(2):e22167. https://doi.org/10.1002/jwmg.22167.
849 43. Worm B, Barbier EB, Beaumont N, Duffy JE, Folke C, Halpern BS, et al. Impacts of
850 Biodiversity Loss on Ocean Ecosystem Services. Science. 2006;314(3):787-90.
851 44. Lewandowsky S, Ecker UKH, Seifert CM, Schwarz N, Cook J. Misinformation and its
852 correction: Continued influence and successful debiasing. Psychol Sci Publ Int. 2012;13(3):106–
853 31. https://doi.org/10.1177/1529100612451018.
854 45. Fazio LK, Brashier NM, Payne BK, Marsh EJ. (2015). Knowledge does not protect
855 against illusory truth. Journal of Experimental Psychology: General. 2015;144(5):993–1002.
856 https://doi.org/10.1037/xge0000098.
857 46. Vraga EK, Kim SC, Cook J, Bode L. Testing the Effectiveness of Correction Placement
858 and Type on Instagram. The International Journal of Press/Politics, . 2020;25(4):632–52.
859 https://doi.org/10.1177/1940161220919082.
860 47. Shulman HC, Dixon GN, Bullock OM, Colón Amill D. The Effects of Jargon on
861 Processing Fluency, Self-Perceptions, and Scientific Engagement. Journal of Language and
862 Social Psychology. 2020; 39(5-6):579–97. https://doi.org/10.1177/0261927X20902177.
863 48. Brandeis LD. Other People's Money. http://louisville.edu/law/library/special-
864 collections/the-louis-d.-brandeis-collection/other-peoples-money-by-louis-d.-brandeis: Btandeis
865 School of Law; 1914.
866 49. Gunsalus CK, Robinson AD. Nine pitfalls of research misconduct. Nature. 2018;557:297-
867 9.

868 50. Mejlgaard N, Bouter L, M., Gaskell G, Kavouras P, Allum N, Bendtsen A, et al.
869 Research integrity: nine ways to move from talk to walk. Nature. 2020;586:358-60.
870 10.1038/d41586-020-02847-8.
871 51. Baker M, Brandon K. Is there a reproducibility crisis? A Nature survey lifts the lid on
872 how researchers view the 'crisis' rocking science and what they think will help. Nature.
873 2016;533:452-4.
874 52. Dance A. Peer review needs a radical rethink. Nature 2023;614:581-3.
875 53. Mukherjee S. The Emperor of All Maladies: A Biography of Cancer. New York:
876 Scribner; 2010.
877 54. Gawande A. The Mistrust of Science. New Yorker. 2016 10 JUNE 2016.
878 55. Seim DL. The Butter-Margarine Controversy and "Two Cultures" at Iowa State College.
879 Annals of Iowa. 2008;67:1-50.
880 56. Johns DM, Oppenheimer GM. Was there ever really a "sugar conspiracy"? Science.
881 2018;359(6377):747-50.
882 57. Mukherjee S. The Gene: An Intimate History. New York: Simon & Schuster; 2016.
883 58. PLoS Medicine Editors. Making sense of non-financial competing interests. PLoS Med.
884 2008;5(9):e199. 10.1371/journal.pmed.0050199.
885 59. Franck G. Scientific Communication--A Vanity Fair? Science. 1999;286(5437):53-5.
886 10.1126/science.286.5437.
887 60. Koot S, Hebinck P, Sullivan S. Science for Success—A Conflict of Interest? Researcher
888 Position and Reflexivity in Socio-Ecological Research for CBNRM in Namibia. Soc Nat Resour.
889 2020;36(5):1-18. 10.1080/08941920.2020.1762953.

890 61. Kricher J. The Balance of Nature: Ecology's Enduring Myth. Princeton, NJ: Princeton
891 University Press; 2009.
892 62. Bolsen T, Druckman JN. Counteracting the Politicization of Science. Journal of
893 Communication. 2015;65(5):745–69. https://doi.org/10.1111/jcom.12171.
894 63. Jasen P. Breast Cancer and the Politics of Abortion in the United States. Medical History.
895 2005;49(4):423–44. https://doi.org/10.1017/S0025727300009145.
896 64. McCright AM, Dunlap RE. Anti-reflexivity. Theory, Culture & Society. 2010;27(2-
897 3):100-33.
898 65. Scheufele D, Krause NM, Freiling I, Brossard D. How not to lose the COVID-19
899 communication war. Slate. 2020. https://slate.com/technology/2020/04/covid19-misinformation-
900 science-communication.html.
901 66. Gustafson A, Rice RE. A review of the effects of uncertainty in public science
902 communication. Public Understanding of Science. 2020; 29(6): 614–33.
903 https://doi.org/10.1177/0963662520942122.
904 67. Chinn S, Sol Hart P. Can't You All Just Get Along? Effects of Scientific Disagreement
905 and Incivility on Attention to and Trust in Science. Science Communication.
906 2021;107554702110544:0544. https://doi.org/10.1177/10755470211054446.
907 68. Department of the Interior. Part 305: Departmental Science Efforts Chapter 3: Integrity of
908 Scientific and Scholarly Activities. In: Office of the Deputy Secretary, editor. Washington, DC:
909 Department of the Interior 2014. p. 1-17.
910 69. Schindler DW. The Impact Statement Boondoggle. Science. 1976;192(4239):509.
911 70. Young TK, Tuteur M. Scientific Research Misconduct vs. Fraud: How to Tell the
912 Difference. Health Care Law Today. 2018;July

913 2018:https://www.healthcarelawtoday.com/20218/05/07/scientific-research-misconduct-vs-
914 fraud-how-to-tell-the-ifference/.
915 71. RW database, Retraction Watch database,
916 https://retractiondatabase.org/RetractionSearch.aspx? accessed 22 October 2025.
917 72. U.S. Department of Agriculture Animal and Plant Health Inspection Services – Wildlife
918 Services (https://www.aphis.usda.gov/sites/default/files/5-aircraft-use-peer-reveiwed.pdf,
919 accessed 12 October 2025).
920
921