The Best Paper You'll Read Today: Media Biases and the Public Understanding of Science

Aydin Mohseni

Department of Philosophy
University of Pennsylvania

Cailin O'Connor

Department of Logic and Philosophy of Science
University of California, Irvine

James Owen Weatherall

Department of Logic and Philosophy of Science
University of California, Irvine

Abstract

Scientific curation, where scientific evidence is selected and shared, is essential to public belief formation about science. Yet common curation practices can distort the body of evidence the public sees. Focusing on science journalism, we employ computational models to investigate how such distortions influence public belief. We consider these effects for agents with and without confirmation bias. We find that standard journalistic practices can lead to significant distortions in public belief; that pre-existing errors in public belief can drive further distortions in reporting; that practices that appear relatively unobjectionable can produce serious epistemic harm; and that, in some cases, common curation practices related to fairness and extreme reporting can lead to polarization.

1. Introduction

The total body of scientific evidence available at a time, as gathered, analyzed, and synthesized by scientists themselves, is the best basis for scientific knowledge. But public beliefs about science—and, indeed, the beliefs of scientists about matters outside their expertise—are rarely formed on the basis of that total

Email addresses: amohseni@upenn.edu (Aydin Mohseni), cailino@uci.edu (Cailin O'Connor), weatherj@uci.edu (James Owen Weatherall)

body of evidence. Instead, the public gains access to scientific evidence via *curators*—individuals, organizations, or, increasingly, algorithms that select research outputs to share, popularize, review, or amplify. Examples of curators include science journalists, science popularizers, science museum curators, textbook writers, policy advisors, industrial propagandists, and social media platforms, among others.

Curation of some sort or another is inescapable in the spread and sharing of scientific knowledge. No one can review the entire body of scientific evidence across every field, so we all depend on curators to select and share relevant parts of it. How curators go about this process, though, crucially shapes how the public comes to understand the state of knowledge in a given domain. Sometimes curators have good epistemic aims. Textbook writers, for instance, typically seek to help students form accurate beliefs. They do this by emphasizing particularly clear work that supports the best beliefs of the scientific community; and neglecting confusing or countervailing evidence. Though this practice systematically distorts the body of evidence presented, we expect it to typically lead readers to good beliefs.

Other curators are principally motivated by non-epistemic goals. Industrial propagandists, for instance, often use scientific evidence to advance their own economic or political interests. Historical evidence reveals that the selection methods used by such actors are often epistemically harmful [33, 25, 38].

Here we consider a type of curation that is more ambiguous both in intent and epistemic effects—science journalism. Journalists are bound by professional norms related to balance, fairness, and truth-telling, which are intended to yield good epistemic effects. Moreover, many journalists are personally motivated to improve reader beliefs by reporting accurately. But media companies and journalists also have a financial motivation to maximize readership, advertising revenue, and so on. These latter goals are, at best, independent of epistemic aims, and in some cases they conflict with them. Something similar can be said about social media sites, which likewise have financial motivations to maximize engagement (even if they also benefit from a reputation for combating misinformation).

To what extent are the norms governing acceptable curation in various contexts sufficient to ensure good epistemic outcomes? How can one assess this? The goal of this paper is to present a simple model to investigate how curation practices may shape public belief about matters of fact. We will focus on journalism—

especially, science journalism—but the model we present may be readily extended to evaluate the epistemic effects of other curators, such as social media algorithms.

Our computational models use probability distributions to represent 1) evidence from the world, 2) reporting on that evidence, and 3) public belief in light of this reporting. We assume that real-world events are sampled from some fixed, unknown distribution. Journalists select some of these events to report, thereby producing a new distribution of reported events. Members of the public update their beliefs in light of this reported evidence. We can then compare our agents' posterior beliefs about the world with the actual distribution to see how they are shaped by reporting practices. While computational models have been widely used to study social phenomena, there is relatively little work using these techniques to study the effects of journalistic practice on public belief.[1] As we will show, modelling of this sort is a useful tool in investigating how public belief might be affected by media. In particular, our models are useful in demonstrating how and when otherwise ideal learners can be led to inaccurate beliefs by curation.

We focus on three curation practices. One is *hyperbole* where curators exaggerate or sensationalize claims to garner attention. The second is *extremity bias* which involves cherry-picking surprising, novel, or extreme events to report. The last is *fair reporting*—the practice of presenting equally weighted evidence from two sides of some issue irrespective of where the preponderance of evidence lies. In addition, we consider how these practices interact with confirmation bias, the widely documented tendency for individuals to accept evidence that supports their beliefs and reject evidence that refutes them [29].

These models yield a series of take-aways. First, all three biases can lead to distorted beliefs, even in cases where learners would otherwise develop an accurate picture of the world. Second, current misunderstandings in public belief can drive these journalistic distortions, thus creating negative feedback effects. This happens when current, inaccurate beliefs shape what is considered extreme, or fair, and thus drive selection of particularly misleading reporting. Third, in the presence of confirmation bias, both extremity bias and fairness are very harmful to the beliefs of media consumers. This is somewhat surprising because both prac-

[1] Some scholars have used economic modelling to address trends in journalism that stem from incentive structures [18, 26, 13]. To our knowledge, though, mathematical and computational models have not previously been developed to explore how these trends affect public belief. The closest predecessors to the present models of which we are aware are the models of journalism discussed in O'Connor and Weatherall [31, Ch. 4] and Weatherall et al. [38].

tices involve selecting true events to report, which is often considered relatively innocuous compared to, say, lying or exaggeration. Last, both fair reporting and extremity bias can lead to polarization in the presence of confirmation bias.

Our paper will proceed as follows. In section 2 we address previous work on journalistic norms and practices, focusing on those we discuss here. In section 3 we describe the basic workings of our model, and then present some initial results in section 4. In particular, we explain how characteristic reporting practices can lead to systematic errors in public belief. In section 5 we consider what happens when individuals exhibiting confirmation bias encounter skewed distributions of reported events. As we show, this combination can lead to polarization. Section 6 concludes by discussing what we can learn from these models, and also their relevance to social media curation. We argue, in particular, that not enough attention is paid to potential epistemic harms from curation, rather than disinformation, on social media platforms.

2. Journalistic Practices

Journalists face a number of incentives and pressures that shape reporting practices [2]. Most notably, journalists are incentivized to draw readership and get attention. This pressure, combined with human biases towards novel, surprising, or extreme information, creates incentives for journalists to write pieces that seem novel, exciting, and extreme. As Edwin E. Slosson, the first editor of Science Service, once put it, "It is not the rule, but the exception to the rule that attracts public attention" Nelkin [quoted in 28, p. 18]. Many authors have argued that these incentives influence what journalists choose to report and how they report on it.[2]

New technologies have arguably amplified these effects. Digital media offers editors and authors unprecedented information about what articles (and, especially, headlines) drive traffic to their publications. This fine-grained data permits fine-tuning of article content and presentation that would have been impossible in

[2] For instance, in addressing an increase in character attacks in political reporting, Bennett [2, p. 376] writes, "...journalists face the increasing pressure to construct more dramatic stories that supposedly appeal to audiences who value being entertained over being informed". Likewise, regarding the importance of such incentives to journalistic practice, Hamilton [21, p. 6] writes, "(N)ews emerges not from individuals seeking to improve the functioning of democracy but from readers seeking diversion, reporters forging careers, and owners searching for profits". Such observations are commonplace in the literature.

a pre-digital landscape. Moreover, the costs of creating new media are lower than ever before. Choi and Yang [8] argue that the proliferation of low-quality online news sites deprecates high-quality journalism as competition for user attention increases. Media organizations are strongly incentivized to generate eye-catching material distinct from what their competitors produce [e.g., 34, 24].

In part to counterbalance these incentives, journalists are expected to adhere to a number of professional norms.[3] Some of the most important include norms of truth-telling and accuracy, transparency, accountability to the public, and fairness. Further norms involve providing context for claims made, promoting a civil exchange of views, and minimizing harm in research. In this paper, we consider distortions that arise both due to the incentives just noted and the norms in place to moderate their effects.

The first distortion we consider is *hyperbole*—reporting that exaggerates or extends facts to make them more extreme, and thus more exciting. In discussing Edwin Slosson, Nelkin [28, p. 18] writes, "That is why, he explained, science is usually reported in short paragraphs, ending in "est". 'The fastest or the slowest, the hottest or the coldest, the biggest or the smallest, and in any case, the newest thing in the world'." This sort of writing may violate explicit journalistic norms against distorting facts or context, at least when it is unclear that the superlatives truly apply, but it is clearly commonplace. We might expect that it should lead members of the public to think events in the world are more extreme than the reality.

A second distortion, also driven by incentives for novelty, is *extremity bias*—selectively reporting only events or findings that are novel or sensational. Note that in principle, as long as such material is accurate and reported fairly, a preference for novel reporting is generally compatible with journalistic norms of practice, since it involves no disortion of facts. But, like hyperbole, it may nonetheless lead readers to perceive extreme events as more common than they are. For instance, reporting over the prevalence of mad cow disease, SARS, and bird flu has been criticized for this reason [18].

We will consider the effects of *fairness* norms. For much of the 20th century,

[3] The Society for Professional Journalists has an established Code of Ethics revised several times in the last century [32, 36]. Muller [27] has argued that new media technologies require a new code of ethics for journalists, though it is not clear that his proposals have been widely accepted in practice.

these norms were implemented in the US via the Fairness Doctrine which required broadcast license holders to cover all sides of a controversial issue. In practice, this often translated to equal air time for different sides of a debate. Bennett [2, p. 375], for instance, points out that, "...the journalistic reliance on all (official) sides of a story is perceived as part of the professional obligation to produce authoritative information". Simon et al. [35, p. 428] make explicit the idea that equal space is equivalent to fairness, "Balance refers to the relative amount of coverage devoted to a particular side of a story. The more one-sided the story, the less balanced it is".[4] Clearly the fairness doctrine still impacts fair reporting practices today.

One might expect journalistic norms, such as fairness, to promote accurate public belief. But part of the reason we focus on fairness here is that it has been widely criticized in the context of reporting on anthropogenic climate change [20, 5, 17]. For decades mainstream journalists reported on the climate "debate" as if it involved two opposing sides with similar evidential grounds. This led to what has subsequently been called "false balance".[5] False balance is especially worrying in the context of science reporting, where in many cases the side with less evidence is also the side that is wrong. Thus in scientific cases fairness systematically overweights reporting on false claims, and arguably harms public belief [38].[6,7]

The three practices we describe here are far from exhaustive. There are many other ways in which traditional media, new media, and social media are responsive to the sorts of incentives and norms we have described.[8] We leave the question

[4] To give further support to these claims, Nelkin [28, p. 19], in discussing fairness in science journalism, writes, "Reporters try to maintain balance by quoting scientific sources representing opposing sides of a controversy. And Eschelman [17] reports that as of 2011, "USA Today...as a matter of policy require[d] that an editorial on a "controversial" topic be paired with an editorial arguing in opposition".

[5] Scrutiny of this particular case has shown that it was, at least in part, a result of a targeted strategy by industrial interests related to oil and gas [4, 33].

[6] Empirical work has shown that even in high prestige papers coverage has over-emphasized the viability of climate denialism [30, 4, 5]. Others have argued that this has indeed influenced public opinion [12].

[7] Relatedly, Nelkin [28] writes that, "Ironically this notion of objectivity is meaningless in the scientific community, where the values of 'fairness', 'balance,' or 'equal time' are not relevant to the understanding of nature, where standards of objectivity require, not balance, but empirical verification of opposing views" (20).

[8] For example, several recent studies have shown that emotional content is more widely propagated on social media, particularly for moralized topics [6, 9]. This suggests that journalists may be incentivized to introduce moral and emotional content in their stories, and that social media platforms are incentivized to promote that content.

of how to model such practices to future work, and focus here on just these three prominent, and arguably impactful, curation practices.

3. The Model

We now present a computational model meant to capture how journalistic practices shape public uptake of information. The basic idea is that events of a given sort—say, daily high temperatures in Phoenix, the maximal windspeed of North Atlantic hurricanes, the average interaction energies of particle accelerators, or the yearly population numbers of Arctic terns—occur sequentially with various values or magnitudes. The values associated with these events reflect some true relative frequencies or objective chances. Journalists convey information about these events by reporting on the value or magnitude of some, but not necessarily all, of the events that occur. Which events they report is determined by the journalistic practices they adopt. For instance, journalists may report on the temperature in Phoenix only when it exceeds some value. Science reporters may only report on Arctic terns when their population explodes or crashes.

Agents in our models—consumers of media—seek to learn about the true relative frequencies of these events' values or magnitudes by reading what journalists write. These agents begin with some prior beliefs about the class of events under consideration and update those beliefs in light of reporting. We then study how accurate agents' posterior beliefs are.

In more detail, our model begins with an event space, representing the range of possible values characterizing some class of worldly occurrences. We assume this space has the structure of the real line, endowed with its standard norm.[9] We assume that the class of worldly occurrences under consideration may be treated as a random variable, with some fixed (univariate) distribution D_A characterizing the likelihood of event-values occurring; occurrent events are understood as a sequence of draws from this random variable. We call the distribution D_A the *actual distribution*, since it models the relative frequency with which different event values actually occur in the world.

Throughout this paper, we assume that D_A is a normal distribution, with mean μ_A and variance σ_A^2.[10] We will allow μ_A to vary in what follows, but with-

[9] Though the detail makes little difference in what follows, we implicitly assume Lebesgue measurable sets form the sigma algebra of subsets to which probabilities may be assigned.

[10] Observe that the assumptions that D_A is normal (and the event space is $\mathbb{R}$) imply that

out loss of generality, we fix $\sigma_A^2 = 1$. Fixing the variance amounts to adopting conventional units in the model; the value of the variance makes no difference to the interpretation of the model, since we are interested only in discrepancies between the actual variance and agents' beliefs about the variance. But the mean of the distribution will matter for the interpretation of the model what follows. This is because some of the journalistic practices we consider involve comparing samples from D_A to a socially accepted "neutral" value, i.e., some value that is unremarkable or unsurprising. This neutral value may or may not reflect the actual mean μ_A. Throughout, we assume that the value $x = 0$ is the fixed socially accepted neutral value; thus, when $\mu_A \neq 0$, it follows that the actual distribution of events is shifted relative to the societal neutral point.[11] The role this feature of the model plays will become more clear presently.

We model journalists' reporting as a process whereby first nature draws a sample from D_A—some event occurs—and then the journalist chooses how and whether to report that result according to the three practices under consideration here. Thus, the journalist transforms a time series of draws from D_A into another time series of reports. This process could be equivalently described as repeated sampling from a transformed distribution, D_R, the *reported distribution*, which results from applying the journalist's rules. Each of the practices we consider here will result in a reported distribution with different qualitative properties. We discuss the details of these transformations presently; first, though, we discuss learning in this model.

As noted above, we consider a scenario where an agent—representing someone

---

the version of the model we consider strictly speaking applies only to cases where arbitrarily large values, positive or negative, are in principle possible; whereas in many natural cases of application, possible values are bounded from above or below. We do not take this to be a severe limitation, however, and we claim that the qualitative results presented below would extend to other event spaces / distributions. It may be interesting to study this model with a broader class of actual distributions, such as log-normal distributions, which would reflect the boundedness-from-below property. We work with normal distributions here because they simplify updating.

[11] One could imagine more complex models in which this neutral value is itself represented by some distribution representing socially accepted expectations about the events under consideration, where those might differ from both individual agents' expectations and the actual distribution of events. One could also allow the neutral value (or distribution) to be endogenous to the model, in the sense that the neutral value reflects, say, the average of the means that a community of agents, all trying to learn from journalists' reporting, attribute to the class of events. We do not consider these variations on the model here, but return to this issue in the conclusion.

who relies on media reports to learn about the world—attempts to update their beliefs in light of reporting. We model this as a Bayesian learning problem. The agent begins with some prior beliefs, which we represent as a normal distribution over the space of events. This distribution, which is characterized by some mean μ_0 and variance σ_0^2, represents the agent's subjective degrees of belief about the relative frequencies or chances of events with various values occurring. The agent is then presented with a time series of event values generated by the journalist, and after each reported event, the agent updates their beliefs using Bayes' rule, under the assumptions that the actual distribution of events is normally distributed and that reported values are exchangeable. Roughly speaking, agents attempt, using Bayesian methods, to identify a "best-fit" normal distribution to the data presented to them (in light of their priors). Note that agent belief updating is a stochastic process in these models: agents are presented with a particular time series and their posterior beliefs may depend on that order of draws.[12]

To summarize, our model involves four distinct distributions over the space of possible event values:

1. The *actual distribution*, representing the frequency with which different event values occur;
2. The *reported distribution*, representing the frequency with which event values are reported (which results from transforming the actual distribution according to rules fixed by journalistic practice);
3. The *agent's prior distribution*, representing an agent's beliefs about relative frequencies or chances prior to learning; and
4. The *agent's posterior distribution*, representing an agent's beliefs after updating in light of reported events.

We now describe how we implement each of the three journalistic practices within the model. Consider first **hyperbole**. Here we assume the media represents events as more extreme than they really are. Within the model, we suppose

[12] More precisely, agents' belief dynamics are represented using a normal-inverse-gamma distribution, which is the conjugate prior distribution for a normal distribution where both the mean and variance are model parameters (and exchangeability is assumed). Agents' beliefs at a time are characterized by four hyperparameters: α, β, μ, and ν, where α is the (total) number of data points the agent has updated on (initially 0); $\beta = \alpha/2$; μ is the estimated posterior mean; and ν is half the sum of squared deviations. Given a new data point x, agents update by: $\alpha \mapsto \alpha+1$; $\beta \mapsto \beta+1/2$; $\mu \mapsto (\alpha\mu+x)/(\alpha+1)$; $\nu \mapsto \nu + \frac{\alpha}{2(\alpha+1)} \times (x-\mu)^2$. The agent's posterior distribution at a time is a normal distribution characterized by $\mu_p = \mu$ and $\sigma_p^2 = \nu/(\beta-1)$ (for $\beta > 1$).

that if a journalist observes an event with value $x \in \mathbb{R}$, instead of reporting x, they instead report $h \cdot x$, where $h > 1$ is the *hyperbole factor.* Thus, if the time series of actual events is $x_1, x_2, \ldots$, then the transformed time series of reported events will be $h \cdot x_1, h \cdot x_2, \ldots$.[13]

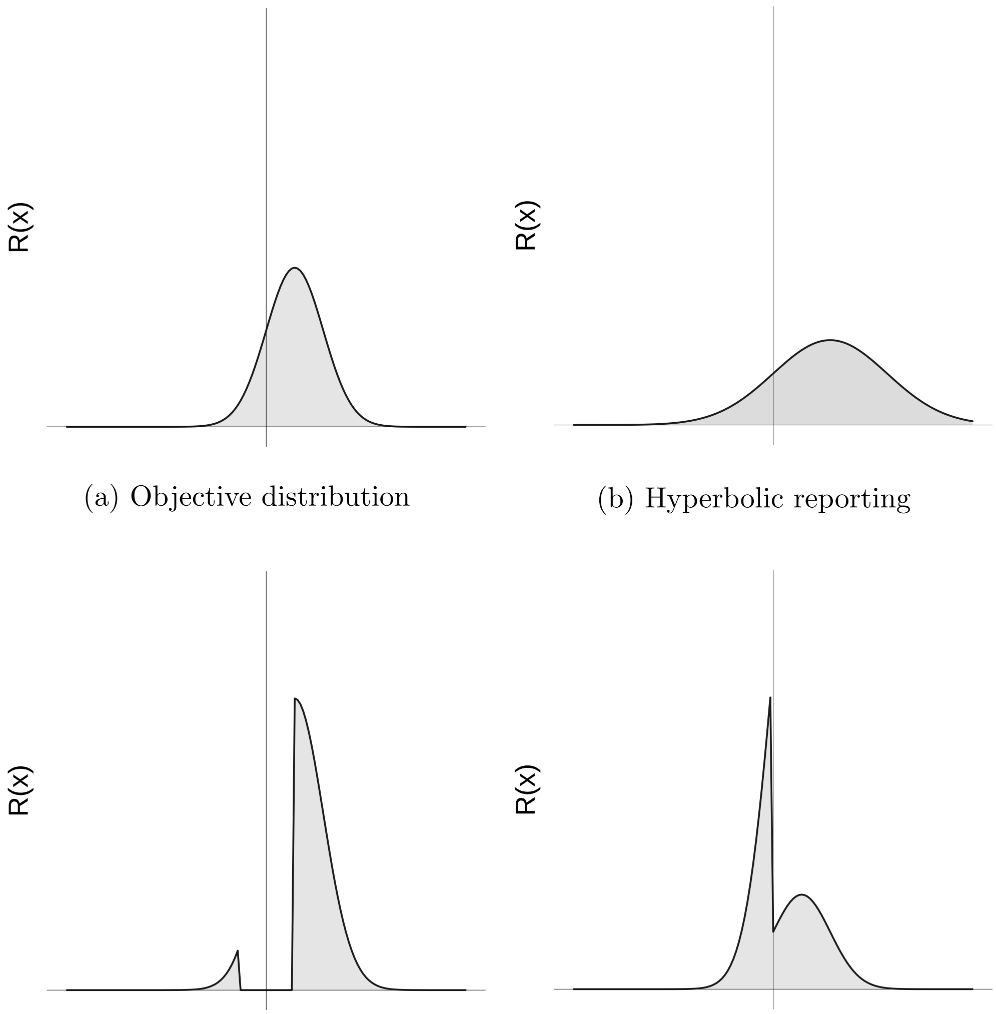
Notice that our model of hyperbole depends on the interpretation of 0 as the societal neutral point. This is because journalists report events with values $x > 0$ as *larger* than they truly are; and events with values $x < 0$ as *smaller* than they truly are (i.e., of greater magnitude, but still negative). Of course in the case where $\mu_A = 0$ the neutral point around which journalists exaggerate is the mean of the actual distribution. But we allow for the possibility that social expectations for neutrality, and therefore the directionality associated with hyperbolization, do not correspond to reality. What journalists and readers consider to be extreme, and thus exciting to report and read, is shaped by social factors.[14] To give an example, there might be an expectation that most fire seasons are relatively extreme compared to what they really are. If so, reporting on an actually extreme fire season may not generate much interest. In order to draw readers, a reporter may be tempted to exaggerate further away from the expectation. This is why hyperbole amounts to journalists drawing from a distribution with variance greater than that of the actual distribution of evidence, and, when $\mu_A \neq 0$, shifting the mean of the reported distribution to a more extreme value (relative to the neutral point). Figure 1b shows qualitatively how hyperbole transforms the distribution D_A.

Now consider **extremity bias** by supposing journalists only report events whose values exceed some threshold value, e (the *extremity parameter*). An extreme event is one whose value has distance at least e from the societal neutral point. Thus, given a time series of actual events, $x_1, x_2, \ldots$, the transformed time series of reported events will consist of precisely those values x_i such that $|x_i| > e$.[15] Figure 1c shows the qualitative results of this transformation. Notice, again, that the neutral point plays an important role in this transformation,

[13] Equivalently, one can think of the reported events as drawn from a normal distribution with mean $\mu_R = h \cdot \mu_A$ and variance $\sigma_R^2 = h^2 \sigma_A^2 = h^2$.

[14] Relatedly, Nelkin [28] points out that the politics, biases, and social influences on journalists play a key role in determining what looks newsworthy.

[15] Equivalently, journalists can be seen as sampling from a random variable who distribution places positive mass only on evidence of magnitude at least e; this distribution is constructed by taking the actual distribution, multiplying it by the characteristic function $\chi_{R/[-e,e]}$, and then renormalizing so that the integral of the distribution over $\mathbb{R}$ equals 1.



(a) Objective distribution (b) Hyperbolic reporting

(c) Extreme reporting (d) Fair reporting

Figure 1: An actual distribution of events D_A (A) with a positive mean $\mu > 0$ contrasted with reported distributions of evidence produced by characteristic distortions: hyperbole (B), extremity bias (C), and fair reporting (D).

since it determines the center of the interval of event values that are not reported. Moreover, the result of the transformation is a bimodal distribution, with both modes located away from the neutral point.

Finally, consider **fairness**. We model this practice by supposing that when journalists report on some class of events, they give equal attention to events whose value is greater than the socially accepted neutral point and ones whose value is less—even if the distribution is not actually symmetric around the neutral point. Journalists give equal weight to "both sides" of an issue by reporting paired event values of opposite sign. If we understand this practice from a time series perspective, journalists will report pairs of results $x_1, x_i, x_{i+1}, x_j \cdots$, where i the least number > 1 such that $(x_i) \neq (x_1)$; j is the least number $> i$ such that $(x_j) \neq (x_{i+1})$; and so on. In other words, they pick an event and then wait to simultaneously report the next event with the opposite sign.[16]

Once again, the qualitative features of the transformation are presented in figure 1d. Observe that yet again, the societal neutral point plays an important role in the transformation. Now, though, the neutral point determines not what counts as exciting or extreme; instead, it determines what counts as a "side" of an issue, by determining the "centrist" position.

Of course, a combination of these distortions can be at play at the same time. We examine them individually and then in interaction. This concludes the basic workings of the model. We will now see what public belief looks like under different scenarios.

4. Results

We use computer simulations to study the model under a range of parameter values, as summarized in Table 1.[17,18] Since agents update sequentially on time

[16] This is approximately equivalent to journalists sampling from a distribution whose probability density function is defined by:

$$g(x) = \begin{cases} f(x)/(2\int_{-\infty}^{0} f(t)dt) & x < 0 \\ f(x)/(2\int_{0}^{\infty} f(t)dt) & x \geq 0 \end{cases}$$

where $f(x)$ is the density function associated with D_A. In other words, one renormalizes the portions of the distribution supported on $x < 0$ and $x \geq 0$ so that half of the total weight lies on each side of the neutral point. We say "approximately equivalent" because the time-series description we have given involves a particular ordering, where values always alternate signs; whereas on the sampling representation, it is merely the case that with probability one, infinite sequences of samples will have equal numbers of positive and negative elements, but with no particular order. We implemented the model using both approaches, and we found qualitatively similar results, though the ordering effects can make a short-term difference in some runs.

[17] For all results reported, two team members independently programmed these models and ran simulations. We verified that results were highly similar, with small differences attributable to small differences in implementation. Code for this model is available at `https://anonymous.4open.science/r/69695669-61ca-4576-acfa-8574a4582da4/` or `https://github.com/jimweatherall/science-curators.git`.

[18] Two notes about our parameters. First, there is no loss of generality in testing only positive means. While negative means make perfect sense in our model, the results should be identical to the ones we gather. Second, agent's initial expectation of variance is realized by initializing agents with $\nu = 1$ and $\beta = 2$, where ν and β are defined as in footnote 12.

Parameter	Role in the Model	Values Studied
	Societal neutral point	0
μ_A	Mean of actual distribution, D_A	$0, .5, 1, 1.5, 2$
σ_A^2	Variance of actual distribution	1
μ_0	Agent's initial expectation of the mean of D_A	$-5, -1, 0, 1, 5$
σ_0^2	Agents initial expectation of the variance of D_A	1
h	Hyperbole parameter	1,1.5,2,2.5,3
e	Extremity parameter	0,.25,.5,.75,1,1.25,1.5
f	Fairness toggle	$\top$, $\bot$

Table 1: Summary of model parameters and values studied.

series data drawn from the reported distribution, the model is stochastic, and early ordering effects can influence agent beliefs. For this reason, we run 1000 simulations for each combination of parameter values, and report averages over these. In each run, we continue the simulation until the agent sees 1000 draws—even in cases where, due to extremity bias or fairness, some data is rejected by the journalist. This way we can be confident the results under different practices reflect the distribution of data as generated by the practices, and not merely the fact that agents in some cases have seen relatively little data.

We report two sorts of results from this model. The first will be qualitative information about features of posterior distributions of belief in light of reporting. The second will correspond to the degree of error in public belief under different conditions. We need a way to quantify this error. At the end of each simulation we measure the (expected) mean and variance of the agent's posterior distribution. We calculate Error in Mean by subtracting the actual mean of the distribution of real world events from the agent's estimated posterior mean. We average this error over all runs of simulation for a parameter setting. Notice, however, that this will incorrectly represent error when some agents overestimate and others underestimate the actual mean. For this reason, we also calculate Squared Error in Mean. We primarily focus on reporting this measure. We calculate Error in Variance and Squared Error in Variance analogously, i.e., by calculating the difference between final estimates of variance and the variance of the actual distribution. We primarily report Error in Variance, rather than Squared Error in Variance, because understanding the direction of this error (i.e., do agents think events from the world are too variable, or too similar) is important.

Before reporting on distortions we observe that, as one should expect, in the absence of any distortative practices, i.e., when the reported distribution coincides with D_A, the agent's posterior distribution closely matches D_A. In other words, agents presented with a representative sample of data are successful in forming beliefs about the actual distribution of events. This is epistemically significant because if we see errors in public belief, we can then attribute them to journalistic curation. Thus using a simplified model allows us to gain control of what effects curation alone can have on belief.

Let us now consider what happens when each of the curation practices discussed above is implemented.

4.1. Hyperbole

Under hyperbole the reported distribution always has higher variance than the actual distribution, as in figure 1b. Whenever the mean of the actual distribution is different from the societal neutral point (i.e., non-zero in our model), hyperbole shifts this mean to make it more extreme.

Under hyperbolic reporting agents nearly always converge to a posterior distribution that closely matches the reported distribution, regardless of their initial beliefs. This means that in most cases, hyperbole is detrimental to the accuracy of the agent's beliefs. Agent belief means are inaccurate whenever the actual mean is non-zero. Figure 2 shows the correlation of different parameters in the model with Squared Error in Mean of the posterior beliefs of agents.[19] In general, the exact values of correlation in this and following figures should not be taken too seriously, as they depend on the particular parameter values we ran. As is clear, though, across parameter variations hyperbole is meaningfully correlated with squared error in mean of the agent's posterior distribution.

In addition hyperbole leads to systematic overestimation of the variance of the actual distribution. Figure 3 shows the correlation of various parameters with Error in Variance. As is evident, agents exposed to hyperbole think actual event vary more than they truly do. This makes sense, since hyperbole increases the variance of the reported distribution, and agents come to fit their posterior beliefs to the distribution of events reported to them. Imagine reporting on the severity of hurricane seasons that always exaggerates either the mildness or the extremity of a season. Readers encountering this reporting should draw the conclusion that hurricane seasons are much more variable than the reality.

[19] We generate these correlation values over our entire data set.

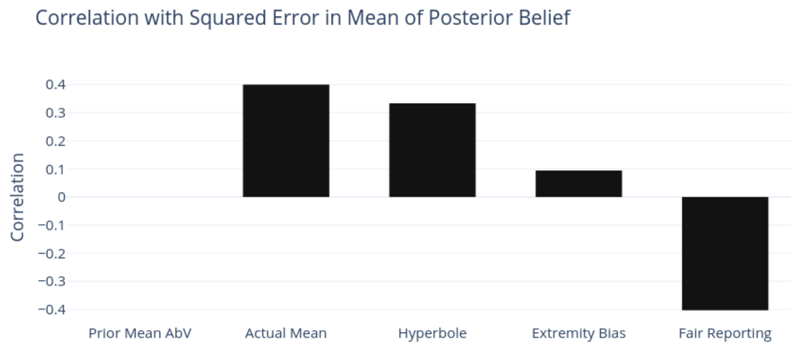


Figure 2: Correlation between various parameters in the model and the average Squared Error in Mean of posterior beliefs. AbV stands for absolute value, which we consider because what matters here is the deviation from the societal neutral point.

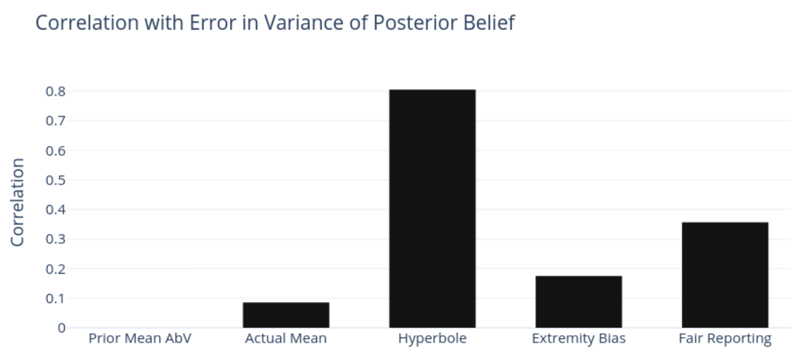


Figure 3: Correlation between various parameters in the model and the average Error in Variance of posterior beliefs.

4.2. Extremity bias

Extremity bias, as noted, creates a reported distribution with two peaks, one on each side of the societal neutral point; with a gap in the middle. Moreover, like hyperbole, this bias will shift the mean of evidence reported to a more extreme value than the actual mean whenever the actual mean is not zero (i.e. does not match the societal neutral point).

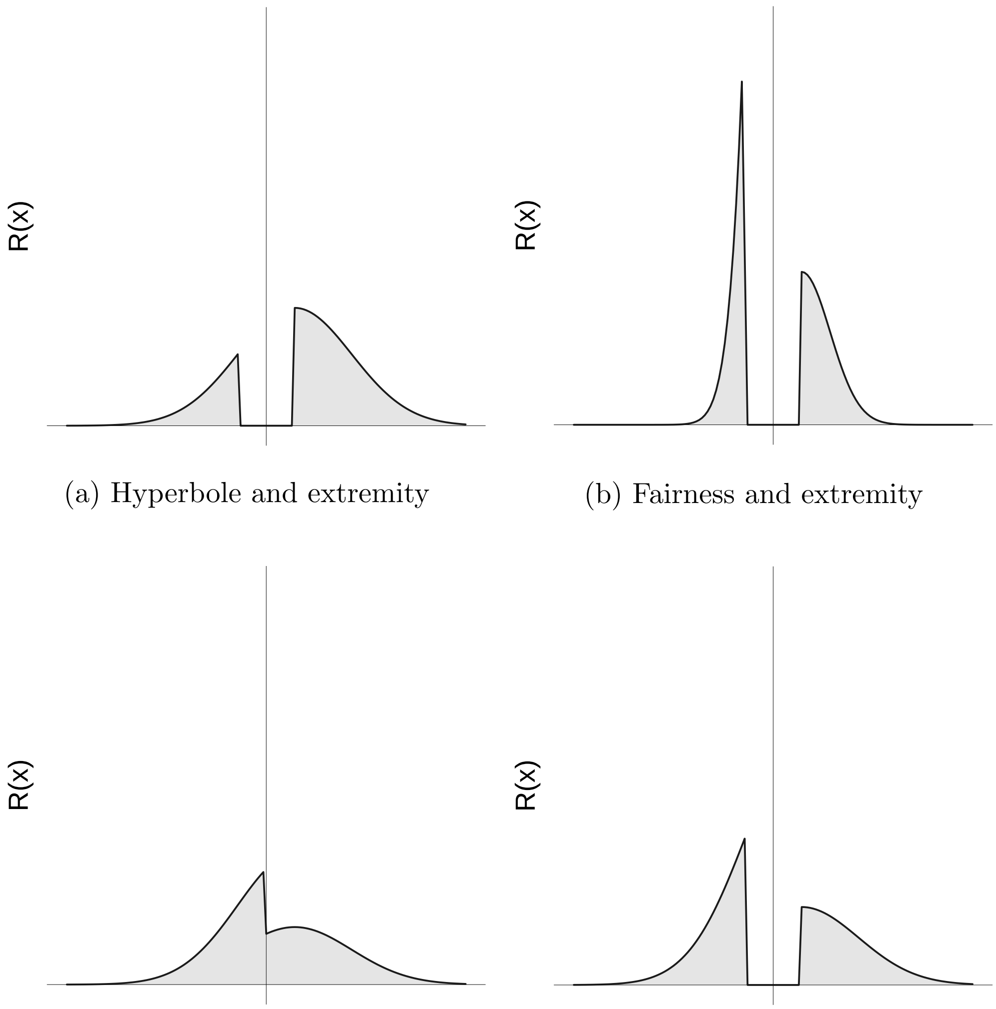
Given that the agent assumes the distribution of events is normal, their posterior distribution will not match reported distributions in the case of extremity bias, because the reported distribution is not normal. Instead, the agent's posterior distribution will be a normal approximation to the reported distribution, i.e., it will tend to have mean and variance very close to those of the reported distribution. For this reason, extremity bias leads to Squared Error in Mean, as is clear in figure 2. As with hyperbole, the estimated mean will always be more extreme than the truth.

With respect to variance, extremity bias can lead to both over- and under-estimations. When there is significant weight on both peaks of the reported distribution, the agent will estimate variance of events to be too high. When the reported distribution is such that almost all events lie on just one side of the societal neutral point, the agent will come to believe that variance is too low. This will tend of happen if the actual mean is far from the societal neutral point. In such cases, extremity bias can almost obliterate data on one side of the distribution. Consider reporting on hurricane seasons that only tracks the mildest and most severe events. Readers will tend to overestimate the variability of hurricane events if they see a good number of reports from both sides. If they only see the most severe events, because no one is interested in mild ones, they will think the world is (invariably) a severe one.

4.3. Fairness

When the societal neutral point coincides with the mean of the actual distribution, D_A, fair reporting does not distort the reported distribution since it is already symmetric. But whenever the actual mean is not 0, fair reporting leads to a doubly-peaked distribution, as in figure 1d. In general this leads to a posterior distribution whose mean is closer to the societal neutral point than the actual distribution. Fair reporting overweights event values on one side of the neutral point and underweights event values on the other side, thus drawing observer mean beliefs towards zero. The further the societal neutral point from the actual distribution, the greater the distortion. Again consider reporting on hurricane season, but where climate change has greatly increased its true severity. Fair reporters, trying for societal neutrality, might share information about mild seasons along with each severe season. Readers will get the impression that mild seasons are more common than the reality.

This said, as is evident in figure 2, across other parameter values in our model fair reporting generally has a positive effect on error, i.e., it is negatively correlated with Squared Error in Mean. This occurs because fairness corrects for errors



(a) Hyperbole and extremity

(b) Fairness and extremity

(c) Hyperbole and fairness

(d) All three distortions

Figure 4: Various distributions with combined reporting effects.

caused by hyperbole and extremity bias, which tend to push the reported mean away from the societal neutral point. In the absence of these other reporting biases, though, fairness does not improve mean belief.

Reported distributions under fair reporting tend to have higher variance than without it, since fairness increase the weight in one tail of the actual distribution (see 1d). This average effect is evident in figure 3.

4.4. Combining Effects

What if journalists engage in more than one of these practices at a time?[20] In many cases, the negative effects of the reporting practices are exacerbated. For instance, hyperbole and extremity bias together create a reported distribution with a more exaggerated mean than either alone (figure 4a). On the basis of this sort of report, the agent tends to develop highly inaccurate posterior beliefs. It is under this sort of regime that fair reporting can sometimes help. It can draw the reported mean back towards the societal neutral point, and thus back towards the actual mean value (see figure 4c).

This said, the combination of fair reporting and the other biases creates a reported distribution that looks very little like the actual distribution. It puts far too much weight on extreme events, and also distorts which extreme events are most common (4b & 4d). This leads agents to strongly overestimate the variance of events in the real world.

4.5. Actual Mean

Figure 2 shows that the absolute value of μ_A, i.e., how far the mean of the actual distribution of events lies from the socially accepted neutral point, strongly predicts Squared Error in Mean in this model. Without any reporting distortions, the location of the neutral point relative to μ_A does not matter. Agents will always end up with correct beliefs about the actual mean because they see reports that accurately track the actual distribution.

On the other hand, reporting practices tend to be most distortative when the neutral point is inaccurate. In these cases all three reporting biases shift the reported mean away from the true one. This leads us to an important take-away from this paper. Widespread errors in public belief, or mistaken assumptions on the part of journalists about what is perceived as "normal", can potentially drive misleading reporting by determining what is perceived as extreme, and what is perceived as fair. This leads to further error in public belief. In other words, current false beliefs can feed into the creation of ongoing false beliefs via journalistic

[20] Note that combining the practices in different orders will generally produce different distributions. Here we attempt to describe qualitative features of combined practices in a way that does not depend on the ordering. But for the purposes of quantitative results, we applied the biases in the following order: hyperbole, extremity bias, fairness. Thus, journalists first exaggerate their drawn sample, then check if it exceeds the extremity threshold, and then search for another sample that, when exaggerated, lies beyond the threshold but with opposite sign.

Parameter	Role in the Model	Values Studied
κ	Bias Parameter	$0, .1, .5, 1$

Table 2: Additional model parameters and values introduced in section 5.

biases. This possibility is not widely understood or acknowledged, and points to an important lever for improving journalistic reporting.

5. Confirmation Bias

Real-world media consumers are subject to various cognitive biases that affect how they take up and learn from data. In this section, we consider how the journalistic practices we have considered interact with one important cognitive bias.

Confirmation bias is the widespread tendency to focus on evidence that supports one's current beliefs. There are, in fact, a cluster of related behaviors under this heading, such as actively seeking out confirming evidence, ignoring disconfirming evidence, and selectively recalling facts during reasoning [29]. Many have argued that this tendency significantly harms belief formation [1, 19]. When it comes to controversial issues like climate change, confirmation bias seems to play an important role in public belief.[21]

We modify the model to include confirmation bias by making the agents more responsive to data they expect to receive than data that is far from their expectation. In this version agents do not update on all events that are reported to them. And the further an event from their current expectation of the mean, the more likely it will be ignored. While this will not capture all the behaviors associated with confirmation bias, it is a reasonable approximation.

More precisely, we introduce a new *bias parameter*, κ, which measures the strength of the agent's confirmation bias. (See Table 2.) Suppose that an agent (at some stage of the model) has estimated posterior mean μ_p. Then she will accept a

[21] Though some have pointed out that it is hard to know exactly what sorts of reasoning biases explain observed behavior. See Druckman and McGrath [15] for a review of these issues. On the issue of climate change in particular, Bolsen and Druckman [3] find that when exposing subjects to the same information, Democrats familiar with the issue, but not Republicans, changed belief to reflect the evidence presented. Zhou [39] found that messages about climate change action did not sway Republican targets, and even sometimes led to backfire effects (i.e., movement in the other direction). In both these studies, participants responded differently to evidence based on their background beliefs.

report $y \in \mathbb{R}$ with probability given by a function of the distance between posterior mean and report $f_\kappa(|y - \mu_p|)$, where the specific function depends on the value of κ. In particular, we stipulate that the probability of acceptance is always 1 when the distance between report and estimated mean is 0. The probability decreases as this distance increases.[22] The probability is always non-negative. And, for $\kappa = 0$, i.e., in the absence of bias, agents accept all data. Thus $\kappa = 0$ coincides with the model discussed in previous sections. Again, the interpretation of the confirmation bias model is just that agents reject data far from their estimated mean with probability related to the distance from that mean.[23]

We now turn to results. As before, we present agents with sequential draws from the reported distribution and allow them to update. Because agents reject some data in this model, we continue the simulation until each agent updates 1000 times. Notice that confirmation bias in this model can create strongly path dependent effects. If agents receive early reports pushing them in one direction, they might reject later reports pulling them another way. For this reason, we see much more variability in outcomes. We again report average errors over model runs and qualitative descriptions of results.[25]

[22] Except when there is no bias, i.e., $\kappa = 0$.

[23] More formally, the one-parameter family of functions f_κ is chosen to have the following qualitative features: (1) for all κ, $f_\kappa(0) = 1$; (2) for all κ, f_κ is non-negative; (3) for all $\kappa \neq 0$, f_κ is monotonically decreasing away from 0, with $f_\kappa \leq f_{\kappa'}$ whenever $\kappa \geq \kappa'$;[24] and (4) when $\kappa = 0$, $f_\kappa = 1$. We realize these qualitative features with the family of functions defined by $f_\kappa(x) = 2 * (1 - g_\kappa(x))$ $\kappa \neq 0$, where $g_\kappa(x)$ is given by the cumulative density function of a normal distribution with mean 0 and variance $1/\kappa^2$. We stipulate that $f_0 = 1$. One way to think about this choice of functions is that the agent effectively constructs a model of the world according to which actual events are expected to be sampled from a normal distribution with mean μ_p and variance $1/\kappa^2$; and then accepts reported data with probability given by the $p-$value of that (unsigned) data. Thus, data that is less likely, according to the model constructed by the agent, has a higher probability of rejection, and data that coincides with the agent's mean, meanwhile, is always accepted. Note that although we explain the behavior of f_κ by supposing the agent constructs a certain model of the world and then uses it to calculate probabilities, we do not mean to suggest that real agents do construct such a model, nor that confirmation bias is a high-rationality decision process. Our model should not be interpreted as making those imputations; we use these functions only because they conveniently realize the qualitative features discussed in the main text.

[25] Although there is increased variability in outcomes, remember that these models were programmed independently, and results gathered independently, by two different team members. Findings were similar indicating that, averaged over 1000 runs, this path dependency did not importantly impact reported results.

5.1. General Results

In the model previously described, agents generally ended up with beliefs reflecting the mean and variance of the reported distribution of events. Once one adds confirmation bias, things change. Agents now ignore some of these reported events, and thus they update on data that no longer reflects the entire reported distribution. Instead, under many conditions agents end up with their belief distribution shifted toward the nearest peak of the reported distribution. A far away peak may be mostly ignored. This tends to distort agents' beliefs compared to a model with no confirmation bias.

In addition, confirmation bias often slows learning down since agents mostly reject disconsonant data that might have a big impact on their current beliefs. This means that even with 1000 reports to update on, agents do not always end up with stable posterior distributions (though in a much longer simulation they should). This effect is more dramatic for prior means that are very far away from μ_A. For these reasons confirmation bias is generally correlated with less accurate posterior means, as shown in figure 5.

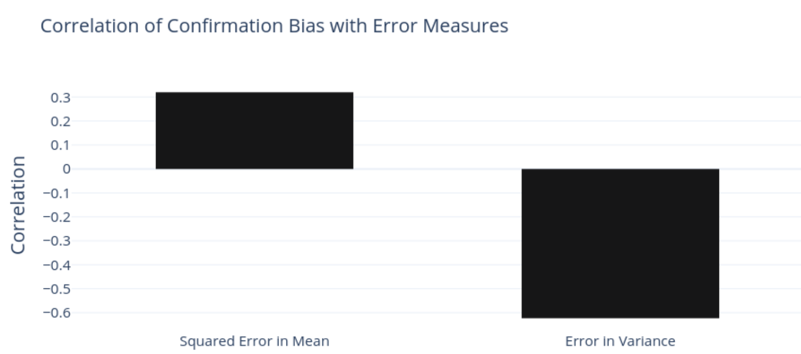


Figure 5: Correlation between confirmation bias and errors in posterior estimates of mean and variance.

As is also evident in this figure, confirmation bias is strongly negatively correlated with Error in Variance. This is because confirmation bias leads agents to ignore data far from their current mean. The data they update on, then, comes from a relatively narrow distribution, and their final beliefs reflect that.[26] This

[26] In the presence of reporting distortions like hyperbole and extremity bias that over-report extreme events, confirmation bias can actually improve estimated variance by correcting for the reporting bias.

might correspond to a climate change denier who, because they believe climate change is not real, does not update their beliefs on reports of increasingly extreme weather events.

5.2. Hyperbole

In the presence of confirmation bias, one might think hyperbolic reporting could sometimes be useful. For example, suppose an individual were convinced that recent fire seasons were relatively mild, when in reality they had been quite serious. Hyperbolic reporting would shift the reported mean of events to a more extreme value. Perhaps this might convince the skeptic?

Our model does not support this possibility. Because hyperbolic reporting flattens the distribution of reported evidence, under this reporting bias, data is shared that is consistent with a wide range of prior beliefs. Agents engaging in confirmation bias tend to update on the data that is closest to their present mean. This means that, in general, hyperbole provides fodder for the tendency to stick with one's current beliefs.

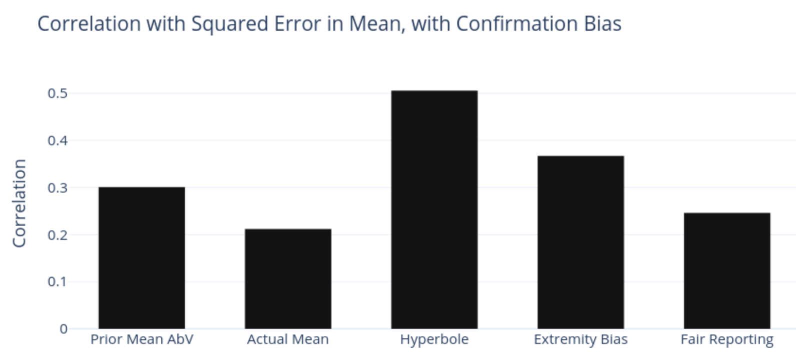


Figure 6: Correlation between various parameters in the model and the average Mean Squared Error of posterior beliefs, in the presence of confirmation bias.

In addition, just as in the model without confirmation bias, by shifting the mean of the reported distribution to an inaccurately high value, hyperbole can pull agent beliefs too far towards the extreme. As figure 6 shows, in the presence of confirmation bias hyperbole still shows a strong positive correlation with Squared Error in Mean. (Results reported are for confirmation bias, $\kappa = 1$. Trends should be similar for other values.)

And again, hyperbole tends to lead to an overestimation of variance in the actual distribution, as is shown in figure 7. But in the presence of confirmation bias, this impact is somewhat mitigated because, as noted, confirmation bias tends to lead agents to underestimate variance.

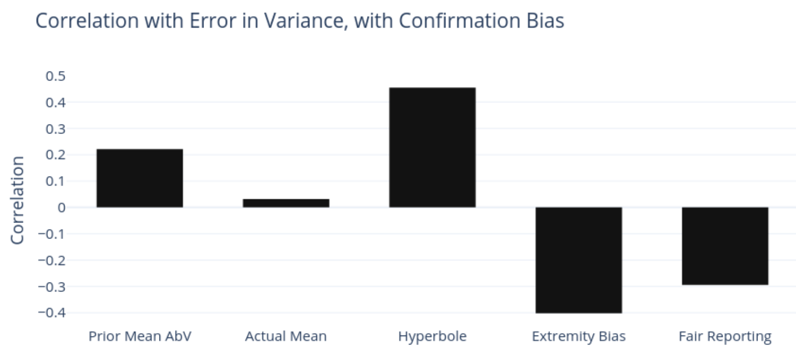


Figure 7: Correlation between various parameters in the model and the average Variance Error of posterior beliefs, in the presence of confirmation bias.

5.3. Extremity Bias

In the presence of confirmation bias, extreme reporting can lead to polarization. We mean this in two senses. First, as noted, agents often develop a posterior distribution with mean close to the nearest reporting peak. So agents with prior means near one peak will head towards that peak, while agents with prior means near the other will head there. If we imagine these individuals as part of a larger population, we expect stable subgroups to emerge with different beliefs. Such subgroups would not arise in the absence of extreme reporting, even with confirmation bias. Figure 8 illustrates this effect.

Second, under this regime an agent who has prior beliefs between the reported peaks can be pulled in either direction, as a consequence of stochastic effects in the reporting time series and the agent's algorithm for accepting data. In other words, individuals in identical epistemic states can end up at different extreme beliefs as a result of path dependency. Figure 9 illustrates this possibility. We might imagine someone who first hears about a number of serious cases of some disease, and forms an impression of the disease as serious. They then start to discount reports that the disease is sometimes mild, and update only on more

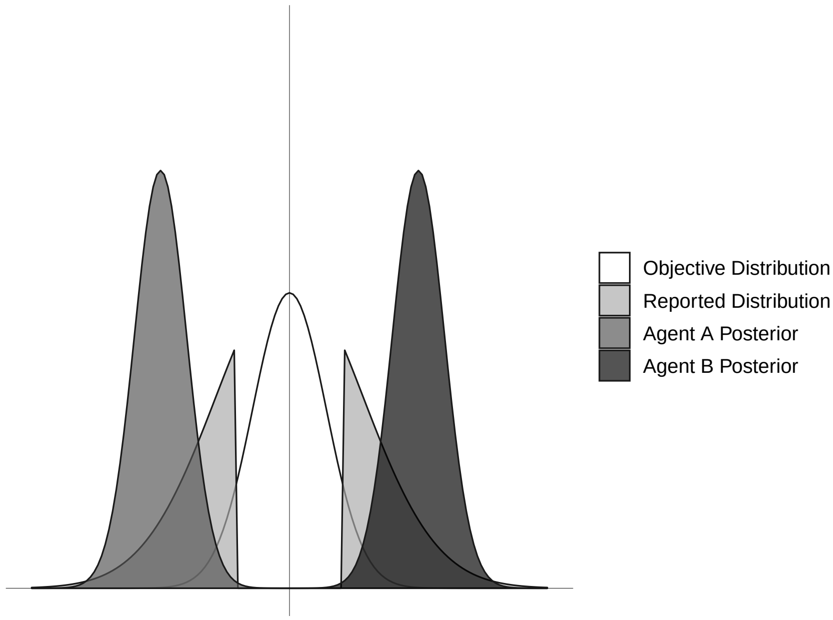


Figure 8: Objective distribution with mean at zero; reported distribution exhibiting extremity bias; and posterior beliefs of agents who started at -1 (Agent A) and 1 (Agent B).

serious reports. Because under extremity bias there are no reports of intermediate (moderate) cases, they will be slowly pulled towards extreme beliefs.

These sorts of effects mean that the combination of extremity bias and confirmation bias leads to highly inaccurate beliefs. As figures 6 and 7 show, extremity bias is strongly correlated with both sorts of error.[27]

We wish to highlight something interesting about these results. Extreme reporting, unlike hyperbole, does not involve falsehood. It is problematic only inasmuch as it involves omission of some relevant events. But generally it is not considered unethical for journalists to report only on interesting or extreme events. When this practice is adopted by an entire journalistic community, though, we see here that it can be highly misleading and drive polarization.

5.4. Fairness

In the model without confirmation bias, fair reporting was negatively correlated with Squared Error in Mean across different parameter choices. In other

[27] Note that it is correlated with underestimation of variance. Those engaging in confirmation bias, and updating only on extreme reports from one end of a distribution, tend to have very narrow variance estimates, since they ignore much of the available data. (See figure 8.)

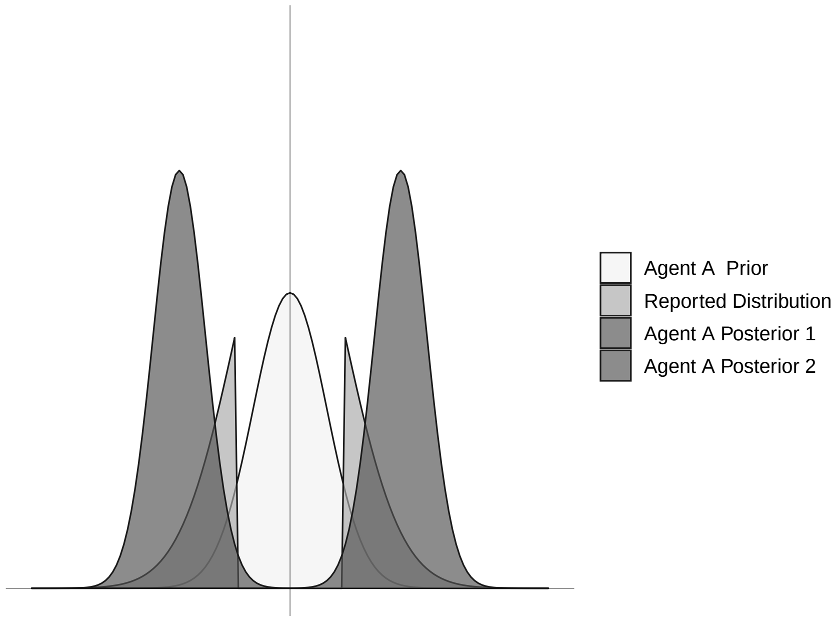


Figure 9: Two divergent posterior beliefs emerging from a single prior belief.

words, on average it improved belief accuracy. This was because fair reporting helped compensate for distortative effects from hyperbole and extremity bias. In the presence of confirmation bias, however, we see just the opposite. Fair reporting is strongly correlated with Squared Error in Mean, as shown in figure 6.

Like extremity bias, fair reporting can lead to a double peaked distribution of reported events. And again, confirmation bias can lead to polarization, where agents develop posterior beliefs that are near just one of these peaks. When that happens, their posterior means often will not be accurate.

In addition, this effect negates the ways that fairness could improve outcomes in the model without confirmation bias. In that version of the model, fair reporting could counterbalance the effects of hyperbolic reporting by drawing the reported mean closer to the neutral point. Under confirmation bias, it does not matter that the reported mean is closer to the actual mean, because agents instead coalesce around the two inaccurate peaks.

Note that there is something surprising about the fact that fairness tends to have such a negative effect on posterior mean here. It is a norm of journalism to report fairly. But in the presence of confirmation bias, where actors ignore reports that are far away from their current beliefs, it can be harmful. This is a potential harm of "false balance" that goes beyond over-weighting misleading evidence. It also allows agents to ignore disconfirming data more thoroughly.

5.5. Combining Effects

The most potently misleading combination in this model involves extremity bias, fairness, and confirmation bias. Under this regime, reported events are extreme, but in addition reporters attempt to represent the extremes fairly. This means that one extreme tends to be strongly over-represented in reporting, as is evident in figure 4. In these cases polarization is especially common as agents tend towards one of two equally weighted extreme peaks of reporting. Hyperbole, when added to these other factors, can also exacerbate error. Because it tends to skew reported distributions towards the extremes, it will tend to make reported extreme peaks even more misleading.

5.6. Priors

As in the model without confirmation bias, if the actual mean of the objective distribution does not line up with the societal neutral point, this tends to predict error. The reasons for this are much the same as in the model with no confirmation bias. In that model, though, prior beliefs do not matter. Agents' beliefs come to match the reported distribution as well as they can, and this is not strongly dependent on their priors.

Once we include confirmation bias, though, prior mean matters because it anchors what the agent is willing to learn about. As we see from figure 6, the absolute value of agents' prior means correlates significantly with Squared Error in Mean. This is because more extreme prior means tend to lead agents towards more extreme eventual beliefs, and tend to make it harder for them to update successfully if the actual mean is far from their priors.

6. Conclusion

We have presented a model where agents (with or without confirmation bias) learn about the world from data curated by an intermediary. We consider the base case where all data is shared, and compare it with three selection procedures inspired by common journalistic practices. We argue that the practices we consider, two of which are generally compatible with existing norms for good journalism, can have serious negative consequences—particularly when agents consuming the reports are subject to confirmation bias. Our models are highly simplified, but this simplicity allows us to isolate the effects of these curation practices, and show that on their own they can distort public beliefs.

We might conclude that ideal curation would faithfully track the distribution of true events in the world. There are cases where this will not be right. Suppose, for instance, that the public is systematically misinformed on some topic and tend to have mean beliefs far from the actual mean. Reporting that exaggerates in the other direction might improve belief over faithful reporting. But generally our models indicate that across a range of conditions reporting that accurately tracks actual distributions tends to do very well.

Of course, this sort of accurate reporting is tricky when news items correspond to single events. For instance, reporting on fires will often involve descriptions of a single (ongoing) fire. Unlike the reports in our model, though, real instances of reporting typically contain much more information than a selection from a distribution. Our analysis here points to the importance of contextualizing single events in a broader picture. In reporting some extreme event, it is incumbent on reporters to also give a broad sense of what the wider distribution of events looks like.

In the case of scientific reporting, many journalists already follow practices of this sort. For instance, some venues no longer print articles reporting the results of single research studies, since they are too likely to be misleading.[28] Instead it is expected that reporters survey an entire body of scientific literature and deliver reporting on the picture that emerges from many studies.[29] Of course, this sort of reporting is time intensive.[30]. This indicates that there are pressures that might prevent the adoption of this sort of practice across news sources. But we can nonetheless label it as a promising one.[31]

Another core take-away from our work here has to do with the existence of

[28] This has been confirmed through personal correspondence with science editors.

[29] This relates to "weight-of evidence" norms that journalists, "...find out where the bulk of evidence and expert opinion lies on the truth continuum and then communicate that to audiences" [16]. For those who advocate this standard of journalistic objectivity, rather than fair reporting, the investigation here provides some further fodder.

[30] Fengler and Ruß-Mohl [18] point out that, "...journalists have to weigh carefully the question of how much time they are going to invest in a story, and they rank 'time pressure' high among the list of negative aspects of the journalistic profession" (675)

[31] See also Weatherall et al. [38] who come to a similar conclusion using a very different model of reporting about scientific data.

societal neutral points that deviate from the central behavior of the real distributions of events. As we point out, when this happens, characteristic reporting distortions can more significantly mislead public opinion. In other words, there is a vicious cycle in which public error can feed into journalistic distortion, which then can feed into further public error.

A third important take-away is that in the presence of confirmation bias, curation practices that may not seem problematic, or which may be adopted with the intent of improving epistemic outcomes, can in fact cause serious distortions. This happens when such practices tend to yield distributions of events with different peaks, about which members of the public can polarize. This is somewhat surprising since hyperbole—which involves outright exaggeration of facts—does not have this particular harmful effect the way extremity bias and fairness do.

Notice, though, that we only consider some reporting biases, and, in particular, we do not address some of the worst ones. For example, highly partisan news sources will sometimes cherry-pick event values to report so severely that only values on one side of the societal neutral point are represented at all. In the presence of such a practice, requirements for fairness will likely significantly improve beliefs. This is to say that while we have shown potential downsides to fair reporting, it may play an important role in preventing even more serious reporting distortions from impacting public belief.

While we focus on more traditional journalism here, we think there are important lessons for an increasingly important curator—social media algorithms. Commentators on social media often emphasize how changes in algorithms could reduce the propagation of false and misleading content. While this is important, we think that this conversation should focus more on the statistical distribution of accurate content that users see. Curation of accurate content can have dramatic impacts on beliefs, as demonstrated here. This worry is addressed somewhat in the literature on echo chambers, where algorithms may play to confirmation bias by presenting confirmatory data to different users [11, 37]. But this is not the only sort of curatory harm algorithms might perpetuate, and wider discussion is needed.

There are several simplifying assumptions of our models that may be worrying. For example, our societal neutral point is always fixed. We might instead have supposed that it was something like an average of the means of the posteriors of a group of agents, which could update as this group changed beliefs in light of reporting. This sort of of endogeniety might seriously alter the sorts of final belief states arising in the model. We think this is a promising extension of the work presented here, though we have not pursued it at this time.[32]

There are other simplifications in our model that may attenuate the applicability of our results. We make strong assumptions about the actual distribution of events (normal) that will not hold for many real world cases. In addition, our agents employ a fixed assumption that the actual distribution is normal. They do not notice or react when the data they receive is unlikely to have been sampled from a normal distribution. This may actually mean that our model underestimates the distortative effects of reporting biases on public beliefs.[33]

Furthermore we do not model our agents as capable of higher level uncertainty about their belief distributions. In reality, people often become confused and uncertain in the light of evidence that is too variable or contradictory. For instance, in psychological studies testing the influence of "false balance", authors have found that subjects report more uncertainty when presented with contradictory expert opinions [14, 10, 23].[34] (Though see Corbett and Durfee [12] who find mixed results.) It may be that in the real world, distortions like extreme reporting, fairness, and hyperbole cause further damage by increasing the amount of seemingly contradictory evidence that audiences see.

As with many such simple models, this one is best understood as a way to explore the implications of various assumptions and generate hypotheses that must be tested and refined empirical. We think this exploration has identified a number of possible effects of common reporting biases that deserve further attention, and highlighted the way that practices that may nonetheless seem innocuous may

[32] Even so, the current fixed neutral point model is relevant to thinking about many cases. First of all, it is useful for thinking about belief updating in the short term. Second the models we consider here are relevant to cases where the neutral point is fixed by exogenous factors. In such cases there may be natural neutral points having to do with the topic at hand. For instance, the general assumption might be that side effects from a new vaccine are likely to be about as bad as side effects from previous, similar vaccines. With topics like climate change, the neutral point is sometimes established around happening vs. not-happening (rather than with a more reasonable estimate of severity). In other cases, outside interests shape these neutral points cynically, again as in the case of climate change.

[33] For example, in the model without confirmation bias extreme reporting is not very harmful when the societal neutral point and the actual distribution match. But this is in part because the learner assumes a distribution similiar to the real world one, rather than a distribution that closely matches the reported one.

[34] This seems to be based in general psychological tendencies towards uncertainty in the face of conflicting evidence [22, 7].

cause epistemic harm.

Acknowledgements

This work is supported by the National Science Foundation under grant number 1922424 "Consensus, Democracy, and the Public Understanding of Science". Thanks to those in our working group for feedback including Matthew Coates, David Freeborn, Yu Fu, Nathan Gabriel, True Gibson, Daniel Herrman, Saira Khan, Ainsley Pullen, Kaetlin Taylor, and Jingyi Wu. Many thanks to the organizers of the Social Minds in Digital Spaces workshop, especially David Barrett and Eric Funkhouser, for the opportunity to present and receive feedback. And thanks to audience members at that workshop for helpful comments.

References

[1] Jonathan Baron. *Thinking and deciding*. Cambridge University Press, 2000.

[2] W Lance Bennett. An introduction to journalism norms and representations of politics. 1996.

[3] Toby Bolsen and James N Druckman. Do partisanship and politicization undermine the impact of a scientific consensus message about climate change? *Group Processes & Intergroup Relations*, 21(3):389–402, 2018.

[4] Maxwell T Boykoff and Jules M Boykoff. Balance as bias: global warming and the us prestige press. *Global environmental change*, 14(2):125–136, 2004.

[5] Maxwell T Boykoff and Jules M Boykoff. Climate change and journalistic norms: A case-study of us mass-media coverage. *Geoforum*, 38(6):1190–1204, 2007.

[6] William J. Brady, Julian A. Wills, John T. Jost, and Jay J. van Bavel. Emotion shapes the diffusion of moralized cotent in social networks. *Proceedings of the National Academy of Science*, 114(28), 2017.

[7] Lyle A Brenner, Derek J Koehler, and Amos Tversky. On the evaluation of one-sided evidence. *Journal of Behavioral Decision Making*, 9(1):59–70, 1996.

[8] Jay Pil Choi and Sangwoo Yang. Investigative journalism and media capture in the digital age. *Information Economics and Policy*, 57:100942, 2021.

[9] Jihyang Choi, Sang Yup Lee, and Sung Wook Ji. Engagement in emotional news on social media: Intensity and type of emotions. *Journalism Mass Communication Quarterly*, 98(4):1017–1040, 2021.

[10] Christopher E Clarke, Graham N Dixon, Avery Holton, and Brooke Weberling McKeever. Including "evidentiary balance" in news media coverage of vaccine risk. *Health communication*, 30(5):461–472, 2015.

[11] Michael Conover, Jacob Ratkiewicz, Matthew Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini. Political polarization on twitter. In *Proceedings of the international aaai conference on web and social media*, volume 5, pages 89–96, 2011.

[12] Julia B Corbett and Jessica L Durfee. Testing public (un) certainty of science: Media representations of global warming. *Science Communication*, 26(2): 129–151, 2004.

[13] Patricia A Curtin. Reevaluating public relations information subsidies: Market-driven journalism and agenda-building theory and practice. *Journal of public relations research*, 11(1):53–90, 1999.

[14] Graham N Dixon and Christopher E Clarke. Heightening uncertainty around certain science: Media coverage, false balance, and the autism-vaccine controversy. *Science Communication*, 35(3):358–382, 2013.

[15] James N Druckman and Mary C McGrath. The evidence for motivated reasoning in climate change preference formation. *Nature Climate Change*, 9 (2):111–119, 2019.

[16] Sharon Dunwoody. Weight-of-evidence reporting: What is it? why use it? *Nieman Reports*, 59(4):89, 2005.

[17] Robert Eschelman. The danger of fair and balanced. *Colombia Journalism Review*, 2014. URL `https://archives.cjr.org/essay/the_danger_of_fair_and_balance.php`.

[18] Susanne Fengler and Stephan Ruß-Mohl. Journalists and the information-attention markets: Towards an economic theory of journalism. *Journalism*, 9(6):667–690, 2008.

[19] Leon Festinger, Henry Riecken, and Stanley Schachter. *When prophecy fails: A social and psychological study of a modern group that predicted the destruction of the world*. Lulu Press, Inc, 2017.

[20] Ross Gelbspan. *The heat is on: The climate crisis, the cover-up, the prescription*. Perseus Books Cambridge, MA, 1998.

[21] James Hamilton. *All the news that's fit to sell: How the market transforms information into news*. Princeton University Press, 2004.

[22] Daniel Kahneman and Amos Tversky. On the psychology of prediction. *Psychological review*, 80(4):237, 1973.

[23] Derek J Koehler. Can journalistic "false balance" distort public perception of consensus in expert opinion? *Journal of Experimental Psychology: Applied*, 22(1):24, 2016.

[24] Jose Mari Hall Lanuza and Cleve V. Arguelles. Media system incentives for disinformation. In Herman Wasserman and Dani Madrid-Morales, editors, *Disinformation in the Global South*, pages 123–139. John Wiley Sons, Hoboken, NJ, 2022.

[25] Stephan Lewandowsky, Toby D Pilditch, Jens K Madsen, Naomi Oreskes, and James S Risbey. Influence and seepage: An evidence-resistant minority can affect public opinion and scientific belief formation. *Cognition*, 188:124–139, 2019.

[26] John H McManus. *Market-driven journalism: Let the citizen beware?* Sage Publications, 1994.

[27] Denis Muller. *Journalism Ethics for the Digital Age*. Scribe, Melbourne, 2014.

[28] Dorothy Nelkin. The culture of science journalism. *Society*, 24(6):17–25, 1987.

[29] Raymond S Nickerson. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, 2(2):175–220, 1998.

[30] Moti Nissani. Media coverage of the greenhouse effect. *Population and Environment*, 21(1):27–43, 1999.

[31] Cailin O'Connor and James Owen Weatherall. *The Misinformation Age: How False Beliefs Spread*. Yale University Press, New Haven, CT, 2019.

[32] Society of Professional Journalists. Code of ethics. 2014. URL `https://www.spj.org/ethicscode.asp`.

[33] Naomi Oreskes and Erik M Conway. *Merchants of doubt: How a handful of scientists obscured the truth on issues from tobacco smoke to global warming*. Bloomsbury Publishing USA, 2011.

[34] Victor Pickard. *Democracy without Journalism? Confronting the Misinformation Society*. Oxford University Press, New York, NY, 2019.

[35] Todd F Simon, Frederick Fico, and Stephen Lacy. Covering conflict and controversy: Measuring balance, fairness, defamation. *Journalism Quarterly*, 66(2):427–434, 1989.

[36] Jane B Singer. Contested autonomy: Professional and popular claims on journalistic norms. *Journalism studies*, 8(1):79–95, 2007.

[37] Cass R Sunstein. # republic. In *# Republic*. Princeton university press, 2018.

[38] James Owen Weatherall, Cailin O'Connor, and Justin P Bruner. How to beat science and influence people: policymakers and propaganda in epistemic networks. *The British Journal for the Philosophy of Science*, 71(4):1157–1186, 2020.

[39] Jack Zhou. Boomerangs versus javelins: how polarization constrains communication on climate change. *Environmental Politics*, 25(5):788–811, 2016.