Can Fabricated Data be Ignored when it is Detected?

by

Adam T. Ramsey

Thesis

Submitted to the faculty of the

Graduate School of Vanderbilt University

In partial fulfillment for the degree of

Master of Science

in

Psychology

May 13, 2022

Nashville, Tennessee

Approved:

Jennifer S. Trueblood, Ph.D.

Geoffrey F. Woodman, Ph.D.

To my wonderful wife, Meg,
for her steadfast support and encouragement

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

As information sharing via social media grows, individuals are increasingly exposed to misinformation that could impact their beliefs. Concerningly, false news often spreads more quickly and broadly online than true news (Vosoughi et al., 2018). The ease with which misinformation is disseminated online produces an environment where a few insistent voices sharing false information can sway much of the populous (Cook and Lewandowsky, 2016). Even those who do not intend to misinform may do so by sharing false news stories simply because they have seen the headline before (Effron and Raj, 2020).

It can be easy to think that identifying misinformation allows one to ignore it, but research into the continued influence effect (CIE; Johnson and Seifert, 1994) has shown that people will utilize information that has been retracted, even if they remember that the information is not legitimate. This effect is difficult to eliminate completely. Invalidated information continues to influence beliefs in circumstances where the retraction was much stronger than the false information (Ecker et al., 2011), and when people are forewarned about misinformation (Ecker et al., 2010).

Interventions to reduce the influence of retracted information include shifting one's focus towards evaluating accuracy when encoding false information (Pennycook et al., 2021). Similarly, reading a debunking message shortly after encountering false information diminishes, but does not eliminate, the CIE (Brashier et al., 2021; Wilkes and Leatherbarrow, 1988). Debunking messages that include truthful information which replaces the retracted information in one's mental model tend to be the most effective (Chan et al., 2017) and have been recommended for news sources, social media sites, and educators (Lewandowsky et al., 2012). However, these interventions apply to situations where an outside source has invalidated a piece of information. It is also important to consider whether individuals can ignore information that they have deemed invalid for themselves.

In the illusory truth paradigm, participants indicate the degree to which they believe different statements, some of which are false. Later, participants are presented with another list of statements, some of which they had previously encountered. Results show that people more

strongly believe previously encountered statements, including false statements (Begg et al., 1992; Hasher et al., 1977). Additional repetitions of a statement increase belief of that statement even further (Hassan and Barber, 2021).

Illusory truth occurs even when one possesses knowledge that contradicts false information (Fazio et al., 2015). The increase in believability occurs for both plausible and implausible statements (e.g., “The Earth is a perfect square"; Fazio et al., 2019). Like the CIE, the illusory truth effect can be diminished, but rarely eliminated, by having participants focus on the accuracy of statements as they initially encounter them (Brashier et al., 2020).

Other research has also observed that ignoring information is not an easy task. 4-6 year old children found it difficult to ignore false information about a previous playdate (Schaaf et al., 2015). Adult jurors have difficulty ignoring inadmissible evidence when deliberating a verdict (London and Nunez, 2000). Even experienced judges who ruled evidence to be inadmissible struggle to ignore that evidence (Wistrich et al., 2004).

The previous literature has mainly investigated whether individuals utilize (CIE) or believe (illusory truth) false factoids, how misinformation spreads online (Effron and Raj, 2020; Lewandowsky et al., 2012; Vosoughi et al., 2018), and how to combat it (Brashier et al., 2021; Pennycook et al., 2021). The current paper extends this literature by investigating how individuals handle misinformation in the form of fabricated data.

News sources are increasingly reporting data to consumers (Westlund and Hermida, 2021), with data journalists often presenting their findings as fact. However, they less often acknowledge limitations such as data collection practices or conflicts of interest in their work. This requires consumers to be vigilant and sample many disparate findings of varying quality in order to estimate the truth of a matter (e.g., Covid-19 vaccine efficacy rates, global temperature change, etc.). Therefore, it is important to understand how people process noisy numerical information as they form beliefs of the underlying truth (Stubenvoll and Matthes, 2021), especially information they deem to be false.

In this paper we describe five experiments investigating whether people could ignore fabricated data when they detected it. We also examine manipulations aimed at helping people identify false data (i.e., visual warning cues). Across our experiments, we show that it is difficult for people to disregard false data even when this data is very easy to detect.

CHAPTER II

EXPERIMENT 1

Participants read imaginary scenarios in which they were presented sequences of values sampled from underlying Gaussian distributions. They attempted to ignore fabricated data while estimating the means of the underlying distributions. The general task presented here was used in all subsequent experiments.

Method

Participants

106 adult participants were recruited online via Amazon's Mechanical Turk platform using the CloudResearch platform. The study was approved by the Institutional Review Board at the authors' university. We intended to have approximately 50 participants per condition and the experiment had two between-subjects conditions. The sample size was determined prior to starting recruitment and was based on previous studies of the CIE and illusory truth effect (Ecker et al., 2011; Hassan and Barber, 2021). Data was analyzed only after all data had been collected. Each participant received $1.00 upon completion of the approximately 30-minute task. This self-paced task allowed participants to decide how many stimuli they viewed on each trial. Due to the nature of the experimental manipulation, participants needed to view at least three stimuli to possibly encounter the outliers of interest to our hypotheses (see Procedure section). Because of this requirement, participants with a median number of stimuli viewed per trial of less than three were excluded from the analyses. This criterion also served to protect against participants who merely clicked through the experiment as fast as possible without dutifully completing the task.

After excluding 21 participants due to low median stimuli viewed per trial, the final data were comprised of the remaining 85 participants. Ages ranged from 20 to 76 years old (mean = 39.88±13.31), and the sample was comprised of 45 female, and 39 male, and 1 non-binary participants.

Materials

Each trial consisted of a sequence of screens, each showing the results from a hypothetical medical study involving 20 patients (i.e., the number of patients out of 20 with negative side effects). To construct these sequences, a list of 51 fictional reports were created for each trial. Each individual report was simply a number between 1 and 20 sampled from a Gaussian distribution with given parameters for that trial (described in greater detail in the following paragraphs). The number of reports available in a given trial was chosen to be 51 somewhat arbitrarily, as it would be a large enough number of reports that participants would most likely cease their information search before exhausting the reports (which is critical to investigating information sampling behavior).

The experiment contained three main trial types: Control trials without outliers, Test trials with outliers, and Catch trials. The 16 Control trials were constructed by sampling from a Gaussian with mean of 8 and a standard deviation of 2. No reports indicating a value further than 2 SD from the mean were included in these trial lists.

The Test trials were designed to investigate the effects of outlier presence and magnitude on information seeking and one's final estimates. There were two sets of 8 test trials created by following the same procedure for generating the control trials, and then randomly inserting an outlier as report three, four, or five. Placing the outlier early in the sequence of reports ensured that participants were likely to see it when one appeared. Low outliers were 1, 2, and 3. High outliers were 13, 14, and 15.

The Catch trials were designed to ensure that there was sufficient variety in the stimuli so that participants would not simply learn that the mean report value was 8. There were 20 catch trials made up of 4 sets of 5 filler trials, with means of either 5 or 11 and SDs of either 1 or 2. These additional trials served as a distraction to help prevent participants from learning the structure underlying the catch and test stimuli.

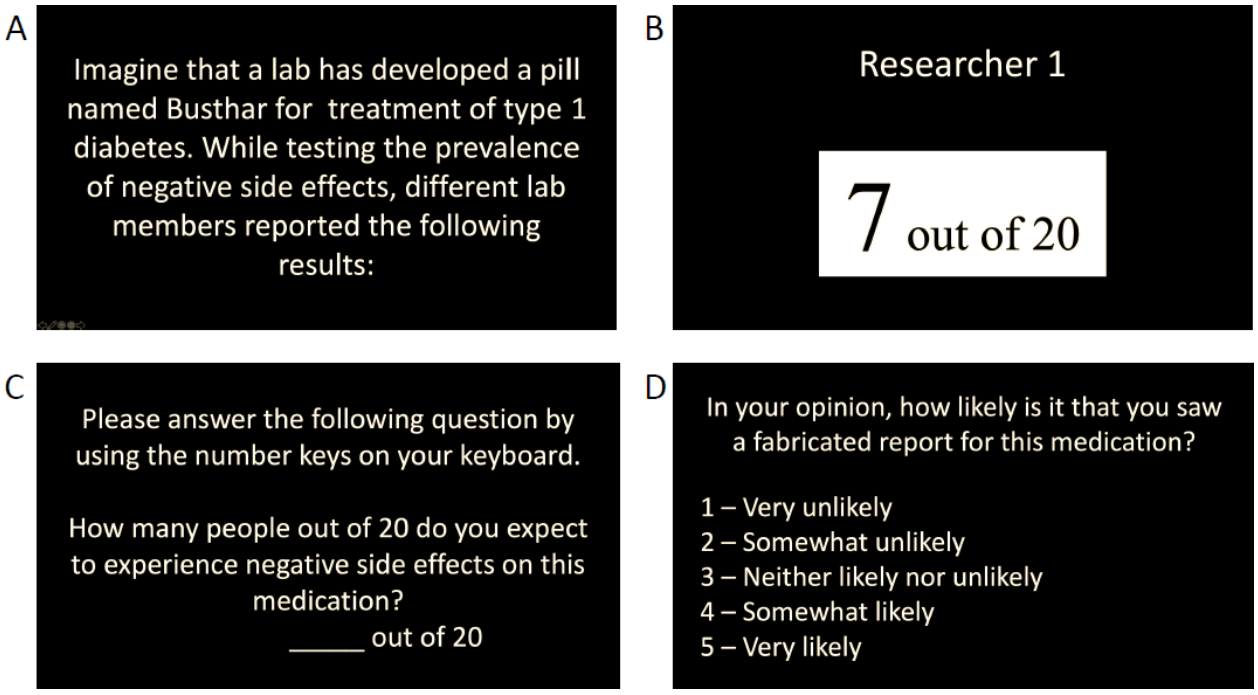

Figure 1

This figure depicts an example of the different screens participants viewed on each trial. Panel A: An example of the orienting story participants read at the start of each trial. The medical condition and fictional drug names were determined randomly for each trial. Panel B: An example of the stimuli participants viewed on each trial. Each stimulus represents one fictional researcher's report, and participants sampled reports until they felt comfortable estimating the underlying true prevalence rate. Panel C: The response screen where participants typed their estimates after terminating their information search. Panel D: The outlier detection response screen. One group in Experiment 1 and all participants in the following experiments indicated how likely it was that they encountered fabricated data at the end of each trial.

Procedure

Participants were randomly assigned to one of two between-groups conditions before the experimental session began. One group of participants provided only their estimates of side effect prevalence rates. The other group provided these estimates, but also indicated their confidence that they detected fabricated data each trial.

Participants were instructed that they would be looking at a number of fictional scenarios in which a medical lab had developed a new medication to treat an ailment. The lab's researchers were each said to have administered the drug to 20 different patients to see how many developed negative side effects. Participants were then shown an example of the reports they would be viewing. Participants were informed that their task was to view the researchers' reports to determine the true underlying prevalence rate of negative side effects for each medication.

While reading the instructions for the task, participants were also made aware of two key points. First, participants were reminded that it is normal for lab members' results to differ from one another since they are different people testing different samples of patients. They were also informed that they may encounter reports from lab members who fabricated their results. Participants were told that fabricated results would be "much higher or much lower than the other reports for a given medication." Participants were specifically instructed to ignore any report that they believed was fabricated in this way. After reading the instructions, a comprehension question assessed whether the participant understood the instructions, and how to determine which reports were fabricated. Any participant who failed this check was removed from the experiment and was not permitted to begin the task.

At the beginning of each trial, participants read a brief story to orient them to the scenario. For example, "Imagine that a lab has developed a pill named Corfenib for treatment of sinus infections. While testing the prevalence of negative side effects, different lab members reported the following results:". Imaginary medication names were created for this experiment to ensure participants could not utilize existing information about real medications to inform their estimates. The medication and condition names were randomly selected each trial, as the specific names were not of interest to the current hypotheses.

After reading the imaginary scenario, the participant was shown the first fictional researcher's report. Each report was presented as a number such as "8 out of 20." Each report remained on the screen for two seconds, after which participants pressed a key to either see another report, or provide their estimate. After seeing the first report, the participant was then shown a screen that allowed them to press the right arrow on the keyboard to see another report, or the spacebar to progress to a screen where they could indicate their estimate and then move on to the next trial. In this way, participants could sample as many reports per trial as they wanted (up to a maximum of 51) until they felt comfortable estimating the true side effect prevalence

rate of the medication. When they were ready to indicate their estimate, they pressed the spacebar and typed in their answer on a response screen. Participants in the Estimate and Confidence group were then asked to rate how likely it was that they encountered fabricated data on that trial on a Likert scale (1 = "Not very likely" to 5 = "Very likely"). Participants repeated this sequential sampling of the reports, choosing their stopping point, and estimating the true prevalence rate of negative side effects for each of 52 scenarios. Participants were given the option of taking a short break after completing half the trials.

Results

All analyses described in this paper were conducted using jamovi computer software (Jamovi project, 2021; R Core Team, 2020). All mixed effects models were fit using the GAMLj jamovi module (Gallucci, M., 2019; Ripley, B., Venables, W., Bates, D. M., Hornik, K., Gebhardt, A., & Firth, D., 2018).

To assess whether the presence of outliers influenced the accuracy of participants' estimates, the estimates were converted to an error score. First, the mean of all values a participant viewed on a given trial excepting any outlier was computed. Then, that mean was subtracted from the participant's estimate for that trial. Thus, the error value indicated the difference between participants' estimates and the true mean underlying a given trial. A positive error value indicated an overestimation, and a negative error value indicated an underestimation.

A linear mixed effects regression model was then constructed to predict participants' error scores. This model estimated an intercept and included Group (Estimate-only, Estimate+Confidence) and Trial Type (No outlier, Low Outlier, High Outlier) as predictors, as well as the Group*Trial Type interaction. We also included by-subject random intercepts.[1] Note that the Catch trials were not included in this analysis since they served only as distractors and were not of interest to the hypotheses of this experiment.

The model indicated that Trial Type significantly affected the accuracy of participants' estimates, $F(2, 2481.9)=53.92$, $p<.001$.[2] Planned comparisons revealed that participants over-estimated when they viewed an outlier higher than the underlying mean (error M=0.398,

[1] By-subject random slopes were not included due to model convergence issues.
[2] Degrees of freedom estimated via GAMLj module in jamovi

SD=1.16) relative to trials without an outlier (error M=0.028, SD=1.03), *b*= 0.371, 95% CI [0.273, 0.468], t(2483)=7.42, p<.001. Participants also under-estimated when they encountered an outlier lower than the underlying mean (error M= -0.198, SD=1.04) relative to trials without an outlier (error M=0.028, SD=1.03), *b*= -0.229, 95% CI [-0.328, -0.131], t(2483)=-4.58, p<.001 (see Figure 2).

The model also revealed that asking participants to provide outlier detection confidence ratings affected estimate accuracy, F(1, 88.8)=7.38, p=.008. Estimates in the Detection+Estimate group (error M= -0.061, SD=1.14) were lower than those in the Estimate-only group (error M=0.147, SD=1.03). However, there was no significant interaction between group and trial type, F(2, 2481.9)=0.752, p=.471.

To establish whether participants were able to detect outliers when they were present, an ordinal logistic regression of outlier detection confidence was conducted. This model predicted the confidence ratings provided by the Estimate + Confidence group, using trial type as a predictor. Confidence ratings were shown to be significantly predicted by trial type. Specifically, when participants encountered an outlier higher than the underlying mean, they reported greater confidence on the 5 point Likert-type scale than when they did not encounter an outlier, *b*= 1.29, 95% CI, [1.01, 1.57], Z=9.05, p<.001. Similarly, when they encountered an outlier lower than the underlying mean, participants' detection confidence ratings were greater on average compared to no-outlier trials, *b*= 1.59, 95% CI [1.30, 1.88], Z=10.76, p<.001. This indicates that participants were more likely to be confident that they detected an outlier when they encountered one, suggesting that participants were generally able to detect outliers when they were present.

The impact of outliers on participants' estimates was not eliminated with their detection, however. We examined trials where a.) an outlier was shown and b.) participants were maximally confident that they detected the outlier. A linear mixed effects regression model was constructed to predict error scores in this subset of trials, using trial type as a predictor along with by-subject random intercepts. The model revealed a significant difference in estimate accuracy in high outlier (error M= 0.335, SD=1.56) and low outlier (error M= -0.164, SD=1.18) trials, *b*= 0.459, 95% CI [0.118, 0.801], t(208.3)=2.637, p=.009.

Finally, to investigate whether encountering fabricated data affected the amount of information participants sought, a linear mixed effects model was constructed to predict the number of reports participants viewed before providing their estimates. This model included trial

type, group (confidence+estimate, estimate-only), and the trial type*group interaction as predictors along with by-subject random intercepts. The model revealed that participants viewed significantly more reports when they encountered outliers above the underlying mean (M= 10.200, SD= 5.880) than on trials without outliers (M= 9.080, SD= 6.15), *b*= 0.646, 95% CI [0.237, 1.060], t(2475.2)=3.094, p=.002. Similarly, participants viewed more images when encountering a low outlier (10.300, SD= 6.08) than no outlier at all, *b*= 0.810, 95% CI [0.399, 1.220], t(2475.2)=3.864, p<.001. Group did not significantly predict the number of reports participants viewed (*b*= -0.143, 95% CI [-2.101, 1.820], t(83.7)= -2.637, p=.887.

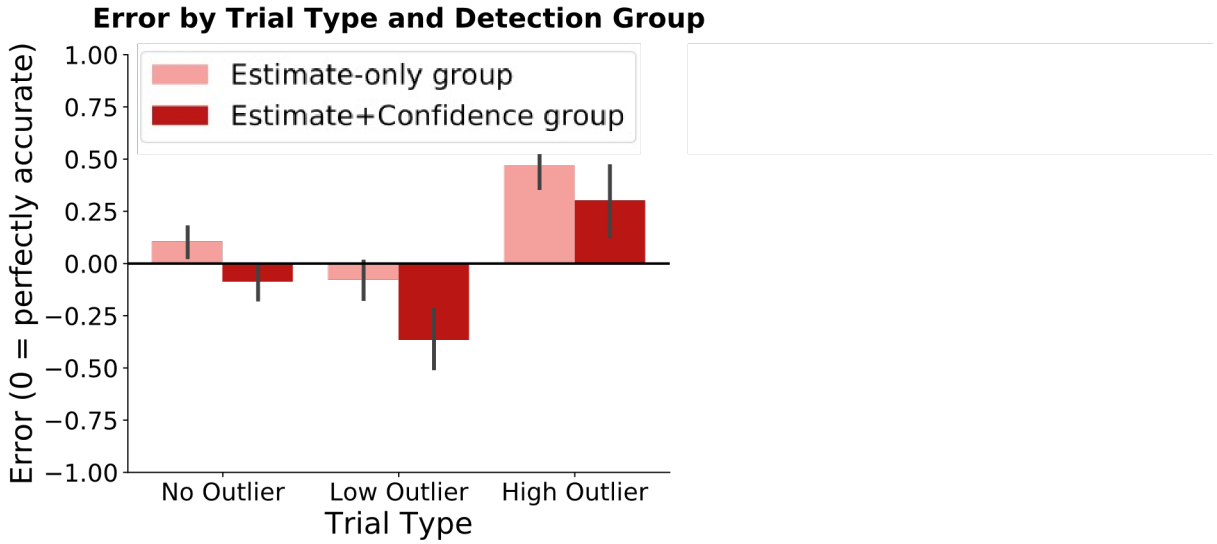

Figure 2

Mean error in each of the experimental conditions of Experiment 1. For each trial, error was computed by first determining the mean of all report values a participant saw in that trial, minus any outliers. This mean was then subtracted from the estimate the participant provided. Thus error is the difference between the participant's estimate and the mean of all non-fabricated reports they viewed on a trial, with positive values indicating an overestimation and negative values indicating underestimation. Overall, estimates were biased in the direction of outliers when they were present. Error bars represent the 95% CI of the means.

Conclusion

Greater outlier detection confidence was associated with greater accuracy, however participants' estimates were still biased in the direction of outliers when they were present. This bias persisted even when participants were most confident that they detected the fabricated data, suggesting that detecting the outliers was not sufficient to fully ignore them.

CHAPTER 3

EXPERIMENT 2

Experiment 2 investigated the use of visual warning cues to alert participants of upcoming fabricated data. Since our previous results indicated that increased outlier detection confidence was associated with greater accuracy, we hypothesized that these warnings would help participants feel more confident in ignoring outliers.

Method

Participants

107 adult participants were recruited online via Amazon's Mechanical Turk platform. The study was approved by the Institutional Review Board at the authors' university. Similar to Experiment 1, we intended to have approximately 50 participants per condition and the experiment had two between-subjects conditions. The sample size was determined prior to starting recruitment and data was analyzed only after all data had been collected. Each participant received $1.00 upon completion of the approximately 45-minute task. As in Experiment 1, participants whose median number of stimuli viewed was less than three were excluded from analysis. After excluding 20 participants due to low median stimuli viewed per trial, the final data were comprised of the remaining 87 participants. Ages ranged from 21 to 72 years old (mean = 41.14±13.58), and the sample was comprised of 50 female, 34 male, and 2 non-binary participants, as well as one participant with an unknown gender.

Materials

In addition to the materials described in Experiment 1, this experiment included a visual warning cue at the beginning of some trials to alert participants that the trial might contain fabricated data. This cue appeared on the same screen as the initial short story that started each trial, and disappeared when the story was no longer on the screen. The cue was a red and white-colored, triangular warning sign with an exclamation point inside it (see Figure 3A). This design and color scheme was chosen to ensure participants did not miss the cue.

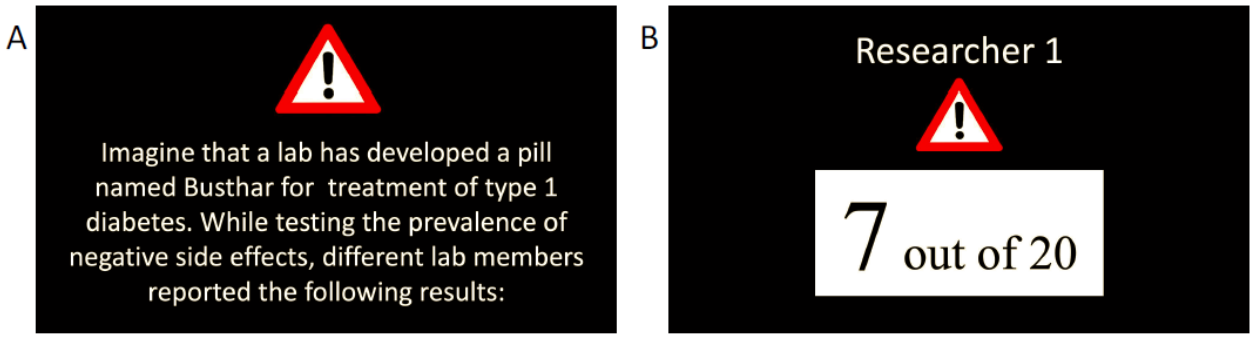

Figure 3

Panel A: An example of the orienting story that participants read in Experiments 2 and 3. The cue at the top of the screen warned that at least one outlier would appear in the reports for that trial. For some participants, this cue was 100% reliable. For other participants, the cue indicated there was a 70% chance that they would see an outlier in the upcoming reports. Participants were randomly assigned to a cue type before beginning the first trial. Panel B: The outlier cue employed in Experiment 5. In this experiment, one group of participants were shown the cue above any report that had been fabricated. The other group was not shown any cue during the task.

Procedure

This experiment (and all subsequent experiments) utilized a slightly modified version of the procedure described in Experiment 1. In this experiment, all participants indicated their confidence that they detected fabricated data after providing their estimates.

Participants were randomly assigned to one of two between-groups conditions before the experiment began. One group was shown warning cues that indicated the presence of upcoming fabricated data 100% reliably. The other group received cues that were 70% reliable.

Participants read instructions that informed them about the nature of the cues, including their reliability. A comprehension check was added to ensure participants understood the instructions. Any participant that did not pass this check was not permitted to begin the experiment.

As in Experiment 1, each trial began with a short story that described the scenario being assessed. In the 100% reliable cue group, the warning cue appeared in the space above the story on every test trial (i.e., the trials with outliers). In the 70% reliable group, the cue had a 70% chance of appearing on outlier trials and a 30% chance of appearing on non-outlier (i.e., control and catch) trials .

All other procedures were identical to those used in Experiment 1.

Results

To assess whether the presence of outliers influenced the accuracy of participants' estimates, the estimates were converted to an error score as in Experiment 1 (error = estimate – mean of non-fabricated values seen). Thus, a positive error value indicated an overestimation, and a negative error value indicated an underestimation.

A linear mixed effects regression model was then constructed to predict participants' error scores. This model estimated an intercept and included Cue Group (70% Cue, 100% Cue) and Trial Type (No outlier, Low Outlier, High Outlier) as predictors, as well as the Cue Group*Trial Type interaction. We also included by-subject random intercepts. Note that the Catch trials were not included in this analysis since they served only as distractors and were not of interest to the hypotheses of this experiment.

The model indicated that Trial Type significantly affected the accuracy of participants' estimates, $F(2, 2493)=55.648$, $p<.001$. Planned comparisons revealed that participants over-estimated when on trials with an outlier higher than the underlying mean (error M=0.335, SD=1.06) relative to trials without an outlier (error M= -0.033, SD=0.992), $b=0.370$, 95% CI [0.278, 0.461], $t(2494.3)=7.898$, $p<.001$. Participants also under-estimated on trials with an

outlier lower than the underlying mean (error M= -0.225, SD=1.05) relative to trials without an outlier (error M= -0.033, SD=0.992), *b*= -0.190, 95% CI [-0.281, -0.098], t(2494.2)= -4.057, p<.001 (see Figure 4A). The model also revealed that there was no significant difference in accuracy between the 70% Cue group and the 100% Cue group, F(1, 84)=1.482, p=.227.

To establish whether participants were able to detect outliers when they were present, an ordinal logistic regression of outlier detection confidence was conducted. This model predicted the confidence ratings using trial type and cue group predictors. Confidence ratings were shown to be significantly predicted by trial type. Specifically, when participants encountered an outlier higher than the underlying mean, they they reported higher confidence on the 5 point Likert scale than when they did not encounter an outlier, *b*= 1.92, 95% CI, [1.732, 2.110], Z=19.90, p<.001. Similarly, when they encountered an outlier lower than the underlying mean, participants' detection confidence ratings were higher on average compared to no-outlier trials, *b*= 2.328, 95% CI [2.132, 2.526], Z=23.16, p<.001. Similar to Experiment 1, these results indicate that participants were more confident that they detected an outlier when they encountered one, suggesting that participants were generally able to detect outliers when they were present.

Cue group also significantly predicted participants' outlier detection confidence ratings. On average, ratings were higher in the 100% cue group than the 70% cue group, *b*= 0.162, 95% CI [0.021, 0.304], Z=2.240, p=.025. This indicates that participants who saw the 100% reliable cue were better able to detect the outliers when they were present than the group that saw the less reliable cue.

As in Experiment 1, the impact of outliers on participants' estimates was not eliminated with their detection. We examined trials where a) an outlier was shown and b) participants were maximally confident that they detected the outlier. A linear mixed effects regression model was constructed to predict error scores in this subset of trials, using trial type, cue group, and the trial type*cue group interaction as predictors along with by-subject random intercepts. The model revealed a significant difference in estimate accuracy in high outlier (error M= 0.282, SD=1.110) and low outlier (error M= -0.178, SD=0.971) trials, *b*= 0.445, 95% CI [0.285, 0.605], t(563.5)=5.463, p<.001. Accuracy did not differ between cue groups in this subset of the data, *b*=0.059, 95% CI [-0.214, 0.331], t(72.8)=0.422, p=.674.

Finally, to investigate whether encountering fabricated data affected the amount of information participants sought, a linear mixed effects model was constructed to predict the

number of reports participants viewed before providing their estimates. This model included trial type, cue group, and the trial type*cue group interaction as predictors along with by-subject random intercepts. The model revealed that participants viewed significantly more reports when they encountered outliers above the underlying mean (M= 10.300, SD= 5.400) than on trials without outliers (M= 9.500, SD= 5.410), *b*= 0.641, 95% CI [0.288, 0.993], t(2490.2)=3.564, p<.001. However, there was no significant difference in reports viewed between low outlier (M=9.670, SD=4.850) and no-outlier trials, *b*= -0.055, 95% CI [-0.407, 0.297], t(2475.2)= -0.305, p=.760. Cue group did not significantly predict the number of reports participants viewed (*b*=1.162, 95% CI [-0.501, 2.825], t(81.6)=1.370, p=.174.

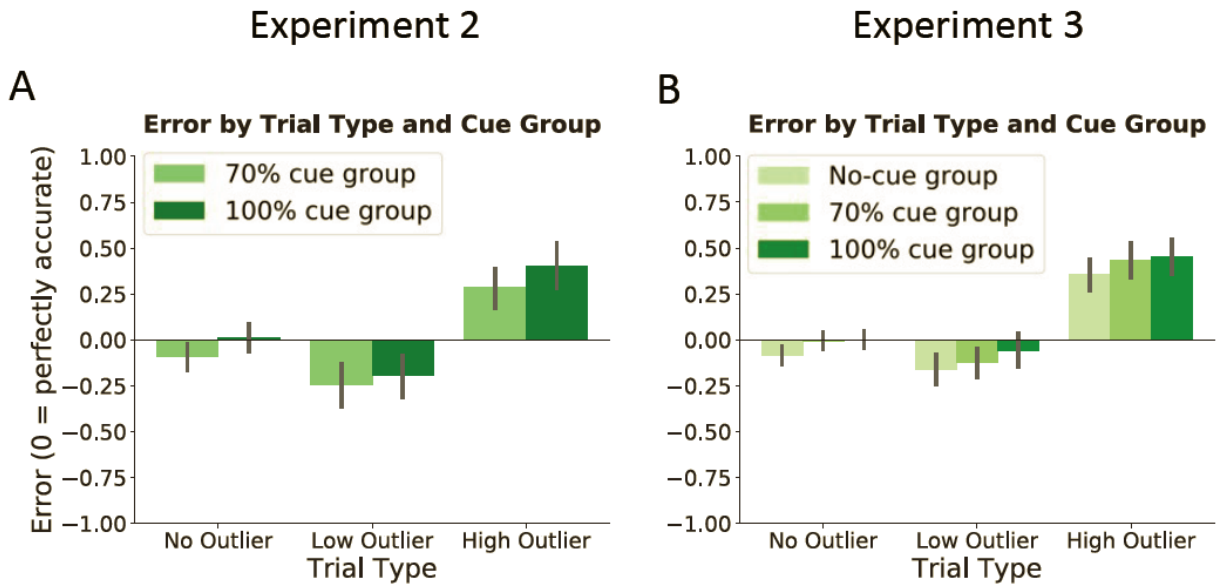

Figure 4

Mean errors for Experiments 2 and 3. Panel A: Participants' estimates in Experiment 2 were biased in the direction of an outlier when it was present. This bias was present in both the 70% and 100% cue groups. Panel B: Results from a pre-registered replication of Experiment 2, including a group which was shown no warning cues. Results from Experiment 2 were replicated, and the no-cue group's estimates were also biased in the direction of outliers. Error bars represent the 95% CI of the means.

Conclusion

The main findings from Experiment 1 were replicated. Furthermore, participants' ability to detect fabricated data was greater if they saw a 100% reliable warning cue than a 70% cue at the start of outlier trials. However, this increase in detection was not sufficient to fully eliminate the impact of outliers on participants' estimates.

CHAPTER 4

EXPERIMENT 3

To confirm the findings of the previous experiments, a pre-registered replication of Experiment 2 was conducted, with a larger sample size.

Method

Participants

382 adult participants were recruited online via Amazon's Mechanical Turk platform. The study was approved by the Institutional Review Board at the authors' university. Sample size for this experiment was increased due to the number of participants that were excluded from Experiments 1 and 2 due to insufficient information search. We intended to have approximately 50 participants per condition and the experiment had 6 between-subjects conditions. The sample size was determined prior to starting recruitment and was pre-registered on AsPredicted.org (#54565) The data were analyzed only after all data had been collected. Each participant received $1.50 upon completion of the approximately 45-minute task. As in the previous experiments, participants whose median number of stimuli viewed was less than three were excluded from analysis. After excluding 68 participants due to low median stimuli viewed per trial, the final data were comprised of the remaining 314 participants. The minimum number of participants in any cell of the experimental design was 45. Ages ranged from 18 to 83 years old (mean = 43.75±14.25), and the sample was comprised of 197 female, 112 male, 1 non-binary, 1 genderqueer, and 1 transmasculine participant, as well as 2 participants of unknown genders.

Materials

The materials used were identical to those used in Experiment 2.

Procedure

Nearly all procedures were identical to Experiment 2, except we added a third group which did not see any warning cues to serve as control. Since this group did not see the cues, their instructions made no mention of cues, and thus did not complete the cue-based comprehension check. We also controlled for the order of responses (estimate and outlier detection confidence) by manipulating response order across participants.

Results

The following analyses were preregistered on AsPredicted.org (#54565). As in Experiments 1 and 2, estimates were converted to an error score (error = estimate – mean of non-fabricated values seen). A linear mixed effects regression model was then constructed to predict participants' error scores. This model estimated an intercept and included Cue Group (No Cue, 70% Cue, 100% Cue), Trial Type (No outlier, Low Outlier, High Outlier), Response Order (Estimate First, Confidence first) and all resulting interactions as predictors. We also included by-subject random intercepts. As before, Catch trials were not included in this analysis.

The model indicated that Trial Type significantly affected the accuracy of participants' estimates, $F(2, 9212)=167.533$, $p<.001$. Planned comparisons revealed that participants over-estimated when they viewed an outlier higher than the underlying mean (error M=0.415, SD=1.27) relative to trials without an outlier (error M= -0.032, SD=1.06), $b=0.437$, 95% CI [0.383, 0.490], $t(9216)=16.125$, $p<.001$. Participants also under-estimated when they encountered an outlier lower than the underlying mean (error M= -0.114, SD=1.14) relative to trials without an outlier (error M= -0.032, SD=1.06), $b= -0.085$, 95% CI [-0.138, -0.032], $t(9223)= -3.128$, $p=.002$ (see Figure 4B). The model also revealed that there was no significant difference in accuracy between the No Cue, 70% Cue, and the 100% Cue groups. F(2,

323)=1.101, p=.334. Additionally, response order was not significantly associated with accuracy, F(1, 323)=0.563, p=.454.

To establish whether participants were able to detect outliers when they were present, an ordinal logistic regression of outlier detection confidence was conducted. This model predicted the confidence ratings using trial type, cue group, and response order as predictors. Participants again reported greater detection confidence when they saw a low (*b*= 2.199, 95% CI [2.098, 2.300], Z=42.527, p<.001) or high outlier (*b*= 0.583, 95% CI [0.472, 0.694], Z=10.355, p<.001) relative to no-outlier trials. Additionally, participants who were shown the 70% reliable cue (*b*=0.175, 95% CI [0.086, 0.264], Z=3.863, p<.001) or the 100% reliable cue (*b*= 0.196, 95% CI [0.105, 0.286], Z=4.221, p<.001) were better able to detect outliers than those who saw no cue. Response order did not significantly predict outlier detection confidence, *b*= 0.019, 95% CI [-0.054, 0.093], Z=0.514, p=.607.

As in previous experiments, the impact of outliers on participants' estimates was not eliminated with their detection. We examined trials where a) an outlier was shown and b) participants were maximally confident that they detected the outlier. A linear mixed effects regression model was constructed to predict error scores in this subset of trials, using trial type, cue group, response order, and the resulting interactions as predictors, along with by-subject random intercepts. The model revealed a significant difference in estimate accuracy in high outlier (error M= 0.357, SD=1.26) and low outlier (error M= -0.092, SD=1.13) trials, *b*= 0.456, 95% CI [0.354, 0.557], t(1827)=8.816, p<.001. Accuracy did not differ between cue groups (F(2, 207)=0.277, p=.758) or response order groups (F(1, 208)=0.118, p=.732) in this subset of the data.

Finally, to investigate whether encountering fabricated data affected the amount of information participants sought, a linear mixed effects model was constructed to predict the number of reports participants viewed before providing their estimates. This model included trial type, cue group, response order, and the resulting interactions as predictors, along with by-subject random intercepts. The model revealed that participants viewed significantly more reports when they encountered outliers above the underlying mean (M= 10.60, SD= 6.36) than when no outlier was present (M= 9.67, SD= 6.25), *b*= 0.575, 95% CI [0.380, 0.770], t(9189)=5.768, p<.001. Similarly, participants viewed more images when encountering a low outlier (M=10.40, SD=6.41) than when no outlier was present, *b*= 0.317, 95% CI [0.122, 0.513],

t(9190)=3.178, p=.001. Neither cue group (F(2, 310)=0.343, p=.710) nor response order (F(1, 310)<.001, p=.992) significantly predicted the number of reports participants viewed.

Conclusion

This preregistered experiment replicated the findings of Experiments 1 and 2. Additionally, we observed that both the 100% and 70% reliable cues increased fabricated data detection ability compared to a no-cue control group, however this was not sufficient to eliminate the biasing effect of outliers. Finally, asking participants to provide detection confidence ratings prior to their estimates did not influence the effect.

CHAPTER 5

EXPERIMENT 4

This pre-registered experiment examined whether the task scenario affected participants' ability to ignore outliers. Additionally, this experiment used larger stimuli values to investigate whether the pattern of effects would be observed with different outlier values.

Method

Participants

134 adult participants were recruited via Amazon's Mechanical Turk platform. The study was approved by the Institutional Review Board at the authors' university. Similar to Experiment 3, we intended to have approximately 50 participants per condition and the experiment had two between-subjects conditions. The sample size was determined prior to starting recruitment and was pre-registered on AsPredicted.org (#64615) The data were analyzed only after all data had been collected. Each participant received $1.50 upon completion of the approximately 45-minute task. As in the previous experiments, participants whose median number of stimuli viewed was less than three were excluded from analysis. After excluding 31 participants due to low median

stimuli viewed per trial, the final data were comprised of the remaining 103 participants. The minimum number of participants in any group was 45. Ages ranged from 22 to 77 years old (mean = 42.51±13.47), and the sample was comprised of 76 female and 27 male participants.

Materials

The materials for this experiment were constructed mostly in the manner described in Experiment 1, however greater values were used for the stimuli. The reports for this experiment indicated a number between 1 and 35 sampled from Gaussian distributions (e.g. "20 out of 35"). The Control trials were constructed by sampling from a Gaussian with mean of 20 and a standard deviation of 2. No reports indicating a value further than 2 SD from the mean were included in these trial lists.

The Test trials were again created by following the same procedure for generating the Control trials, and then randomly inserting an outlier as report three, four, or five. Low outliers were 13, 14, and 15. High outliers were 25, 26, and 27. We designed the stimuli so that the low outliers in Experiment 4 had the same value as high outliers in Experiments 1-3.

As in previous experiments, Catch trials were designed to ensure that there was sufficient variety in the stimuli so that participants would not simply learn that the mean report value of the test and control trials was 20. There were 4 sets of 5 filler trials, with means of either 17 or 23 and SDs of either 1 or 2.

Procedure

This experiment mainly utilized the procedure from Experiment 2, except a different between-group manipulation was employed. Before the experiment began, participants were assigned to one of two between-groups conditions. One group assessed the rate of negative side effects, as in the previous studies. The other group instead assessed the proportion of patients that would show improved health for each medication. For this group, any language referencing side effects was modified to describe health improvements. This included the instructions, the short story at the start of each trial, and the response screens. Only the language was

manipulated; the program that governed the presentation of numerical stimuli was identical for both groups.

Results

The following analyses were preregistered on AsPredicted.org (#64615). As in the previous experiments, the estimates were converted to an error score (error = estimate – mean of non-fabricated values seen). A linear mixed effects regression model was then constructed to predict this score. This model estimated an intercept and included Task Scenario (Side Effects, Health Improvements), Trial Type (No outlier, Low Outlier, High Outlier), and the Task Scenario*Trial Type interaction as predictors. We also included by-subject random intercepts. Note that the Catch trials were not included in this analysis, similar to previous experiments.

The model indicated that Trial Type significantly affected the accuracy of participants' estimates, $F(2, 3048)=77.615$, $p<.001$. Planned comparisons revealed that participants over-estimated when they viewed an outlier higher than the underlying mean (error M=0.254, SD=1.390) relative to trials without an outlier (error M= -0.160, SD=1.070), $b=0.418$, 95% CI [0.318, 0.517], $t(3050)=8.197$, $p<.001$. Participants also under-estimated when they encountered an outlier lower than the underlying mean (error M= -0.476, SD=1.280) relative to trials without an outlier (error M= -0.160, SD=1.070), $b= -0.320$, 95% CI [-0.419, -0.220], $t(3047)= -6.305$, $p<.001$ (see Figure 5A). The model also revealed that there was no significant difference in accuracy between the task scenarios, $F(1, 104)=0.542$, $p=.463$.

To establish whether participants were able to detect outliers when they were present, an ordinal logistic regression of outlier detection confidence was conducted. This model predicted the confidence ratings using trial type and task scenario as predictors. Confidence ratings were shown to be significantly predicted by trial type. Specifically, when participants encountered an outlier higher than the underlying mean, they reported higher confidence than when they did not encounter an outlier, $b= 1.399$, 95% CI, [1.237, 1.562], $Z=16.88$, $p<.001$. Similarly, participants reported higher confidence when they encountered an outlier lower than the underlying mean compared to no-outlier trials, $b= 1.585$, 95% CI [1.423, 1.750], $Z=19.05$, $p<.001$. This again indicates that participants were more likely to be confident that they detected an outlier when

they encountered one, suggesting that participants were generally able to detect outliers when they were present.

Task scenario also significantly predicted participants' outlier detection confidence ratings. On average, ratings were higher in the side effects group than the benefits group, *b*= 0.175, 95% CI [0.049, 0.302], Z=2.71, p=.007. This suggests that participants who judged side effect rates were better able to detect the outliers than those who judged health improvement rates.

As in previous experiments, the impact of outliers on participants' estimates was not eliminated with their detection. We examined trials where a) an outlier was shown and b) participants were maximally confident that they detected the outlier. A linear mixed effects regression model was constructed to predict error scores in this subset of trials, using trial type, task scenario, and the trial type*task scenario interaction as predictors, along with by-subject random intercepts. The model revealed a significant difference in estimate accuracy between high outlier (error M= 0.149, SD=1.53) and low outlier trials (error M= -0.449, SD=1.25), *b*= 0.628, 95% CI [0.343, 0.913], t(355.3)=4.321, p<.001.

The model also revealed a main effect of task scenario on accuracy in this subset of the data. On average, error scores were higher in the side effects group (error M= -0.037, SD=1.26) than in the benefits group (error M= -0.331, SD=1.60), *b*= 0.327, 95% CI [0.010, 0.646], t(45.8)=2.015, p=.050. Importantly, there was no significant interaction between trial type and task scenario, *b*= 0.206, 95% CI [-0.364, 0.776], t(355.3)=0.708, p=.479.

Finally, to investigate whether encountering fabricated data affected the amount of information participants sought, a linear mixed effects model was constructed to predict the number of reports participants viewed before providing their estimates. This model included trial type, task scenario, and the trial type*task scenario interaction as predictors, along with by-subject random intercepts. The model revealed that participants viewed significantly more reports when they encountered outliers above the underlying mean (M= 10.30, SD= 5.60) than when no outlier was present (M= 9.59, SD= 5.89), *b*= 0.386, 95% CI [0.065, 0.707], t(3045)=2.359, p=.018. Similarly, participants viewed more images when encountering a low outlier (M=10.40, SD=6.06) than when no outlier was present, *b*= 0.582, 95% CI [0.262, 0.901], t(3045)=3.569, p<.001. Number of images viewed did not differ between the side effects and benefits groups, *b*= -0.313, 95% CI [-2.136, 1.510], t(102)= -0.336, p=.737.

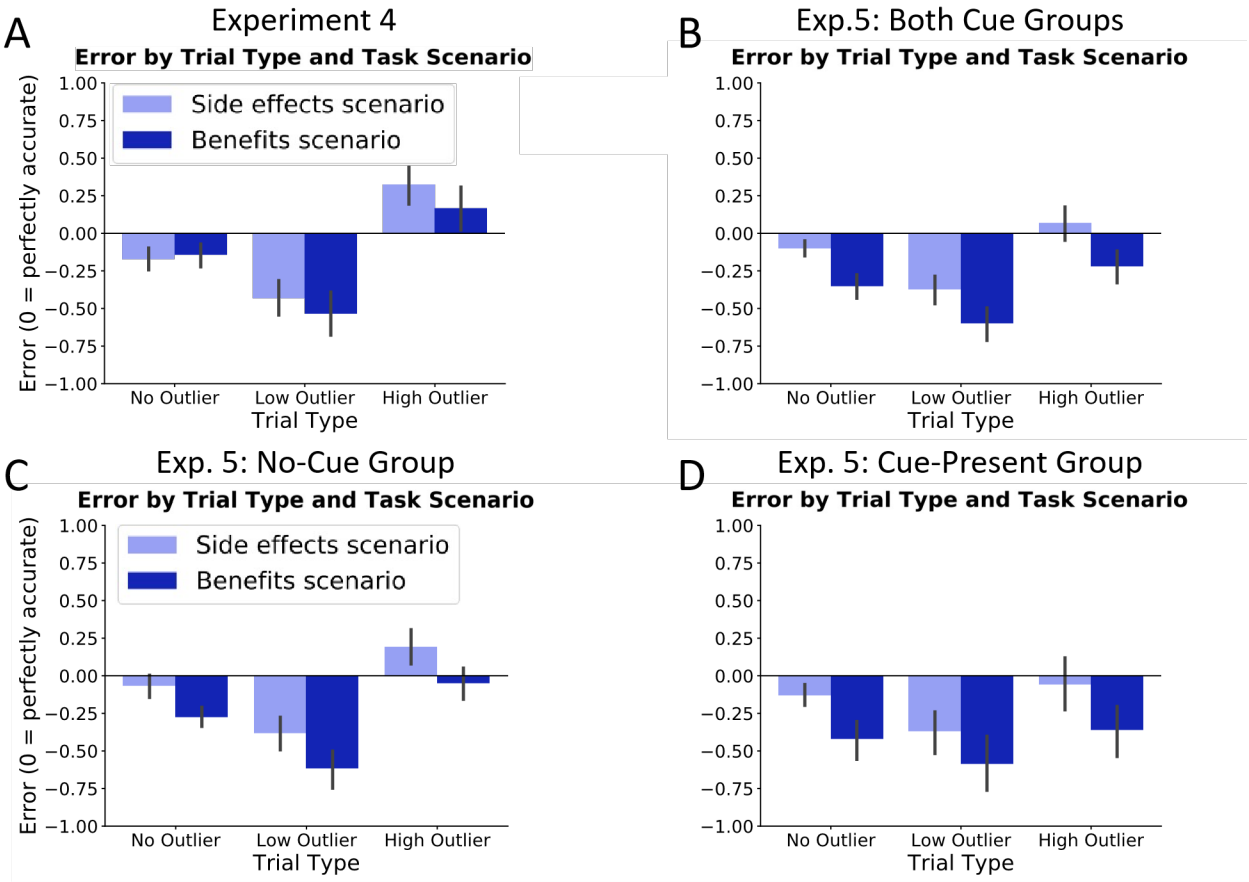

Figure 5

Mean errors for Experiments 4 and 5. Panel A: Estimates in Experiment 4 were biased towards outliers, indicating participants were not able to fully ignore them. This was true regardless of whether participants were estimating the underlying rates of health improvements or negative side effects. Panel B: Mean errors for Experiment 5 for both cue groups combined. Panel C: Results for the no-cue group in Experiment 5. Panel D: Results for the cue-present group in Experiment 5. Error bars represent the 95% CI of the means.

Conclusion

Despite the increased magnitude of the stimuli, participants' estimates were again biased in the direction of outliers, replicating the findings from Experiments 1-3. Additionally, task scenario did not have much effect on participants' ability to ignore outliers. In trials with the highest detection confidence, we observed a main effect of task scenario, but no interaction.

Thus, while the Health Benefits group provided lower estimates overall, both groups' estimates were biased in the direction of outliers.

CHAPTER 6

EXPERIMENT 5

To confirm the results from Experiment 4, we conducted a pre-registered replication. To further investigate whether the biasing effect of fabricated data could be eliminated, we added a stronger cuing manipulation than those employed in Experiments 2&3, in which individual outlier stimuli were cued.

Method

Participants

274 adult participants were recruited via Amazon's Mechanical Turk platform. The study was approved by the Institutional Review Board at the authors' university. Similar to Experiments 3 and 4, we intended to have approximately 50 participants per condition and the experiment had four between-subjects conditions. The sample size was determined prior to starting recruitment and was pre-registered on AsPredicted.org (#67768) The data were analyzed only after all data had been collected. Each participant received $1.50 upon completion of the approximately 45-minute task. As in the previous experiments, participants whose median number of stimuli viewed was less than three were excluded from analysis. 43 participants were excluded due to low median stimuli viewed per trial. An additional participant who did not understand the task provided an estimate of zero whenever a cue was presented. This participant's data were excluded (this exclusion was not pre-registered). The final data were comprised of the remaining 230 participants. The minimum number of participants in any cell was 52. Ages ranged from 19 to 78 years old (mean = 40.94±14.42), and the sample was comprised of 156 female, 72 male, and one non-binary participant, as well as one participant with an unknown gender.

Materials

This experiment utilized the same stimuli values from Experiment 4, as well as the side effect/health improvement manipulation. In addition, this experiment also added the visual warning cue from Experiments 2 and 3 (see Figure 3B) above any outlier inserted into the trials. The cue appeared simultaneously above the report it was flagging, and disappeared when the report left the screen. Note that is different from Experiments 2 and 3 where the cue was only shown at the start of the trial. Importantly, this cue was 100% accurate and always indicated that the flagged report had been fabricated.

Procedure

Before the experiment began, participants were randomly assigned to either the side effects or health improvement assessment group as in Experiment 4. Participants were also randomly assigned to one of two cue conditions. One group did not see any cues at all, which serves as a replication of Experiment 4. The other group was shown visual warning cues above any report that had been fabricated.

Participants were provided instructions with language appropriate to their condition assignment. Participants who would see the warning cues were instructed that they would see the triangular cue above any report that had been fabricated. They were informed that the cue was 100% accurate, and that they should ignore any report with which the cue appeared. A comprehension question ensured that participants understood what the cue meant. Any participant who did not pass the comprehension check was excluded from participation.

Participants assigned to the No Cue group completed the trials exactly as described in Experiment 4. In the Cue Present group, the visual warning cue appeared on screen whenever an outlier was shown and remained on the screen as long as the report was on the screen (2 seconds). All other procedures were identical to Experiment 4.

Results

The following analyses were preregistered on AsPredicted.org (#67768). As in the previous experiments, estimates were converted to an error score (error = estimate – mean of non-fabricated values seen). A linear mixed effects regression model was then constructed to predict participants' error scores. This model estimated an intercept and included Task Scenario (Side Effects, Health Improvements), Trial Type (No outlier, Low Outlier, High Outlier), Cue Group (No Cue, Cue Present), and all resulting interactions as predictors. We also included by-subject random intercepts. Note that the Catch trials were not included in this analysis, similar to the other experiments.

The model indicated that Trial Type significantly affected the accuracy of participants' estimates, $F(2, 6800)=54.427$, $p<.001$. Planned comparisons revealed that participants over-estimated when they viewed an outlier higher than the underlying mean (error M= -0.078, SD=1.570) relative to trials without an outlier (error M= -0.225, SD=1.420), $b= 0.161$, 95% CI [0.094, 0.229], $t(6801)=4.678$, $p<.001$. Participants also under-estimated when they encountered an outlier lower than the underlying mean (error M= -0.488, SD=1.480) relative to trials without an outlier (error M= -0.225, SD=1.420), $b= -0.255$, 95% CI [-0.322, -0.187], $t(6801)= -7.358$, $p<.001$ (see Figure 5B). The model also revealed that on average, error scores were higher in the side effects group (error M= -0.125 , SD= 1.30) than the benefits group (error M= -0.380, SD=1.64), $b= 0.264$, 95% CI [0.024, 0.504], $t(230)=2.160$, $p=.032$.

There was also a significant interaction between trial type and cue group, $F(2, 6800)=6.800$, $p=.001$. An examination of the simple effects in Table 1 reveals that error scores in the No Cue group show the pattern of over- and under-estimation observed in the previous experiments. However, the group for which each outlier was clearly cued only marginally showed over-estimation in high outlier trials relative to no-outlier trials, $b= 0.083$, 95% CI [-0.010, 0.176], $t(6803)=1.760$, $p=.079$.

Table 1

Simple effects parameter estimates for Experiment 5, trial type

			95% CI				
Cue Group	Contrast	Estimate	Lower	Upper	t	df	p
No cue	High Outlier - Control	0.239	0.141	0.337	4.77	6799	<.001
	Low Outlier - Control	-0.325	-0.424	-0.227	6.46	6801	<.001
Cue present	High Outlier - Control	0.083	-0.010	0.176	1.76	6803	.079
	Low Outlier - Control	-0.184	-0.277	-0.091	3.88	6801	<.001

To establish whether participants were able to detect outliers when they were present, an ordinal logistic regression of outlier detection confidence was conducted. This model predicted the confidence ratings using trial type, task scenario, and cue group as predictors. Confidence ratings were shown to be significantly predicted by trial type. Specifically, when participants encountered an outlier higher than the underlying mean, they reported greater confidence than when they did not encounter an outlier, b= 0.626, 95% CI, [0.518, 0.734], Z=11.38, p<.001. Similarly, when they encountered an outlier lower than the underlying mean, participants' detection confidence ratings were higher on average compared to no-outlier trials, b= 0.692, 95% CI [0.583, 0.802], Z=12.43, p<.001. This again indicates that participants were more likely to be confident that they detected an outlier when they encountered one, suggesting that participants were generally able to detect outliers when they were present.

Cue group also significantly predicted participants' outlier detection confidence ratings. On average, ratings were higher in the Cue Present group than the No Cue group, b=1.042, 95% CI [0.953, 1.130], Z=23.08, p<.001. This suggests that participants who saw warning cues on all outiers were better able to detect the outliers than those who did not see any cues.
Task scenario did not significantly predict outlier detection confidence, b= -0.050, 95% CI [-0.136, 0.036], Z= -1.14, p=.255.

As in previous experiments, the impact of outliers on participants' estimates was not eliminated with their detection. We examined trials where a) an outlier was shown and b) participants were maximally confident that they detected the outlier. A linear mixed effects

regression model was constructed to predict error scores in this subset of trials, using trial type, task scenario, cue group, and all resulting interactions as predictors, along with by-subject random intercepts. The model revealed a significant difference in estimate accuracy between high outlier (error M= -0.027, SD=1.74) and low outlier (error M= -0.429, SD=1.46) trials, *b*= 0.502, 95% CI [0.364, 0.640], t(1486)=7.128, p<.001. The model also revealed a main effect of task scenario on accuracy. On average, error scores were higher in the side effects group (error M= -0.135, SD=1.74) than in the benefits group (error M= -0.326, SD=1.48), *b*= 0.392, 95% CI [0.006, 0.777], t(178)=1.993, p=.048.

There was also a significant interaction between trial type and cue group, *b*= -0.444, 95% CI [-0.721, -0.168], t(1486)= -3.154, p=.002. An examination of simple effects revealed that while error scores were significantly higher in high outlier trials than low outlier trials for both cue groups, the difference between high and low outlier trials is larger if no cue is shown (*b*= 0.724, 95% CI [0.485, 0.964], t(1502)=5.93, p<.001) than if all outliers are cued (*b*= 0.280, 95% CI [0.142, 0.418], t(1431)=3.99, p<.001).

Finally, to investigate whether encountering fabricated data affected the amount of information participants sought, a linear mixed effects model was constructed to predict the number of reports participants viewed before providing their estimates. This model included trial type, task scenario, cue group, and all resulting interactions as predictors, along with by-subject random intercepts. The model revealed that, on average, participants viewed significantly more reports when they encountered outliers above the underlying mean (M= 11.00, SD= 7.43) than when no outlier was present (M= 10.30, SD= 7.60), *b*= 0.354, 95% CI [0.110, 0.599], t(6796)=2.842, p=.004. Similarly, participants viewed more reports when encountering a low outlier (M=11.10, SD=7.72) than when no outlier was present, *b*= 0.448, 95% CI [0.203, 0.693], t(6796)=3.578, p<.001.

The model also revealed a significant interaction between trial type and cue group, F(2, 6796)=5.060, p=.006. Examination of the simple effects reveals that participants who saw no cues viewed more reports when an outlier was present (F(2, 6795)=12.191), however participants who were shown a cue with each outlier looked at the same number of reports whether or not an outlier was present, F(2, 6796)=0.180, p=.835.

There was also a significant interaction between trial type and task scenario, F(2, 6796)=4.335, p=.013. Examination of the simple effects reveals that participants in the side

effects group viewed more reports when a high outlier (*b*= 0.712, 95% CI [0.366, 1.058], t(6797)=4.036, p<.001) or low outlier (*b*= 0.484, 95% CI [0.136, 0.832], t(6797)=2.728, p=.006) was present relative to no-outlier trials. Participants in the benefits group, however, only viewed more reports when a low outlier was present, (*b*= 0.411, 95% CI [0.065, 0.757], t(6797)=2.331, p=.020. When a high outlier was present, this group viewed an equal number of reports as in the no-outlier trials, (*b*= -0.004, 95% CI [-0.349, 0.342], t(6797)= -0.020, p=.984.

Conclusion

Results from the no-cue group replicated those of Experiment 4, which are consistent with Experiments 1-3. However, the group who saw a cue with each outlier only marginally overestimated when they saw a high outlier. This suggests that cuing the fabricated data directly may have allowed participants to ignore it. It should be noted, however, that the cue did not eliminate underestimation in low outlier trials.

CHAPTER 7

GENERAL DISCUSSION

Across five experiments, participants' estimates were biased in the direction of outliers. This bias persisted across two different ranges of stimuli values. Overall, participants' information seeking slightly increased when they saw an outlier, but this did not affect the bias. Warnings of upcoming fabricated data increased participants' ability to detect outliers, but did not eliminate the influence of the fabricated data on estimates. When outliers were directly cued at presentation, participants still underestimated when they saw a low outlier, however overestimation in the presence of a high outlier was only marginal. Consistent with previous research on misinformation, it appears that numerical misinformation is difficult to ignore even with intervention.

These findings suggest that people may not be able to fully ignore fabricated data once they have encountered it. This is consistent with the continued influence effect (CIE) and the illusory truth literature, showing that it is difficult for individuals to ignore or disbelieve invalid

information that they have encountered (Hasher et al., 1977; Johnson and Seifert, 1994; Wistrich et al., 2004). Our work expanded the traditional CIE paradigm by having participants decide for themselves which pieces of information (data) were invalid. Even when participants actively detected fabricated data and made their final response shortly thereafter (approx. 10-25sec), they were still influenced by it.

One explanation of these findings could relate to memory retrieval errors. According to exemplar theories of memory, individuals in our task store instances in memory representing each stimulus they encounter. When participants are asked to make judgments, they sample values from memory and combine them to generate an estimate (André et al., 2021). In this account, fabricated data would be encoded, but would have a potentially lower (but non-zero) activation weight than other values encountered on a trial. When participants generate their estimates, the outlier values could be sampled and incorporated into the final judgment. producing estimates biased towards outliers.

It should be noted that the current experiments did not ascertain the exact moment participants detected an outlier. Participants only indicated their detection confidence at the end of the trial. Thus, we could not determine in which trials participants detected the outlier upon seeing it, which might determine whether they encode the value. Similarly, we cannot say in which trials participants encoded the outlier as valid, and later determined it to be fabricated after seeing more values in the set. Future work could probe participants' detection after each stimulus to investigate this issue.

Another possible explanation of our findings is that our outlier values were not surprising enough. Filipowicz et al. (2018) found that participants who would usually show increased belief updating to surprising stimuli instead showed no updating when new information was extremely surprising. They suggested that their participants may have judged information to be too surprising to be legitimate and ignored it. Our experiments used outliers that were 2.5 to 3.5 SD from the mean in order to mimic deceptive statistics that one may encounter in their daily life. Future work could employ more extreme outliers to investigate the boundary conditions for incorporating outliers into one's beliefs.

It should also be noted that our experiments utilized a narrow range of values, presented in a format that might not reflect data seen in the real world. For example, people often encounter data in the form of percentages, such as in demographic information cited by

politicians (Prévost and Beaud, 2015), or they encounter large numbers such as mortality predictions for the COVID-19 pandemic (Allyn, 2020). Future work should also focus on investigating whether our findings persist in different numerical formats and magnitudes, and with different task contexts.

The current experiments add to the literature concerning how we deal with misinformation. Much of the current CIE and illusory truth work examines how we share and spread misinformation, and how we may curtail that spread (Chan et al., 2017; Lewandowsky et al., 2012; Pennycook et al., 2021). Few studies have examined how people handle numerical misinformation (but see Stubenvoll and Matthes, 2021), and our work contributes to the literature by illuminating that people have difficulty ignoring misinformation in the form of fabricated data, even when they do not believe the data is legitimate. The influence of bad data on our beliefs suggests a vulnerability to sloppy data reporting or outright fabricated data/statistics in individuals' voting, consumer, and health decisions.

REFERENCES

Allyn, B. (2020, March 29). *Fauci estimates that 100,000 to 200,000 americans could die from the coronavirus*. https://www.npr.org/sections/coronavirus-live-updates/2020/03/29/823517467/fauciestimates-that-100-000-to-200-000-americans-could-die-from-the-coronavirus

André, Q., Reinholtz, N., & De Langhe, B. (2021). Can consumers learn price dispersion? evidence for dispersion spillover across categories. *Journal of Consumer Research*.

Begg, I. M., Anas, A., & Farinacci, S. (1992). Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General*, *121*(4), 446.

Brashier, N. M., Eliseev, E. D., & Marsh, E. J. (2020). An initial accuracy focus prevents illusory truth. *Cognition*, *194*, 104054.

Brashier, N. M., Pennycook, G., Berinsky, A. J., & Rand, D. G. (2021). Timing matters when correcting fake news. *Proceedings of the National Academy of Sciences*, *118*(5).

Chan, M.-p. S., Jones, C. R., Hall Jamieson, K., & Albarracın, D. (2017). Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological science*, *28*(11), 1531–1546.

Cook, J., & Lewandowsky, S. (2016). Rational irrationality: Modeling climate change belief polarization using bayesian networks. *Topics in cognitive science*, *8*(1), 160–179.

Ecker, U. K., Lewandowsky, S., Swire, B., & Chang, D. (2011). Correcting false information in memory: Manipulating the strength of misinformation encoding and its retraction. *Psychonomic bulletin & review*, *18*(3), 570–578.

Ecker, U. K., Lewandowsky, S., & Tang, D. T. (2010). Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory & cognition*, *38*(8), 1087–1100.

Effron, D. A., & Raj, M. (2020). Misinformation and morality: Encountering fake-news headlines makes them seem less unethical to publish and share. *Psychological Science*, *31*(1), 75–87.

Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General*, *144*(5), 993.

Fazio, L. K., Rand, D. G., & Pennycook, G. (2019). Repetition increases perceived truth equally for plausible and implausible statements. *Psychonomic bulletin & review*, *26*(5), 1705–1710.

Filipowicz, A., Valadao, D., Anderson, B., & Danckert, J. (2018). Rejecting outliers: Surprising changes do not always improve belief updating. *Decision*, *5*(3), 165.

Gallucci, M. (2019). *Gamlj: General analyses for linear models* [jamovi module]. https://gamlj.github.io/

Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of verbal learning and verbal behavior*, *16*(1), 107–112.

Hassan, A., & Barber, S. J. (2021). The effects of repetition frequency on the illusory truth effect.

Cognitive Research: Principles and Implications, *6*(1), 1–12.

Jamovi project. (2021). *Jamovi* (Version 1.6) [Computer Software]. https://www.jamovi.org

Johnson, H. M., & Seifert, C. M. (1994). Sources of the continued influence effect: When misinformation in memory affects later inferences. *Journal of experimental psychology: Learning, memory, and cognition*, *20*(6), 1420.

Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological science in the public interest*, *13*(3), 106–131.

London, K., & Nunez, N. (2000). The effect of jury deliberations on jurors' propensity to disregard inadmissible evidence. *Journal of Applied Psychology*, *85*(6), 932.

Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, *592*(7855), 590–595.

Prévost, J.-G., & Beaud, J.-P. (2015). *Statistics, public debate and the state, 1800–1945: A social, political and intellectual history of numbers*. Routledge.

R Core Team. (2020). *R: A language and environment for statistical computing* (Version 4.0) [Computer Software (R packages retrieved from MRAN snapshot 2020-08-24)]. https://www.R-project.org

Ripley, B., Venables, W., Bates, D. M., Hornik, K., Gebhardt, A., & Firth, D. (2018). *Mass: Support functions and datasets for venables and ripley's mass.* [R package]. https://cran.r-project.org/package=MASS

Schaaf, J. M., Bederian-Gardner, D., & Goodman, G. S. (2015). Gating out misinformation: Can young children follow instructions to ignore false information? *Behavioral sciences & the law*, *33*(4), 390–406.

Stubenvoll, M., & Matthes, J. (2021). Why retractions of numerical misinformation fail: The anchoring effect of inaccurate numbers in the news. *Journalism & Mass Communication Quarterly*, 10776990211021800.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151.

Westlund, O., & Hermida, A. (2021). Data journalism and misinformation. *The routledge companion to media disinformation and populism* (pp. 142–150). Routledge.

Wilkes, A., & Leatherbarrow, M. (1988). Editing episodic memory following the identification of error. *The Quarterly Journal of Experimental Psychology*, *40*(2), 361–387.

Wistrich, A. J., Guthrie, C., & Rachlinski, J. J. (2004). Can judges ignore inadmissible information-the difficulty of deliberately disregarding. *U. Pa. L. Rev.*, *153*, 1251.