Are You Serious?
Handling Disagreement When Annotating Conspiracy Theory Texts

Ashley Hemm, Sandra Kübler†, Michelle Seelig, John Funchion,
Manohar Murthi, Kamal Premaratne, Daniel Verdear, Stefan Wuchty
University of Miami, †Indiana University
ashleyhemm@miami.edu, skuebler@iu.edu,
{mseelig,jfunchion,mmurthi,kamal,dverdear,s.wuchty}@miami.edu

Abstract

We often assume that annotation tasks, such as annotating for the presence of conspiracy theories, can be annotated with hard labels, without definitions or guidelines. Our annotation experiments, comparing students and experts, show that there is little agreement on basic annotations even among experts. For this reason, we conclude that we need to accept disagreement as an integral part of such annotations.

1 Introduction

In typical linguistic annotation projects, such as for part of speech tags or for dependency syntax, we assume that there is a single correct analysis, which can be determined reliably by trained annotators (e.g. Kübler and Zinsmeister, 2014). For phenomena such as hate speech or conspiracy theories (CTs), we tend to follow the same model without reflecting on the tasks. We assume that we know what hate speech or CTs are, asking our annotators to proceed without providing a definition of the phenomenon or guidelines. Another way of creating corpora of annotated data consists of using search terms to find examples of the phenomenon, where the search term (or the website from which the example was retrieved) serves as an approximation of the gold standard annotation. For example, a research team from the RAND Corporation, investigating the automatic detection of "conspiracy theory language" (Marcellino et al., 2021), used search terms to find examples of CT texts from social media. They created a machine learning classifier for which such texts were to be distinguished from "a baseline sample of 'normal' non-conspiracy talk". The CT texts covered alien visitation, vaccine dangers, the origins of COVID-19, while the non-CT texts were sampled from topics such as 'sports', 'movies', 'holidays', etc. Similarly, Miani et al. (2021) created a large corpus of CT texts, LOCO. It provides CTs and mainstream texts on the same topics. To gather CT texts, they used a list of CT websites based on scores from mediabiasfactcheck[1]. To retrieve mainstream documents, the authors used Google to search for the seeds. Work by Mompelat et al. (2022), who reannotated parts of the LOCO corpus, shows that the original distinction of conspiracy and mainstream texts is reasonably reliable, but they also found that in some cases, it was difficult to decide whether a CT was perpetuated in a text. Note that any bias introduced by annotation or corpus creation will automatically be perpetuated into any machine learning model trained on such a corpus.

In the work presented here, we delve deeper into the distinction of CT and mainstream texts. In the first experiment, we asked students to annotate texts to determine how CTs were propagated. They were first trained on a set of texts, after having been provided with annotation guidelines. The results showed low inter-annotator agreement (IAA). In a follow-up study, we decided to redo the experiment with CT experts. These results were slightly improved but still did not meet the threshold of acceptable IAA, which raises the question of whether we need better annotation guidelines or whether the CT phenomenon is highly subjective. After a thorough evaluation of the two separate sets of annotations (students and experts), we come to the conclusion that the latter is the case. This suggests that we cannot expect a single gold standard annotation, and consequently cannot use IAA as a measure of the annotation quality. We finally provide a discussion of the consequences of our findings for annotation projects concerned with highly subjective phenomena.

Our work is closely aligned with other work on annotator disagreement and perspectivist approaches to NLP. Thus, our insights are not novel in NLP; we are adding to the discussion by

[1] https://mediabiasfactcheck.com/conspiracy/

Proceedings of The 18th Linguistic Annotation Workshop (LAW-XVIII), pages 124–132

adding conspiracy annotation/detection to the list of subjective tasks, which require perspectivist approaches.

2 Disagreement in Annotation

Early work on disagreement in linguistic annotations (Passonneau, 2004; Poesio and Artstein, 2005) introduced Krippendorff's alpha (Krippendorff, 2019) as a metric to measure inter-annotator agreement and introduced the notion of explicit and implicit ambiguity in annotations, the latter referring to ambiguity revealed through annotator disagreement. More recent work has started looking into disagreement in annotations beyond measuring it, instead accepting it as a necessary phenomenon in the annotation of subjective tasks for a range of tasks: for example, POS tagging (Plank et al., 2014), textual inference (Pavlick and Kwiatkowski, 2019), sexism (Almanea and Poesio, 2022), toxicity (Sap et al., 2022), and hate speech detection (Akhtar et al., 2020; Mostafazadeh Davani et al., 2022).

3 Annotating Conspiracy Theories

A CT refers to a story or narrative that claims a small group or 'deep state' has control over the government and is involved in harmful activities aimed at causing widespread harm (e.g. Enders et al., 2021). This includes doubting scientific evidence of climate change and vaccinations, believing that elections are rigged, and fearing that immigrants, Black individuals, or Jewish people pose a threat to the rights, freedom, and culture of white people (Enders et al., 2021). Empirical evidence-based research from a variety of disciplines seeks to explain why people believe in certain CTs (e.g. Daniel and Harper, 2022; Lewandowsky et al., 2013; Uscinski et al., 2020, 2022).

Neville-Shepard (2018) and Serazio (2016) assert that creators of CT text use language familiar to audiences attracted to conspiracism, and conspiratorial text is grounded in a small amount of 'evidence' that encourages a 'leap of faith' by the audience to reach conclusions. This suggests that conspiratorial discourse caters to people in the know and, therefore, does not explicitly convey a premise, rather, the audience completes the argument based on prior knowledge. Reyes and Smith (2014) further assert that creators of CT text depend on audiences already familiar with similar ideas. This means that these theories tap into a broader culture of belief in conspiracies and act as a way for people to find others who share their beliefs and reinforce their convictions. However, this makes it difficult to classify conspiratorial text because, without an explicit claim, it is up to the audience to 'leap' to conclusions based on the familiar tropes presented in the text.

Mompelat et al. (2022) have reannotated parts of the LOCO corpus (Miani et al., 2021) to determine how reliable the automatic corpus collection was. They started with a simple definition of what they considered a CT text, namely a text that propagated a *conspiracy belief*, defined as: "A conspiracy belief is the belief that an organization made up of individuals or groups was or is acting covertly to achieve some malevolent end. It depicts causal narratives of an event as a covert plan orchestrated by a secret cabal of people or organizations instead of a random or natural happening" (Seelig et al., 2022). Their first round of annotations showed high IAA for mainstream texts, but the IAA for CT was 0.47. As a consequence, they adjusted the guidelines and added that in order for a text to be considered a CT text, the following had to hold: "A document is considered CT if and only if such a belief is manifested in the text via specific expressions" (Mompelat et al., 2022). They also identified a set of textual and verbal cues that triggered a reading of conspiracy, e.g., all caps texts, paraphrases, questions. The revised guidelines resulted in a higher IAA for conspiracy texts of 0.70. However, when they used the same annotation scheme for a different conspiracy theory, the results for CT texts was considerably lower (0.58), thus showing that robust annotations are difficult, even with trained annotators.

4 The Annotation Study

We conducted an annotation study to determine which circumstances (in terms of training) we need to reliably annotate a range of phenomena. More specifically, the annotations covered identifying similarities in main themes, structures, rhetorical forms, and tropes. For the annotation samples, we revisited the LOCO corpus to draw a sample of CTs. We only extracted documents identified as a conspiracy, representing a broad range of topics, using the seeds Covid-19, Pizzagate, Climate change, JFK assassination, 9/11, Illuminati, and Flat Earth. Our non-experts are graduate students without prior knowledge of the litera-

ture on CTs. The experts are researchers who have worked on CTs for at least 2 years. We report both Fleiss' kappa (Fleiss, 1971) and Krippendorff's alpha (Krippendorff, 2019).

The annotation scheme is based on past research on hate speech, CT text, and populist rhetoric (Bastos and Farkas, 2019; Rieger et al., 2021; Seelig et al., 2022). We used the following questions:

1. Presence of a CT in the text (e.g., causal narratives of an event as a covert plan orchestrated by a secret cabal of people/organizations).
2. Main or dominant CT narrative (e.g., Flat Earth, Moon Landing, White Genocide, etc.).
3. Treatment of the CT: document supports, endorses, and/or reinforces a CT; or refutes or debunks it.
4. Leap of faith: narrative takes accepted facts and makes a leap of faith to reach conclusions that are not supported by the facts.
5. Type of argument: syllogism (a deductive scheme of a formal argument consisting of a major and a minor premise and a conclusion) or enthymeme (an argument in which one premise is not explicitly stated).
6. Sentiment of narrative (e.g., positive, neutral, or negative).
7. Pathos: appealing to audience's emotions (e.g., humor and sarcasm; inspiration and hope; sadness; sympathy and pity; courage and strength; hatred; love; fear; anger).
8. Logos: rational basis for an argument/reason (e.g., statistics; recorded evidence; historical data or facts; studies, surveys, or academic papers; personal experience/testimony; hearsay; or not applicable).
9. Ethos: the credibility of the speaker or poster (e.g., celebrity; authority figure; credible or public figure; animals; inanimate objects; a person in the street excluding celebrity/ authority/credible figures).
10. Fearmongering (e.g., mentioning fatalities caused by natural disasters, crime, acts of terrorism, civil unrest, or accidents).
11. Emotional spectrum: use of emotional words (e.g., afraid, excited, sweet, and jealous), exclamation marks (e.g., they are crying!), or emojis.
12. Real-world issues (e.g., politics, economy, military conflict, crime, local affairs, weather, public health, education, protest, ethnocultural minorities, or terrorism).

	Non-experts		Experts	
	κ	α	κ	α
CT present	0.10	0.37	-0.12	-0.71
Main CT present	0.22	0.58	0.07	0.25
Treatment of CT	0.13	0.46	0.01	-0.01
Leap of faith	0.09	0.11	0.10	0.02
Type of arg.	0.03	0.34	-0.09	0.28
Sentiment	0.02	0.30	-0.05	0.05
Pathos	0.07	0.33	0.06	0.51
Logos	0.03	0.23	0.01	0.29
Ethos	0.04	0.28	0.02	0.41
Fearmongering	0.15	0.48	-0.07	-0.28
Emotional spect.	0.06	0.34	-0.13	-0.38
Real-world	0.13	0.43	0.03	-0.10

Table 1: Inter-annotator agreement for CTs comparing non-experts and expert.

We first trained the non-expert annotators on a sample of 11 conspiracy documents selected from the seven CTs (Covid-19, Pizzagate, Climate change, JFK assassination, 9/11, Illuminati, and Flat Earth). After training, they independently annotated a random sample of 472 CT texts representing the same conspiracy topics. The annotations were conducted by two MA students and three PhD students without prior knowledge of research on CTs. The results are shown in Table 1. The analysis yielded poor IAA, even for the most basic question of whether a CT was present ($\kappa = 0.10$).

Given the low IAA of the non-experts, we conducted a similar, but smaller experiment with a group of 4 experts as annotators. We used the same annotation scheme on a sub-sample of 25 CTs from the same sample the students annotated. The results of this experiment are shown in the second column in Table 1. We notice that there are several negative values. Similar to the findings of Mompelat et al. (2022), the reason for this can be found in the very high expected values. Neither metric is useful for data with very high agreement and small sample size (Zhao et al., 2013). If we interpret these values as reasonably high agreement, we see that the experts tend to agree on whether a CT is present, how the CT is treated, and on fearmongering, the emotional spectrum, and real-

	κ	α
CT present	0.403	0.827
Main CT present	0.298	0.217
Treatment of CT	0.383	0.795

Table 2: Inter-annotator agreement for CTs for experts using the simplified annotation scheme.

world issues present. The remaining numbers are lower in comparison to the non-experts, including for the questions which CT was present and its treatment, thus showing that prior research experience on CTs is not helpful.

As part of this second round of annotations, we added questions to capture experts' certainty of their annotation for items flagged controversial (e.g., very certain, pretty sure, not sure, I have no idea/guessed). The answers show that while the experts never guessed, the majority were only pretty sure (61-84%) or not sure (3-18%).

Due to the lack of agreement for non-experts and experts, we decided to simplify the coding scheme to the first three questions, but to clarify and extend the guidelines, to see if more explicit instructions would increase IAA. The modifications were based on a discussion of the experts of which uncertainties they faced during the annotation process. We used another sub-sample of 10 CTs from the same sample the students annotated. The results are based on 6 expert annotators and are shown in Table 2. We see that these numbers are higher than for the previous experiment, but the agreement is still far from what is generally considered reliable: when asked whether the text contains a CT, we obtained $\kappa = 0.403$; for "main conspiracy present?" $\kappa = 0.298$; and for "treatment of conspiracy" $\kappa = 0.383$.

During the last experiment, we also asked the larger group of experts to describe any difficulty they had determining the answers. We show sample responses in Table 3. These responses show that even experts struggle with basic questions such as whether a CT is present, which we interpret as an indication that there do not exist clear boundaries.

Q1	Text read as if an excerpt from a news story [Text 1]
Q1	The overall passage read as if not true, but hard to discern a specific CT
Q1	Unsure if this is a CT or simple a dispute [Text 2]
Q2	sounds like a movie plot [Text 2]
Q2	This one was difficult because it's describing a real thing that happened in language that's a bit bombastic, and also acknowledges an offshoot that may or may not actually exist. [Text 3]
Q2	Illuminati is mentioned, but the main text assumes the reader knows the Illuminati are perpetrating mind control and other atrocities.
Q3	needs fact checking
Q3	narrative was about combating the spread of COVID-19 [Text 1]
Q3	Unsure about this one - it mentions misinformation but is it a CT?
Q3	It engages with CTs and seems to endorse them, but it is more about getting you to pay attention and stay.

Table 3: Sample responses describing difficulties in answering questions. Text numbers refer to Appendix A.

5 Consequences for Annotation Projects

Our results above show clearly that it is extremely difficult to reach high IAA on even basic questions such as whether a text contains a CT, even when experts are used as annotators. It is possible that agreement metrics can be increased by further extending the instructions for annotations and by training annotators to respond in a specific way to specific texts. However, such a setup may encourage annotators to annotate what experimenters want to hear instead of annotating what they see as being present in the text. A closer look at the texts and the annotations shows that these decisions depend on prior knowledge and on how the text is interpreted. If we streamline the annotations too rigidly, then we create the possibility that annotators try to guess what the experimenters want to see as answers, thus clouding legitimate interpretations of the text. For example, the decision whether Text 3 in Appendix A propagates a CT will depend not only on how much the annotator knows about the case, but also on how much they trust the source of this text.

A closer look at the texts and the annotations allows us to conclude that the annotations are and need to be subjective. We cannot have a single gold standard annotation; rather, we must be prepared to accept a range of answers. This conclu-

sion leads to several consequences, partly for the definition of the problem, and partly for modeling the problem computationally.

1. Disagreement between annotators is not a sign of lack of clarity in the annotation scheme but a direct consequence of the phenomenon to be annotated. One reason is that belief in CTs is not monolithic, but rather faceted, where individuals belief in subsets of factoid of a range of CTs. Another reason is that we need to model the preception of standard readers, and their interpretation will depend on prior knowledge as well as prior bias.
2. We cannot expect a single "correct" answer; rather, we need to accept ranges of answers. This is in line with other subjective tasks such as hate speech annotation or sentiment annotation.
3. Metrics such as Fleiss' kappa and Krippendorff's alpha cannot be used to evaluate the quality of annotations. More specifically, such tasks cannot be evaluated based on consistency.
4. Machine learning (ML) approaches to model the phenomenon should not define it as a classification task but instead need to predict the range and distribution of answers.
5. ML models based on gold standard annotations may be severely biased.
6. The lack of agreement requires a shift in machine learning paradigm, taking learning from disagreement (Mostafazadeh Davani et al., 2022; Uma et al., 2021) more seriously since the variability in annotations can significantly affect the task (classification vs. predicting a distribution). Thus, it needs to be integrated more closely in the training regime.

6 Limitations

Our comparative annotation study is based on a small number of annotators since it was supposed to serve as a pilot stud for a larger annotation project. However, the students went through a thorough training session, and the number of experts is naturally limited by availability. All expert annotators are also co-authors on this report, our expertise ranges over a wide array of fields, which ensures a wide disciplinary stance.

7 Ethical Considerations

Working with CTs tends to be difficult for the annotators. For this reason, we concentrated on a set of different CTs that are less prone to explicit hatred that is often present in CTs targeting specific minority groups (e.g., white genocide). However, despite this careful selection, the chosen texts can contain content that may upset annotators.

Acknowledgments

This work is based on research supported by US National Science Foundation (NSF) Grants #2123635 and #2123618.

We are grateful to all our annotators and to the reviewers for their comments, which helped strengthen the paper.

References

Sohail Akhtar, Valerio Basile, and Viviana Patti. 2020. Modeling annotator perspective and polarized opinions to improve hate speech detection. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, pages 151–154.

Dina Almanea and Massimo Poesio. 2022. ArMIS - the Arabic Misogyny and Sexism Corpus with annotator subjective disagreements. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 2282–2291, Marseille, France.

Marco Bastos and Johan Farkas. 2019. "Donald Trump is my President!": The internet research agency propaganda machine. *Social Media + Society*, 5(3).

Lauren Daniel and David J. Harper. 2022. The social construction of conspiracy beliefs: A Q-methodology study of how ordinary people define them and judge their plausibility. *Journal of Constructivist Psychology*, 35(2):564–585.

Adam Enders, Joseph Uscinski, Casey Klofstad, Michelle Seelig, Stefan Wuchty, Manohar Murthi, Kamal Premaratne, and John Funchion. 2021. Do conspiracy beliefs form a belief system? Examining the structure and organization of conspiracy beliefs. *Journal of Social and Political Psychology*, 9(1):255–271.

Joseph L. Fleiss. 1971. *Statistical Methods for Rates and Proportions*. John Wiley.

Klaus Krippendorff. 2019. *Content Analysis: An Introduction to Its Methodology*. SAGE.

Sandra Kübler and Heike Zinsmeister. 2014. *Corpus Linguistics and Linguistically Annotated Corpora*. Bloomsbury.

Stephan Lewandowsky, Klaus Oberauer, and Gilles E. Gignac. 2013. NASA faked the moon landing – therefore, (climate) science is a hoax: An anatomy of the motivated rejection of science. *Psychological Science*, 24(5):622—633.

William Marcellino, Todd C. Helmus, Joshua Kerrigan, Hilary Reininger, Rouslan I. Karimov, and Rebecca Ann Lawrence. 2021. Detecting conspiracy theories on social media. Technical Report RR-A676-1, The RAND Corporation.

Alessandro Miani, Thomas Hills, and Adrian Bangerter. 2021. LOCO: The 88-million-word language of conspiracy corpus. *Behavior Research Methods*.

Ludovic Mompelat, Zuoyu Tian, Matthew Luettgen, Amanda Kessler, Aaryana Rajanala, Sandra Kübler, and Michelle Seelig. 2022. How "loco" is the LOCO corpus? Annotating the language of conspiracy theories. In *Proceedings of the Sixteenth Linguistic Annotation Workshop*, Marseille, France.

Aida Mostafazadeh Davani, Mark Díaz, and Vinodkumar Prabhakaran. 2022. Dealing with disagreements: Looking beyond the majority vote in subjective annotations. *Transactions of the Association for Computational Linguistics*, 10:92–110.

Ryan Neville-Shepard. 2018. Paranoid style and subtextual form in modern conspiracy rhetoric. *Southern Communication Journal*, 83(2):119–132.

Rebecca Passonneau. 2004. Computing reliability for coreference annotation. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC)*, Lisbon, Portugal.

Ellie Pavlick and Tom Kwiatkowski. 2019. Inherent disagreements in human textual inferences. *Transactions of the Association for Computational Linguistics*, 7:677–694.

Barbara Plank, Dirk Hovy, and Anders Søgaard. 2014. Linguistically debatable or just plain wrong? In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pages 507–511, Baltimore, MD.

Massimo Poesio and Ron Artstein. 2005. The reliability of anaphoric annotation, reconsidered: Taking ambiguity into account. In *Proceedings of the Workshop on Frontiers in Corpus Annotations II: Pie in the Sky*, pages 76–83, Ann Arbor, MI.

Ian Reyes and Jason K. Smith. 2014. What they don't want you to know about planet X: Surviving 2012 and the aesthetics of conspiracy rhetoric. *Communication Quarterly*, 62(4):399–415.

Diana Rieger, Aanna Sophie Kümpel, Maximilian Wich, Toni Kiening, and Groh Groh. 2021. Assessing the extent and types of hate speech in fringe communities: A case study of alt-right communities on 8chan, 4chan, and Reddit. *Social Media + Society*, 7(4).

Maarten Sap, Swabha Swayamdipta, Laura Vianna, Xuhui Zhou, Yejin Choi, and Noah A. Smith. 2022. Annotators with attitudes: How annotator beliefs and identities bias toxic language detection. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5884–5906, Seattle, WA.

Michelle Seelig, John Funchion, Ruth Trego, Sandra Kübler, Daniel Verdear, Stefan Wuchty, Joseph Uscinski, Casey Klofstad, Manohar Murthi, Kamal Premaratne, and Amada Diekman. 2022. The dark side of Twitter: A framing analysis of conspiratorial rhetoric stoking fear. Presentation at the 72nd Annual ICA Conference: One World, One Network? Online.

Michael Serazio. 2016. Encoding the paranoid style in American politics: "Anti-establishment" discourse and power in contemporary spin. *Critical Studies in Media Communication*, 33(2):181–194.

Alexandra Uma, Tommaso Fornaciari, Dirk Hovy, Silviu Paun, Barbara Plank, and Massimo Poesio. 2021. Learning from disagreement: A survey. *Journal of Artificial Intellingence Research*, 72.

Joseph Uscinski, Adam Enders, Amanda Diekman, John Funchion, Casey Klofstad, Sandra Kübler, Manohar Murthi, Kamal Premaratne, Michelle Seelig, Daniel Verdear, and Stefan Wuchty. 2022. The psychological and political correlates of conspiracy theory beliefs. *Scientific Reports*, 12(21672).

Joseph E Uscinski, Adam M Enders, Casey Klofstad, Michelle Seelig, John Funchion, Caleb Everett, Stefan Wuchty, Kamal Premaratne, and Manohar Murthi. 2020. Why do people believe covid-19 conspiracy theories? *Harvard Kennedy School Misinformation Review*, 1(3).

Xinshu Zhao, Jun S Liu, and Ke Deng. 2013. Assumptions behind intercoder reliability indices. *Annals of the International Communication Association*, 36(1):419–480.

A Appendix

Text 1:

Mainland China reported 30 new coronavirus cases on Saturday, up from 19 a day earlier as the number of cases involving travellers from abroad as well as local transmissions increased, highlighting the difficulty in stamping out the outbreak.

The National Health Commission said in a statement on Sunday that 25 of the latest cases involved people who had entered from abroad, compared with 18 such cases a day earlier.

Five new locally transmitted infections were also reported on Saturday, all in the southern coastal province of Guangdong, up from a day earlier.

The mainland has now reported a total of 81,669 cases, while the death toll has risen by three to 3,329.

Though daily infections have fallen dramatically from the height of the epidemic in February, when hundreds of new cases were reported daily, Beijing remains unable to completely halt new infections despite imposing some of the most drastic measures to curb the virus spread.

The so-called imported cases and asymptomatic patients, who have the virus and can give it to others but show no symptoms, have become among Chinas chief concerns in recent weeks. The country has closed off its borders to almost all foreigners as the virus spread globally, though most of the imported cases involve Chinese nationals returning from overseas.

The platinum standard of advanced multivitamin formulations is back in stock! Order Vitamin Mineral Fusion at 50% off with double Patriot Points and free shipping today!

Text 2:

Between 542 and 66 million years ago — long before the "supervolcano" became part of Yellowstone's geologic story — the area was covered by inland seas.

NPS/Jim Peaco

Most of Earth's history (from the formation of the earth 4.6 billion years ago to approximately 541 million years ago) is known as the Precambrian time. Rocks of this age are found in northern Yellowstone and in the hearts of the Teton, Beartooth, Wind River, and Gros Ventre ranges. During the Precambrian and the subsequent Paleozoic and Mesozoic eras (541 to 66 million years ago), the western United States was covered at times by oceans, sand dunes, tidal flats, and vast plains. From the end of the Mesozoic through the early Cenozoic, mountain-building processes formed the Rocky Mountains.

During the Cenozoic era (approximately the last 66 million years of Earths history), widespread mountain-building, volcanism, faulting, and glaciation sculpted the Yellowstone area.

Magma (molten rock from Earth's mantle) has been close to the surface in Yellowstone for more than 2 million years. Its heat melted rocks in the crust, creating a magma chamber of partially molten, partially solid rock. Heat from this shallow magma caused an area of the upper crust to expand and rise. The Yellowstone Plateau became a geomorphic landform shaped by episodes of volcanic activity. Stress also caused rocks overlying the magma to break, forming faults and causing earthquakes. Eventually, these faults reached the deep magma chamber. Magma oozed through these cracks, releasing pressure within the chamber and allowing trapped gases to expand rapidly. A massive volcanic eruption then occurred along vents, spewing volcanic ash and gas into the atmosphere and causing fast super-hot debris (pyroclastic) flows on the ground. As the underground magma chamber emptied, the ground above it sunk, creating the first of Yellowstones three calderas.

This diagram shows the general ideas behind two theories of how magma rises to the surface. Adapted with permission from Windows into the Earth by Robert Smith and Lee J. Siegel, 2000.

Researchers found that the changes leading up to an eruption may happen in a matter of decades rather than thousands of years in advance as previously thought.

Based on minerals from the last major eruption, the Supervolcanoes are characterized as volcanic centers that have had eruptions that covered more than 240 cubic miles. The US has two: one in Yellowstone and another in Californias Long Valley. An eruption could emit ash that would expand over 500 miles. The eruption would likely cover the ground with as much as 4 inches of gray ash, which could be detrimental to crops growing in the Midwest. Another less worrisome concern is the 1,000 degree F molten lava that could ooze out. Gases, including sulfur dioxide, which contributes to acid rain would be spewed from the supervolcano and the global cooling issues associated with reflecting sunlight away from the Earth are also concerns.

But there are other supervolcanos in the world with sooner predictions than Yellowstones. Campi Flegri, a name that aptly translates as burning fields, is in a critical state, according to researchers in Italy. It consists of a vast and complex network of underground chambers that formed hundreds of thousands of years ago, stretching from the outskirts of Naples to underneath the Mediterranean Sea. Though its last eruption was in 1538, its due for an eruption soon. It would be a minor event compared to the 72 cubic miles of molten rock it spewed in its most notorious eruption 39,000 years ago, called Campanian Ignimbrite, that likely contributed to the extinction of the Neanderthals.

Fortune website article reported that if the Yellowstone supervolcano erupts, it could shoot out more than 1,000 cubic kilometers of rock and ash into the air. Thats 250 cubic miles. Thats more than three times as large as the Campanian Ignimbrite eruption in Italy, which created a sulfurous cloud that floated more than 1,200 miles away to hang over Russia. Thats 2,500 times more material than Mount St. Helens expelled in 1980, killing 57 people. An eruption at Yellowstone would result in a cloud of ash more than 500 miles wide, stretching across nearly the entire western United States.

NASA has a plan to neutralize supervolcano threats however. They believe the most viable solution could be to drill up to 6 miles down into the supervolcano, and pump down water at high pressure. The circulating water would return at a temperature of around 662F, thus slowly day by day extracting heat from the volcano. And while such a project would come at an estimated cost of around $3.46 billion, it comes with an enticing catch which could convince politicians to make the investment. It would become a source of geothermal energy. But there are considerable risks, too. It could trigger the eruption its meant to save us from.

Historically, four types of volcanic events have taken place in Yellowstone (you may click on each one to learn more):

1. Caldera Forming Eruptions – 2.1 and 1.3 million years ago

2. Lava Flows – about 30 between 640,000 and 70,000 years ago

3. Earthquakes – 1000 to 3000 yearly; last notable quake was in 1959

4. Hydrothermal (Steam) Explosions – small explosions in the 20th century; a dozen or so major explosions between 14,000 and 3,000 years ago

The likelihood of an eruption in the near future is still low. However those who instigate such a project will never see it to completion, or even have an idea whether it might be successful within their lifetime. Cooling Yellowstone in this manner would happen at a rate of 3.2808399 feet a year, taking of the order of tens of thousands of years until just cold rock was left.

Featured Image: Yellowstone harbours a giant magma chamber that will blow one day if we dont act (Credit: iStock)

Text 3:

The victim was kept in a chemically induced sleep for weeks and subjected to rounds of electroshocks, experimental drugs and tape-recorded messages played non-stop.

CBC News recently reported that the Canadian government reached an out-of-court settlement of $100,000 with Allison Steel, the daughter of Jean Steel, a woman who was subjected to horrific brainwashing experiments funded by the CIA.

The settlement was quietly reached in exchange for dropping the legal action launched by Allison Steel in September 2015. The settlement includes a non-disclosure agreement prohibiting Steel from talking about the settlement itself. However, the existence of the settlement and its total amount appeared in public accounts released by the federal government in October.

Jean Steels ordeal began in 1957, at the age of 33. She was admitted at the Allan Memorial Institute in Montreal after being diagnosed with manic depression and delusional thinking.

In the following months, Steel became a victim of CIA-funded MKULTRA experiments conducted by Dr. Ewen Cameron.

Camerons experiments aimed to de-pattern the victims mind through intense trauma in order to re-pattern it afterward. In other words, he was researching the basis of Monarch Programming the mind control program that is often discussed on Vigilant Citizen.

Cameron believed a combination of chemically induced sleep for weeks at a time, massive electroshock treatments, experimental hallucinogenic drugs like LSD and techniques such as psychic driving through the repeated playing of taped messages could de-pattern the mind, breaking up the brain pathways and wiping out symptoms of mental illnesses such as schizophrenia. Doctors could then re-pattern patients. However, the de-patterning also wiped out much the patients memory and left them in a childlike state. In some cases, grown adults forgot basic skills such as how to use the bathroom, how to dress themselves or how to tie their shoes.

CBC News, Federal government quietly compensates daughter of brainwashing experiments victim

Hundreds of pages detail the horrific experiments Jean Steel was subjected to.

According to a report written by Cameron, Steel was kept in a chemically induced sleep for weeks. One series lasted 29 days. A second lasted 18 days. The sleep therapy was accompanied by a series of electroshocks. She was extremely confused and disoriented but much more co-operative, Cameron wrote in his report. Nurses notes on her charts detail repeated doses of sodium amytal, and how Steel would pace the hall and rail about feeling like a prisoner: Its just like being buried alive. Somebody please do something. This was all said screaming at the nurse and doctor, one note said.

Steel then began to exhibit bizarre behavior. Her daughter recounts:

When you wanted to talk with her about something emotional she just could not do it, Steel said. Her emotions were stripped. It took away her soul. Her mother would sit alone in the dark, writing codes and numbers on the walls. One time I came home and the ceiling was spray-painted with red swirls all over it, Steel said. She would take wallpaper and cut out little sections of it and she would pin it to the whole room.

While MKULTRA is viewed by mass media as a shameful episode of the past, it is also part of our present. The program still exists in a much more refined version under the name of Monarch programming.

Heres an interesting 1980 documentary about MKULTRA experiments in Canada produced by the CBC: