The Tentativeness and Trustworthiness of Science

by

Jocelyn Robbins

A thesis submitted to the

Faculty of the School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Bachelors of Arts in Physics

Department of Physics

April 12, 2023

Committee Members:

Thesis Advisor:	Dr. Bethany Wilcox, Physics
Honors Council Representative:	Dr. Paul Beale, Physics
External Faculty:	Dr. John Keller, Astrophysical and Planetary Sciences

Robbins, Jocelyn (B.A. Physics)

The Tentativeness and Trustworthiness of Science

Thesis directed by Dr. Bethany Wilcox

The goal of this project is to explore how educational experiences influence the understanding of the tentativeness and trustworthiness of science. Previous researchers created instruments to measure how an individual's educational experience impacts their understanding of the tentativeness of science and identified different factors that affect trust in science. However, none of these studies have looked into the relationships between these variables. To address this relationship, we created a survey consisting of a combination of open-ended questions and likert-style items. After distributing the survey and collecting the data, we used statistical hypothesis testing to determine what, if any, correlations existed between the items and different demographics groups. Contrary to what we initially expected, the results found that formal education in science was not a significant predictor of an individual's ideas about the tentativeness and trustworthiness of science, rather, political party affiliation was the strongest predictor of an individual's response. Further research in this area could explore the relationships of intersectional identities and look at effects of different types of education.

Dedication

To my friends and mentors who supported me throughout this undertaking and my cats, Marcy and Gumbo.

Acknowledgements

This project would not have been possible without Dr. Bethany Wilcox who let me explore my own research project and helped in every way possible. Special thanks to Wilcox's Monday Meeting group, the Physics Education Group, and Emily Allen Walsh for proof reading and revising. I am also grateful for the Physics Education Group for showing the world of education research to me. Final thanks to Mark Suter for acting as another reviewer and giving me unconditional support.

Contents

Tables

Table

Figures

Figure

Chapter 1

Introduction

Change in science reflects a natural progression as new ideas and theories allow us to gain a deeper understanding of our world. Recent events, including the COVID-19 pandemic, emphasized the importance of understanding how people reconcile the tentativeness and trustworthiness of science. For the purpose of this study, we used the phrase ‘tentativeness of science’ to encompass the changing nature of science and the willingness of scientists to modify scientific ideas and results. The phrase ‘trustworthiness of science’ refers to one’s ability to trust scientific results and make informed decisions derived from those results. This study explored one potential consequence of misunderstanding the tentativeness of science: a lack of trust in science. We examined the relationship between science and trust from the perspective of educational experience to determine any potential points of intersection between the three factors. How is trust in science impacted by its iterative nature and does one’s educational experiences play a role in this?

That question guided us through creating a survey and distributing it across the University of Colorado Boulder’s campus. We made the survey from Likert-style statements, referred to as items, and open-ended questions that were original or sampled from validated studies. We collected data on different social identifiers and, even though our primary focus was the impact of educational experience, we acknowledged the potential power other lived experiences possess over the reconciliation of the tentativeness and trustworthiness of science. Once we distributed the survey and accumulated enough data, we used descriptive statistics and statistical hypothesis testing to discover any potential correlations between social identifiers and individuals’ answers to items. We then explored the trends uncovered from the claims and made interpretations using available data from both our survey and other studies. We summarized our work and considered possible future work in the conclusion. In the rest of this section, we cover the motivation behind our work and review any prior literature in this field of study.

1.1 Motivation

COVID propelled the tentativeness of science into the spotlight as scientists and media managed emerging new evidence and resource management. Before, the general population may not have had sufficient reason to investigate the inner workings of experiments, but ongoing scientific-based changes to COVID policy increased awareness. Instead of accepting new information and policies based on science, we saw denial and mistrust. We recognized the multiple influencing factors in this period but the online reactions made us wonder about the role the tentativeness of science played in their distrust. Even though our curiosity about COVID initiated this study, we chose not to focus the survey on its impacts, as we wanted to expand the scope to trustworthiness of science in general.

We focused on the tentative nature of science in general, as science is ever changing with new discoveries or revised data and the tendency to change is a facet of the nature of science (NOS). While popular scientific curricula like the Next Generation Science Standards provide information about NOS in an appendix, they do not feature the information in the main curriculum, so it is unclear how teachers might cover these concepts. This uncertainty inspired curiosity about the impact educational experiences have on an individual's understanding surrounding the tentativeness of science. Does this potential lack of education impact an individuals' trust if schools do not cover the tentativeness of science? This question cemented our main goal of exploring the intersection between the tentativeness and trustworthiness of science and education.

While reviewing the literature surrounding trust in science, we noticed a lack of research about what makes people distrust science as several studies addressed only how to build trust in science. The studies emphasized what is done in the sphere of science and experiments , such as reproducibility tests and fostered integrity. There are also very few studies that explicitly explore the influence of education on trust in science.

1.2 Literature Review

Related to the Tentativeness of Science

The idea of teaching more than hard science in a science class has existed before the turn of the 20th century, and the Next Generation Science Standards (NGSS) contributed the most recent curricula to this argument. They focused on cross cutting concepts, scientific and engineering practices, and disciplinary core ideas. After the framework was released in July 2011, the committee published regular updates. One of the updates was Appendix H - Understanding the Scientific Enterprise: The Nature of Science. The goal of Appendix H was to introduce a scientifically literate person to the nature of science [7]. Appendix H breaks the multifaceted concept of NOS into eight segments that are interspersed throughout NGSS concepts and practices.

The third segment stated that “scientific knowledge is open to revision in light of new evidence”. This understanding was incorporated into the scientific and engineering practices and was rephrased for each level of school ranging from K-2 to highschool. The learning outcomes are as follows:

- K-2: “Science knowledge can change when new information is found”
- 3-5: “Science explanations can change based on new evidence.”
- Middle school: “Scientific explanations are subject to revision and improvement in light of new evidence.”
- High school: “Most scientific knowledge is quite durable but is, in principle, subject to change based on new evidence and/or reinterpretation of existing evidence.” [7]

A study in 2016 analyzed and critiqued the Appendix H and rated each of the eight segments and the respective provided examples as either acceptable or unacceptable. To gain an acceptable rating, the exemplar had to be both accurate and related to the category. A panel of five expert science educators with a background in NOS determined the accuracy of the examples along with two other questions; how frequently do the NOS categories appear in the overarching curricula, and how complete is the NGSS version of NOS compared to a broader list created by a NOS focused meta-analysis. The tentativeness of science was rated acceptable across the three categories but there are critiques concerned with how this topic is taught in schools. The main concern cited was the lack of assistance when it came to teachers learning about NOS and implementing NOS in their classroom [20].

An instrument, the Student Understanding of Science and Scientific Inquiry (SUSSI), tested certain aspects of the NOS and went through many revisions and validation tests. The instrument measured students' development of scientific knowledge and their beliefs inherent to NOS by including both qualitative and quantitative questions. Researchers have used SUSSI in many studies such as testing the effectiveness of historical storytelling or assessing preservice elementary teachers' views about NOS. The entire instrument focuses on seven essential elements in K-12 science education which includes the tentativeness of scientific knowledge. Their definition of the tentativeness of science can be found below.

> "Scientific knowledge is both tentative and durable. Having confidence in scientific knowledge is reasonable while realizing that such knowledge may be abandoned or modified in light of new evidence or reconceptualization of prior evidence and knowledge. The history of science reveals both evolutionary and revolutionary changes." [17]

Related to the Trustworthiness of Science

The National Science Board under the National Science Foundation collected data from multiple sources about the attitudes and interests in science over the past four decades. The May 2022 report included a measured high confidence in scientists from a majority of American adults who also reported they recently sought information about medicine or disease [3]. The report does not state how Americans looked into the information and what sources used and trusted. One of National Science Board's sources is the Wellcome Global Monitor (WGM) which measured global attitudes towards science in 2018.

The WGM report measured trust from 140 different areas and countries and, while we were concerned about the trust of science in America, there were overarching trends we considered important. The WGM identified the two main factors that influenced personal levels of trust in science as "learning science at school or college and confidence in key national institutions" [8]. The statistical analysis only explained 15% of the variations in trust but other potential factors were location, access to information, and economic status.

America, with about 1000 participants, ranked as one of the top countries who trusted science. 25% of respondents reported a high level of trust in scientists and 13% of respondents felt less trusting [8]. The only other options were "medium trust" or "no opinion". The questions WGM asked included measuring trust that scientists reported accurate information, scientists worked for the benefit of the public, and scientists were honest about their funding. This report was created in 2018 so COVID had no influence on the data. However, a repeat study conducted by WGM in 2020 found that, on the global scale, trust levels increased from 34% in 2018 to 43% by the end of 2020 [9].

As a global study, the WGM's identified drivers of influenced trust in science were broad. While the study asked about confidence about national institutions, including confidence in the national government, it did not ask about political parties. In America, ideological thinking along party lines has become prevalent over the past few decades. According to the Pew Research Center, "the overall share of Americans who express consistently conservative or consistently liberal opinions has doubled over the past two decades from 10% to 20%" [12]. Political divisions influence many aspects of life but a recent study conducted at University of Chicago in 2021 measured a 30-point gap between Democrats and Republicans when asked about their confidence in the scientific community, which is up from a 9-point gap in 2018 [14].

1.3 Leading Questions

With a completed literature review and stated motivation, we crafted research questions that lead us through the rest of the project.

- Does education influence how the tentative nature of science is understood?
- What causes people to distrust science?
- How does personal educational experience impact the ability to reconcile the tentativeness and trustworthiness of science?

Chapter 2

Methods

In this chapter, we presented the methods used in our study of the tentativeness and trustworthiness of science. This included the creation of the survey, the context surrounding distribution, and the ways in which we analyzed the data. The survey creation process included using questions and items from previously validated instruments and creating our own. The context section explored the distribution methods we used and a breakdown of the final data set population. We also described the rubric creation process for the qualitative section and the statistical strategies we employed for the quantitative data.

2.1 Survey Creation

This section explores the creation of our survey. We designed the survey using existing items from validated measurement tools and original items and questions that engaged with our specific research questions. Appendix D has the full survey.

Format

The survey consisted of four open-ended questions and, because of two conditional items that relied on participants' prior responses, ten to eleven Likert-scale items. These represent the two popular question formats in a survey: open and closed-ended questions, each with their own affordances and constraints. Open-ended questions allowed participants to generate their own responses which can be more representative of the participants' beliefs than close-ended questions, as we did not limit their answers. However, the open-ended questions were more time consuming to analyze

[10].

Closed-ended questions were quicker to analyze, as we converted individuals' answers to numbers. Some examples of closed-ended questions are multiple choice selection, true/false, and the Likert-scale which we used in this survey. The Likert-scale we used provided the participants with five reactions to a statement, ranging from strongly disagree to strongly agree with neither agree nor disagree in the middle. This model measured both direction and intensity of attitude towards the statement.

Questions & Items from Outside Sources

While there were no previous studies that addressed the exact purpose of our survey, some studies addressed specific aspects of our leading questions and provided validated inventories we used. The SUSSI survey, mentioned in the literature review, measures beliefs related to the nature of science and scientific tentativeness. As we wanted to measure the individual's understanding of scientific tentativeness, we implemented both open-and-closed ended questions from the SUSSI survey. This included one open-ended question and five Likert-style items, as well as the coding rubric for the first open-ended question. Due to the previous validation and potential for comparison across studies, we changed nothing about the items and questions listed below.

- I1: Science is subject to change.
- I2: It is possible that scientific theories may be completely replaced by new theories in light of new evidence.
- I3: It is possible scientific theories may be changed because scientists reinterpret existing observations
- I4: It is possible scientific theories are subject to ongoing testing and revision.
- I5: It is possible scientific theories based on accurate experimentation will not be changed.
- OQ1: With examples, explain why you think scientific theories do not change OR how (in what way) scientific theories change.

After the section which measured scientific tentativeness, the survey explored participants' attitude towards science. The second half of the Likert-scale items evaluated personal connections to science and levels of trust. Items six, seven, and eight came from The National Science Board's annual survey mentioned in the literature section. We changed the wording of Item 8 during the revision process as the original wording stated "An absolute truth exists and science is the key to discovering it" [3]. We discuss this revision in the revision section below.

- I6: Science is reliable and trustworthy.
- I7: Science, in general, makes life better for the average person.
- I8: Science is an important method for discovering our natural world

Original Questions & Items

We created original questions and items for this survey, and in this section, we described the motivation for these new questions and provided the prompts (listed below). The ninth item addresses one of our leading questions for the study. We explored how educational backgrounds influence beliefs about the tentativeness of science and, in specific, if academic failure played a role. Did the education system link change with failure and what are the repercussions of that connection? So the ninth item stated that being incorrect and asked to change is a sign of failure. The tenth item explored a connection between credibility for personal ideas and the changing nature of science as a potential consequence of the tentative nature of science.

Open-ended questions two and three addressed a different consequence. We asked about participants' reactions to a trusted source reporting on something scientific changing. The one difference between OQ2 and OQ3 is that OQ2 asked about a non-personal scientific idea changing (e.g., general relativity) while OQ3 asked about something scientific that plays into a personal belief (e.g., eating meat affects the planet). We used the same rubric to compare answers of individuals and examined trends. The last open-ended question explored how a participant's education influenced their understanding of the tentativeness of science.

- I9: Education enforces that being incorrect and asked to change is a sign of failure.
- I10: The fact that science is sometimes incorrect provides credibility to personal ideas.
- OQ2: If a reputable scientific source stated that a well-known theory (i.e. general relativity or big bang theory) is incorrect, how would you react to the finding?
- OQ3: If a source you trust stated that a personal core belief (e.g., eating meat is bad for the planet, aliens exist) is scientifically incorrect, how would you react to the finding?
- OQ4: How has your educational experience impacted your opinion regarding the changing/static nature of science?

Once participants finished the Likert-scale section, they were presented with a final item which depended on their previous answers. Item 1 states that science is subject to change and, depending on a participant's answer, they were presented one of two Likert-style items. If an individual agreed that science changes, they evaluated the statement "science is more trustworthy because it changes". If they disagreed, the phrase was "science is more trustworthy because it does not change". These items target the intersection between the tentativeness and trustworthiness of science.

- I11: Science is more trustworthy because it changes.
- I12: Science is more trustworthy because it does not change.

Revisions

The Physics Education Research (PER) group at CU Boulder helped revise the survey. We presented the original questions and discussed them at length to finalize wording and motivations. OQ3, the open-ended question about changing science that impacted a personal belief, was a point of contention. Originally, we gave the participants examples that were more politically charged which was removed to avoid prompting individuals to think about politics. We also entertained a potential ethical element but that was removed because it was not in the scope of our study.

Recommendations and revisions removed the need for ethical or political questions and as the motivation was to compare how people reacted to changing science based on personal connection.

We also made small changes to the wording of several Likert-style items, such as Item 8. The National Science Board framed science as the most important tool for discovering an absolute truth and the PER group disliked the ultimatum. We also disliked the placement of science in competition with other important beliefs about the world. We reworded it to measure the value of science without competing interests. The PER group debated over Item 9 and the difference between 'wrong' and 'incorrect'. We settled on 'incorrect' as 'wrong' implied a moral judgment while the connotation of 'incorrect' was academic in nature. We made an important change to the conditional items which previously stated "science is trustworthy because it changes". The shift to "science is more trustworthy because it changes" avoided the feeling of an ultimatum and allowed for participants to consider other factors.

The PER group focused on revisions for the structural details of the survey. We discussed allowing participants to write comments, the addition of a back button, and the distribution plan. A study proved that using comment boxes had the chance to increase engagement [19] and gave participants the space to individualize their answers. We debated adding a back button as it allowed participants more freedom to express themselves, but also allowed them to change their answers to the open-ended questions as participants answered those first. In the end, we considered the initial reaction from participants more valuable and removed the back button.

Demographics

At the end of the survey, we asked participants about several facets of their identity; gender, race and ethnicity, religious beliefs, political parties, and educational experiences related to science. These questions came last to avoid stereotype threat, which can impact how salient identity groups respond. The only demographic question that did not have a 'prefer not to say' option asked about the participant's educational background.

Our motivations behind the demographic questions were two-fold: to discover ideas held by different populations and to test relationships between social identifiers and their respective understanding of the tentativeness and trustworthiness of science. Our research questions did not explicitly identify gender and race as differentiating variables. However, we wanted to uncover any potential patterns that different social identifiers embodied, hence asking for race and gender. Our data is limited to a primarily Colorado collegiate location and was not fully representative so demographic data allowed us to quantify limitations.

The second motivation explained our questions that request religious beliefs, political parties, and educational background in science. We prioritized including educational backgrounds, as one of our leading questions for this study was how education plays a role in understanding the tentativeness of science. We did not ask for a specific type of science education other than an inclusive collegiate identity, so 'science vs non-science' was a self-identified demographic. As discussed in the prior literature section, political parties play a large role in confidence surrounding science, and we were curious if those patterns would emerge in our study. For religion, we were curious to see if the responses would reflect the theory of religion and science being at odds with each other.

2.2 Context

This section outlines the distribution and response rates of the study.

Distribution

After we created the survey and acquired IRB approval, we started distributing the survey to students via email solicitation. The study targeted people over the age of eighteen because students younger than eighteen must have provided parental permission to engage in Human Subjects Research. Another reason for excluding this population is the concern that they may not have robustly established ideas about the tentativeness and trustworthiness of science to reliably complete the survey. Our main distribution strategy was emailing professors with an introduction and an announcement to post on their class website. The whole email can be found in the Appendix B and the following is the announcement seen by students.

> Recent world events, including the COVID-19 pandemic have emphasized the importance of understanding how people understand the nature of science and trust-

> worthiness of scientific claims. As part of an effort to better understand students' ideas around this, researchers at CU have developed an anonymous survey to solicit ideas. This survey will close mid Feb 2023 and students who complete the survey will have the option to enter into a drawing for one of five 20 dollar electronic gift cards. Due to the anonymity of the survey, instructors will not be informed about participation.

The distribution of emails occurred during the Fall 2022 semester and the beginning of the Spring 2023 semester and resulted in 1541 professors contacted. We emailed some professors twice, during both the fall and spring semesters. During the fall semester, we sent custom emails on a regular schedule to different sections of CU Boulder, such as undergrad natural science classes or business classes. However, because we customized the emails to each professor, we contacted fewer professors. We started by emailing science professors and, because of time consumption, did not email an equivalent number of non-science professors. By the end of the semester, we had an underwhelming number of non-science students in our data set. In order to explore the effect of educational experiences, we needed to change the proportions of science and non-science students in our data set.

Over winter break, we changed the distribution plan to improve the representation of non-science majors. Instead of working through the classes at CU Boulder one at a time, which allowed for the custom emails to professors, we emailed as many professors as possible in a mass email. The only parameter was the class had to be on Boulder campus because we had access to reliable contact information. We targeted both grad and undergrad classes and invited the professors to take the survey. This new approach was successful in recruiting additional non-science respondents.

Response Breakdowns

When we closed the survey, we had 398 responses that were broken down into the various demographic categories. An overwhelming majority of responses were in Colorado which might have contributed to patterns seen in the demographic breakdown in Tables 2.1-2.5.

Science	Non-Science
294	104

Table 2.1: Science and non-science population breakdown (SNS)

We asked the science vs non-science demographic question from Table 2.1 in two different ways, both 'Are you working on a science degree' and 'Have you received a science degree' as the population of the survey is not limited to only students. We aggregated the results. The term 'science degree' was defined by the participants who took the survey.

Religious	Not religious
80	299

Table 2.2: Religious and non-religious population breakdown (R)

We used an open-ended question to ask participants about their religious beliefs which was later coded by hand into one of these two categories. The population breakdown can be seen in Table 2.2.

White	People of Color
281	99

Table 2.3: Racial and ethnicity population breakdown (RE)

We posed the race and ethnicity question as a multiple select option with an other category. Colorado's population is historically white and with CU Boulder as the main population for this survey, we did not receive a large number of responses from any individual racial category. To address this, we aggregated all respondents of color together. We acknowledge that being a race or ethnicity other than white is not a monolithic experience and the distinctions between the experiences of different groups are real and valid. However, to ensure sufficient statistical power, we collapsed these groups here; our results should be interpreted carefully with this limitation in mind. The result of this aggregation can be seen in Table 2.3.

Man	Woman	Non-binary/gender non-conforming
181	187	16

Table 2.4: Gender identity population breakdown (G(All) and G(M/F))

This gender demographics question was single select with an 'other' or prefer not to answer option. We ran into a similar issue with gender that we did with race. There was a small population of non-binary or gender non-conforming responses (seen in Table 2.4) which might lower the power of statistical claims.

Independent	Democratic	Republican	Other
130	154	26	39

Table 2.5: Political party population breakdown (P(All), P(Def), & P(ID))

The breakdown of political parties in Table 2.5 was intriguing and we asked the question as a single select with an 'other' and 'prefer not to answer' option. When registered as an Independent, people in Colorado are given the ability to vote in the primary election of their choice, which potentially leads to the high number of citizens registering as an Independent. In the results section, we used statistical tests to explore the relationship between Independents and Democrats to understand the impacts of this phenomenon.

We would also like to mention the positionality of the main person working on the data analysis as a limitation. They are a white, non-binary, Democrat who is approaching the data as unbiased as possible but recognizes they do not have not lived the experiences of the different populations in this survey.

2.3 Qualitative Rubrics

In the next three sections, we discussed the methods used to analyze data from our open-ended questions and Likert-scale items. We used two different strategies as the data came from open-ended and closed-ended questions. We developed rubrics for the qualitative questions and prepared the quantitative data to ensure that our statistical claims could be powerful.

Rubric Creation

When making the rubric, we could have used something known as an ‘expert rubric’. To create an expert rubric, we would have consulted with experts in the field and recorded their answers. These would have been considered the “best” and the other rubric categories contain different misunderstandings. However, with the nature of our study, we opted out of an expert rubric due to the dependence on value judgment. We wanted to avoid stating that anyone is morally or intellectually better or worse because of their beliefs about the tentativeness so our rubrics did not use expert grading.

As stated in the survey creation section, we took the first qualitative question (OQ1) from the SUSSI instrument. OQ1 asked participants if they believed science changed and to provide examples. The instrument provided not only the question itself but an accompanying rubric. We did not change the question and the rubric underwent only one modification. We only changed the titles of coding categories, as the original used phrases like *naïve* and *expert* [17]. As mentioned above, we avoided using expert-like rubrics, category name change reflected this decision. The rubric’s categories are *no change* (NC: science does not change over time), *change* (C: science changes with new technologies or new evidence), *reinterpreted change* (RC: science changes when previous data is reinterpreted), and *other* (O: no response or can not be categorized).

Our motivation behind the second and third qualitative questions (OQ2 /OQ3) was exploring how participants responded to science changing in different situations. OQ2 asked participants about their reactions when science changed whereas OQ3 asked about changing science that personally affected the participant. We kept the rubric the same to compare the data. An initial idea was to analyze each question on two different scales, one measuring where the participant falls on their acceptance of changing science and the other recording their initial skepticism towards the news. However, we recognized that the question did not prime people to answer in a way that provided enough information. This initial attempt is an example of *a priori* coding, in which we asked a question with a specific grading scheme in mind. Another option for rubric creation is emergent coding, a technique that requires familiarity with the responses before creating a code.

Moving forward, we used a mix of *a priori* and emergent coding to create the rubric for OQ2 and OQ3. Keeping in mind the motivation to understand the responses of how people understood the tentativeness of science in a non-academic setting, the first rubric coded responses as either *reject*, *question*, or *accept*. As we coded the responses, a different pattern emerged. There were responses that accepted the changes to science, "I would be surprised but generally inclined to accept it" or questioned it, "I would view it with some skepticism" however, numerous responses went further with their reasoning. Participants would repeatedly emphasize looking beyond the first source. Whether it was investigating other scientists' thoughts or the source of funding, the pattern of 'with sources' had emerged.

The motivation behind OQ2 and OQ2 and the following modifications helped us complete the rubric: *reject* (R: rejects the new science), *question* (Q: questions or is skeptical about the new science), *question with sources* (QWS: questions or is skeptical about the new science AND mentions different sources), *accept* (A: accepts the new science), *accept with sources* (AWS: accepts the new science AND mentions different sources), and *other* (O: no response or can not be categorized). The lack of *rejection with sources* was purposeful. While we coded the responses, not a single participant indicated they would reject the changing science even after looking at other sources. We also defined what qualifies as 'with sources'. The response needs to mention investigating other sources to either learn more, which would qualify as a *QWS*, or further convince them which is *AWS*. This does not include the participant saying they would read the initial source. A tricky aspect of the rubric is the difference between *QWS* and *AWS*. A couple of data points seem to fit in either category and we made the final decision that, to qualify as an *AWS* response, it has to refer to acceptance or some derivation of acceptance.

We encountered an interesting complexity when coding OQ3. As a reminder, OQ3 differed from OQ2 by making the changing science related to a personal belief. We prompted participants to think of a scientific belief with examples like eating meat or the existence of aliens and the statement 'is scientifically incorrect'. So participants commented on their reactions to both their personal beliefs and science changing. It became clear there were two aspects we could capture -

the belief, the science itself, or both. For example, a participant answered, "Depending on what was stated, I would believe it but possibly ignore it." In the end, we created this question to measure how people reacted to science changing when it personally affected them, so we noted the belief aspect of responses but did not factor it into the coding. As for participants who responded to how the changing science would influence their belief but did not mention their reaction to the science itself, we coded those as acceptance. If an individual mentioned how their belief would shift, we assume they did so with an acceptance that the science is correct. Approximately 35% of the data points that we coded as acceptance were based on this assumption. An example follows: "I would trust them, but depending on the information may continue to live my life unchanged. Like the example that eating meat is bad for the planet, I am still going to eat meat."

The last rubric created for OQ4 used emergent coding. We created OQ4 as a response to one of the main questions that defined the study: what role does education play in the relationship between the tentativeness and trustworthiness of science? Unlike the previous questions, which were taken from a previous survey or created with a rubric in mind, OQ4's creation did not come with a rubric. So the categories - *no change* (NC: education had no impact on their understanding of changing science), *unspecified change* (UC: education had an impact but the impact is unclear), *tools* (T: education provided tools for understanding science), *static science* (S: education taught them science does not change), *changing science* (C: education taught them science changes), and *both* (B: education taught them science does and does not change) - emerged while we read through the responses.

The *NC* category came from people who answered along the lines of "Not at all, I've had a relatively constant understanding of how I should think about things and come to my beliefs". Their educational experience did not impact their beliefs. Other responses also did not address the impact their education experience had on their beliefs about changing science but differed from *NC*. That led to the unspecified change and tools category. The *UC* category existed for responses that indicated education affected their view of science, but either lacked detail or did not address the nature of science prompt. We noticed a need for a new category after a pattern emerged from the *UC* group. While some people remained unspecific, there were several responses that mentioned that their educational influences gave them tools to understand science. Examples include an increase in critical thinking, investigating funding and sources, and understanding manipulated data. So while the *tools* category does not address the prompt, it is a point of interest. The *static science* category exists for people who stated their educational experience taught them science does not change, the *changing science* category opposes that, and the *both* category is for people who expressed both concepts in their answers. All of the rubrics along with prime examples can be found in Appendix C.

Each rubric went through an interrater reliability test (IRR). The IRR test took multiple reviewers ratings for the same question and tested the consistency through percent agreement. There were two reviewers and while the percent agreement test lacked in its ability to calculate chance or acknowledge the intensity of the agreement, it worked for what we needed. The higher the percent agreement, the higher the consistency between the two raters and the potential reliability of the rubric. The results follow in Table 2.6 and OQ1 had the highest internal consistency. Since OQ1 was taken from a previous instrument that had been tested for reliability, the high percent agreement did not shock us. OQ2 and OQ3 were the lowest of the four with a large discrepancy between *question*, *question with sources*, and *accept*. As a reminder, OQ2 and OQ3 asked for participants reaction to changing science. For both OQ2 and OQ3, the reviewer coded more responses into *question* than *question with sources* and fewer responses overall into *accept*.

IRR	% Agreement
OQ1	90%
OQ2	75%
OQ3	72%
OQ4	85%

Table 2.6: IRR Percent Agreement

2.4 Data Types

There are two types of data that results from a Likert-scale item offered - ordinal or interval. These are two of four data types, the other two are ratio and categorical. Open-ended questions generate categorical data as it defines data that is typically qualitative and sorted into arbitrary categories. Ratio data are not considered in this thesis. For data to be considered ratio it must be numerical, continuous, and have a true zero. The magnitude of ratio data can be compared and fractions concerning variables are possible. Likert-scale data points are not continuous nor do they have a true zero. Categorical and ratio data define two extremes of data definitions from least restrictive to greatly restrictive. Ordinal and interval have their places along this scale but arguments can be made Likert-scale results as either.

Ordinal data is closer to categorical data. The categories are still discrete but are now placed in a definite order which makes it suitable for a Likert-scale. While the data is not numerical, there is an order from strongly disagree to strongly agree. It would make sense to argue that a person who selected "strongly agree" could have also selected agree. Interval takes these ordered categories and assumes an equal spacing between the options. Treating likert-scale data as interval is quite a big assumption but also allows statistical tests to be run like average and variance. However, it is unlikely that a person taking the survey would have equal spacing between strongly disagree to disagree and disagree to neutral. The assumption that a person does value that spacing equally was never prevalent in our data so the Likert data was treated as ordinal.

The next question around that data analysis was to treat the data set as a five point scale or collapse it down into a three point scale. Allowing participants to choose between strongly agree or agree decreases the default selection of "neutral". However, for the purpose of this study, the difference between strongly agree and agree don't fundamentally shift the claims we might make. Collapsing down to a trichotomous scale has also been shown to have a negligible impact on the reliability or validity of the data [23]. There is also a tendency for the likert-scale to overestimate extreme measurements, the strongly agree and strongly disagree options [1]. So this limitation is removed when the data are collapsed because the extreme measurements are combined with the less extreme. The motivation to collapse down to a three point scale is to have more power in any claims we make from the data and will be explored further in the results section.

2.5 Quantitative Methods

The majority of the data analysis for this study fell into the realm of statistical hypothesis testing. Statistical hypothesis testing is a type of analysis that decides whether the data supports a chosen hypothesis or rejects it. This strategy is ideal for determining if there is a relationship between two variables and in this case, the variables refer to specific populations' responses. The ability and power of the hypothesis depends on the type of test being used which will be explored later in this section. We also used descriptive statistics, which do not make specific statistical claims about the data but help to describe trends and patterns in the dataset. This was valuable as we explored the relationship between education and the tentativeness and trustworthiness of science.

The first step we took was deciding between interval or ordinal data to narrow down the potential statistical hypothesis tests. Interval data required a collapsed likert scale and used t-tests or ANOVAs. However, as mentioned in the data types section, we could not prove equal spacing between *agree*, *neutral*, and *disagree*. So interval data tests were short lived and our attention turned back to ordinal data tests.

With all the data about social identities that we gathered, there were a lot of combinations to test. Educational background, race and ethnicity, and religion were all dichotomous yes/no variables which made it easier to determine if there were correlations between participants' responses and various aspects of their self-reported identity. Gender identity and political ideology were a bit more complicated. As mentioned in the context section, we had a fair amount of responses from male and female identifying people but statistically fewer non-binary/gender non-conforming people. The statistical power of tests is highest when the variables have equal sample sizes so the tests were run with just male as well as female identities and all identities so we could see if the size plays a significant role. For political parities, we ran tests only with Independents and Democrats as they were the largest categories and tests concerning Independents, Democrats, and Republicans (which excluded participants who identified with an "other" party), and a test with all parties.

Before moving on to the conducted tests, there are some vocabulary terms that should be addressed. The null hypothesis always predicts that there is no relationship between two variables. It can either be accepted or rejected depending on the alpha value. An alpha value is the threshold for statistical significance, commonly 0.05. This translates to the p-value which is the probability that the two variables don't have a relationship. If the calculated p-value is under 0.05, that means there is a only 5% chance that the null hypothesis is correct and the combination is determined to be statistically significant.

With the knowledge that we were working with ordinal data, questions to explore with our variables, and the vocabulary to support us, the next step was finding the proper test. The default test to use in correlation testing is the chi-square test of independence.

$$\chi^2 = \sum \frac{(obs - exp)^2}{exp} \tag{2.1}$$

Our chi-squared test needed a data set that represented the population, ordinal data points that were not continuous, and at least five points of data in each combination. This last proved difficult but we continued working with the chi-square test for a rough idea of what claims could be made. The chi-square test worked by calculating expected values from the observed measurements and comparing them to a value from a chi-squared distribution. We determined the distribution from our chosen alpha value and the degrees of freedom from both variables. If the sum of the expected calculations were higher than the distribution value, we rejected the null hypothesis.

As mentioned above, one of the chi-square test requirements is for each combination of variables to have $n = 5$. Only a couple of our data sets fulfilled this requirement so we needed another test. Fisher's exact is a statistical hypothesis test that has the power to reject the null hypothesis but is set up so even small sample sizes are valid. Even with combinations where n is less than five, the p-value is statistically powerful. While chi-squared calculations of expected values and the test value was relatively simple, Fisher's exact uses binomial coefficients and factorial operators and the 2x2 version of it can be seen below.

$$p = \frac{\binom{a+b}{a}\binom{c+d}{c}}{\binom{n}{a+c}} = \frac{\binom{a+b}{b}\binom{c+d}{d}}{\binom{n}{b+d}} = \frac{(a+b)!\,(c+d)!\,(a+c)!\,(b+d)!}{a!b!c!d!n!} \tag{2.2}$$

The a, b, c, d, n refer to variables that seen in the table below.

w	x	c = w + x
y	z	d = y + z
a = w + y	b = x + z	n = a + b + c + d

Most of our data requires tables bigger than 2x2 which makes the p-values no longer doable by hand. These multivariate generalized calculations become unwieldy but possible with the statistical software package R.

After calculating the p-values for 128 statistical tests, we addressed the errors that arose with multiple comparisons. The more null hypotheses that we tested, the higher the probability that we calculated false positives. The Holm-Bonferroni test corrects for this by adjusting the criteria for individual hypotheses by calculating a new p-value for each hypothesis that takes into account the number of tests overall.

$$p_k < \frac{\alpha}{m+1-k} \tag{2.3}$$

The m refers to the amount of p-values being calculated and the k refers to it's position in a sorted order. The α is the 0.05 value that was discussed above. At first, we calculated the Holm-Bonferroni with every test to find the most significant p-value in the entire data set. However, using the test on every item at once was too limiting so we went item by item to find the true statistically significant relationships.

The only question remaining after Fisher's exact tests and Holm-Bonferroni corrections was what the tests are claiming. Was there a chance that the statistical hypothesis tests were looking at differences between strongly agree and agree? How much did we value that difference? In the end, we argued that we wanted to know if a population has statistically significant relationships between people who fall in either of the *agree* categories, *neutral*, and people who fall in either of the *disagree* categories. This was done by collapsing the data to a three point scale instead of a five point scale. We also removed any potential to over or under-estimate the intensity of a participant's response because we no longer consider strength. As covered in the data type section, collapsing the likert-style data has quite a few advantages when making statistical claims and we explored the effects of collapsing the data in the next chapter.

Chapter 3

Results

This chapter of the thesis stated the results and interpretations of the statistical claims. We started by arguing for a collapsed data set as it proves more statistically powerful and cleaned up our data. After presenting the overall likert responses, we made claims from the statistical hypothesis tests and found trends in the Likert-style data which we then interpreted. We displayed the results from the open-ended questions along with intriguing patterns, engaging examples, and descriptive statistics that connect the questions.

3.1 Collasped Data

This section explored the differences between the statistical data when the data is not collapsed (Table 3.1) and collapsed (Table 3.2). The collapsed Likert data combines the strongly agree/agree and strongly disagree/disagree into two respective categories. This granted the statistical tests the power to ignore differences between intensity and provided a larger N to most variable combinations, increasing the power. The change proved statistically significant.

The items that start with I in the tables correspond to the Likert-style items while OQ represent the open-ended questions. The labels on the right correspond to the population group breakdown which can be found in the methods section. For brevity, we've included the meanings here: SNS: science background, R: religion, RE: race & ethnicity, G(M/F): gender breakdown only included men and women, G(All): gender breakdown with all identities, P(ID): political parties with only independents & democrats, P(Def): defined political parties (i.e. independents, democrats, &

republicans), P(All): all political parties.

Table Color Scheme		
Signficant in both 5pt & 3pt	**Insignificant → significant**	**Significant → Insignificant**

p-values: 5 point scale																
	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	OQ1	OQ2	OQ3	OQ4
SNS	0.1	0.3	0.6	0.3	0.5	0.2	0.4	0.4	0.1	1.0	1.0	1.0	1.0	1.0	1.0	1.0
R	0.2	0.6	0.3	1.0	1.0	1.0	1.0	0.1	0.7	0.6	1.0	1.0	1.0	1.0	1.0	0.1
RE	1.0	0.8	0.3	**1.0**	1.0	1.0	1.0	0.3	0.2	0.7	1.0	1.0	1.0	1.0	1.0	0.6
G(M/F)	0.2	0.3	0.3	0.5	1.0	**0.001**	**0.001**	0.1	**0.02**	0.4	0.1	1.0	0.1	1.0	1.0	1.0
G(All)	1.0	0.3	0.8	1.0	1.0	**0.002**	0.4	0.4	**0.3**	**1.0**	0.7	1.0	0.7	1.0	1.0	1.0
P(ID)	0.2	0.3	0.3	0.4	0.6	**0.01**	**0.003**	0.3	**0.5**	0.5	0.2	1.0	0.2	1.0	0.2	0.2
P(Def)	1.0	0.6	**0.3**	1.0	0.4	**0.4**	**0.01**	0.3	0.7	**1.0**	**0.03**	1.0	0.06	1.0	0.4	0.02
P(All)	1.0	0.3	0.6	**0.004**	0.7	0.3	0.8	0.4	0.7	1.0	**0.01**	1.0	0.1	1.0	0.6	0.03

Table 3.1: Uncollasped p-values for Comparison

p-values: 3 point scale																
	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	OQ1	OQ2	OQ3	OQ4
SNS	0.2	1.0	0.6	1.0	1.0	0.1	0.2	1.0	0.1	1.0	0.7	1.0	1.0	1.0	1.0	1.0
R	1.0	1.0	0.1	1.0	0.7	1.0	0.2	1.0	1.0	0.5	0.2	1.0	1.0	1.0	1.0	0.1
RE	1.0	0.2	0.3	**0.002**	0.7	0.1	0.4	1.0	1.0	1.0	0.6	0.8	1.0	1.0	1.0	0.6
G(M/F)	1.0	1.0	0.5	1.0	1.0	1.0	0.5	1.0	0.5	0.1	1.0	0.9	1.0	1.0	0.6	1.0
G(All)	1.0	1.0	0.6	1.0	1.0	1.0	0.5	1.0	**0.04**	**0.04**	1.0	0.9	1.0	1.0	0.6	1.0
P(ID)	1.0	1.0	0.6	1.0	1.0	**0.01**	0.3	0.3	**0.04**	0.2	1.0	1.0	0.2	1.0	1.0	0.2
P(Def)	0.3	0.7	**0.048**	1.0	1.0	**0.03**	0.1	1.0	0.1	**0.04**	0.4	1.0	0.4	1.0	1.0	**0.02**
P(All)	0.5	0.7	0.1	1.0	0.3	0.1	0.1	1.0	0.2	0.1	0.6	0.8	0.6	1.0	1.0	**0.03**

Table 3.2: Collasped p-values (Black Highlights w/ White Text are Statistically Significant)

As seen above, some of the relationships between variables that were statistically significant changed when we collapsed the data. Some combinations went from being statistically significant to insignificant (dark gray) and some combinations experienced the reverse (light gray). There are only three p-values that stayed significantly significant during the collapse (I6:P(ID), OQ4: P(Def) & P(All)) which we shaded with black.

Given that these p-values were already corrected to address multiple statistical tests (Holm-bonferroni), we did not believe this change was driven by random statistical fluctuations. Instead, p-values that went from significant to insignificant could be influenced by differences between individuals selecting agree or strongly agree, a distinction that is lost in the 3pt scale. I1: G(M/F) is an example of this trend, where a quick chi-squared test results in a p-value $= 3 * 10^{-6}$ (Chi-Squared) for the difference between men and women selecting agree vs. strongly agree. This might reflect a difference between men and women in the intensity of their beliefs, or it could reflect a difference in men and women's threshold for what strongly agree means. When we adopted the 3pt data scale, we removed the confounding variable of intensity. Another example of intensity; the p-value for I7:P(ID) goes from statistically significant at $p = 0.003$ (Fisher's Exact, HBF) to not significant at $p = 0.3$ (Fisher's Exact, HBF). We ran tests for significant differences in intensity and both the difference between strongly agree and agree and the difference between strongly disagree and disagree were significant.

There were also relationships that became statistically significant when the data was collapsed that were not significant on the 5pt scale. Collapsing the data gives the Fisher's Exact test more statistical power because of the larger N in each individual bin. The test used the extra power to parse out differences in the data that were previously hidden. For the reasons above, we used a 3pt scale for the rest of the results and interpretations.

3.2 Likert Results

This section discussed the results from all the Likert-style items. We introduced some overall trends and limitations of the data before stating claims. These claims were backed up by statistically significant p-values, information we found in the literature review section, and quotes from students. We then explored the trends related to the claims and proposed potential interpretations.

As seen in Figure 3.1, eight items had over 80% of our participants agree with the statements. This limited our statistical analysis because the N in different variable combinations is lower. It also indicated that, for the overall population, the statement was not controversial and therefore, the item does not act as a predictor. Due to this pattern, items nearer to the top of Figure 3.1 had less statistically significant p-values, the only expectation was I4 which we discussed later in

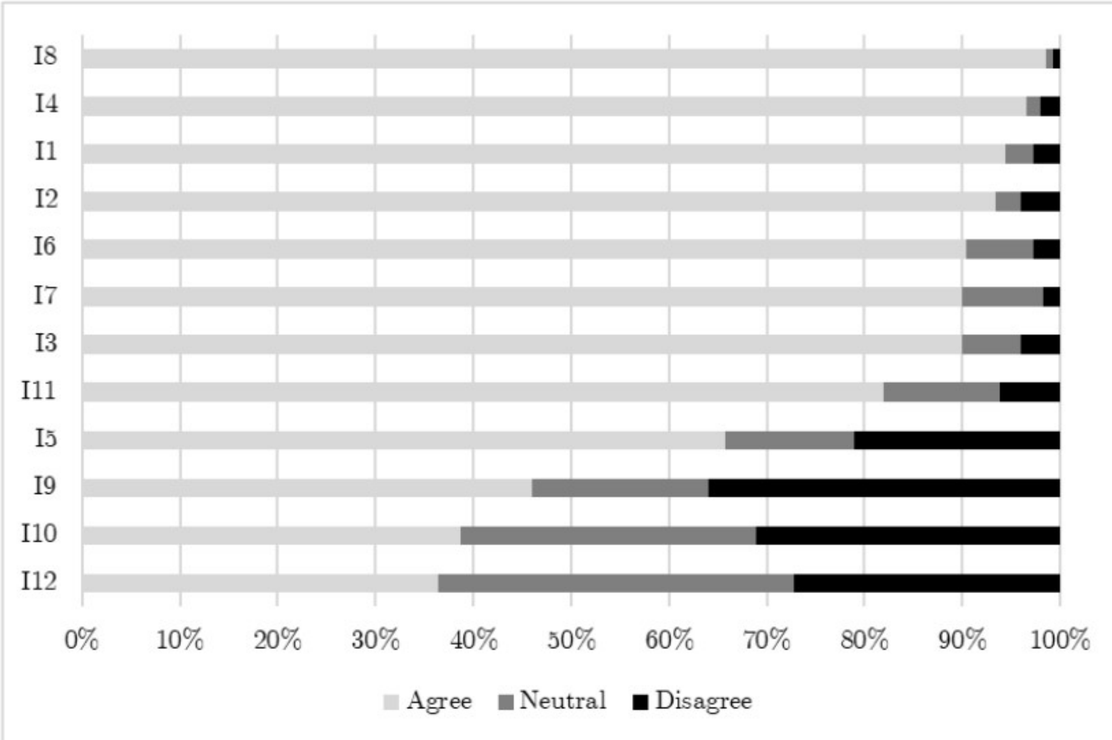

Figure 3.1: Collapsed Likert Response Rates

this section. Questions at the bottom of Figure 3.1 were more likely to be statistically significant in certain populations, especially political parties. I12 is an outlier as this was our second conditional question and only 10 people answered it. So it appeared statistically valuable, but it held less power than any other question.

Educational Background

In this section, we discussed the findings related to one of our leading questions: how does personal educational experience impact the ability to reconcile the tentativeness and trustworthiness of science? In the split between students who identified as having or pursuing a degree in science relative to those who did not, we saw no statistically significant differences on any of our questions. This suggests that, for our population, engagement with formal education in science is not a significant predictor of students' ideas about the tentativeness and trustworthiness of science.

We explored a couple of ideas as to why we did not detect any correlations. Our main idea was that the concept of changing science was not covered differently in science versus non-science classes for this population. Or, the differentiator between science and non-science was based on the participants' self-identified college experience. Perhaps the K-12 NGSS standards related to the nature of science were influential and uniform, and college education did not change their beliefs. For example, the NGSS argues that students should be taught the scientific method and learn that science is an iterative process. OQ4, the open-ended question which asked about educational experience, did not specify a time frame and there were no emerging patterns surrounding K-12 versus college education. Figure 3.2 was from OQ4 and it depicted the difference between the observed and expected frequencies from people who we coded into the *changing science* category. There is not a large difference between the two frequencies which implies a lack of correlation between educational influences and beliefs about the tentativeness of science. This lent credence to the idea that college classes did not treat the topic of scientific changes differently so individuals' attitudes towards science were not affected by their classes.

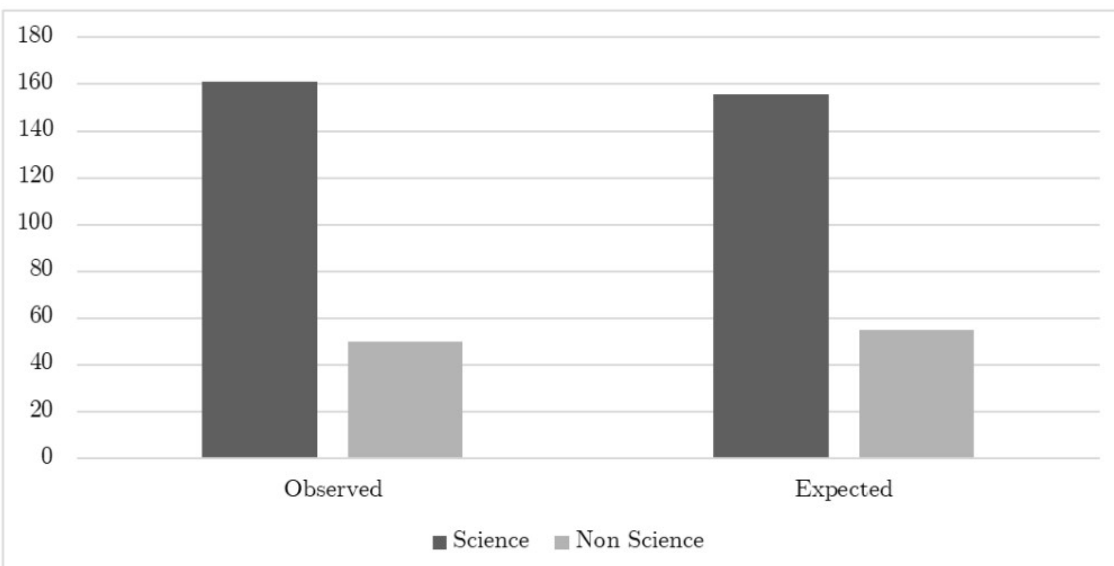

Figure 3.2: Frequencies of *Changing Science* Selection from OQ4

Or, as 21 people stated in OQ4, educational influences did not happen in the formal classroom. There were a variety of individuals who mentioned family or outside lab work. We added quotes that represent the family and lab work:

- "My educational experience hasn't impacted it as much as my mom has. She is a scientist and has worked at NASA so I hear a lot of stories from her and my beliefs in science and the changing nature of science stem from her."

- "My formal educational experience hasn't, to the degree that my clinical work experience has. Clinical trials often inform clinical practice, and practice norms are frequently re-examined."

Political Parties

After we recognized the lack of educational experience as a predictive factor, we took a holistic approach focused on overarching patterns while we analyzed the p-values. With this mindset, we noticed the main social identifier that resulted in statistically significant differences on multiple questions was related to political parties. The questions I3, I6, I10, and OQ4 all have p-values less than 0.05 which made them statistically significant.

I3, (p(P(Def)) = 0.048 Fisher's Exact/HBF) focused on the idea of science changing due to reinterpretation of existing evidence. The specific political party breakdown involved in this difference is P(Def) which includes Independents, Democrats, and Republicans. The normalized data reveals a similar breakdown between Independents and Democrats. Republicans, while they still reported *agree* more than any other answer, tended to report feelings of neutrality more than the other parties.

I3	I	D	R
Disagree	5%	3%	0%
Neutral	5%	3%	23%
Agree	91%	94%	77%

Table 3.3: Percentage of Normalized Frequencies of P(Def) for I3

The visible trend we found from Table 3.3 was more Republicans reporting feeling neutral about the tentative nature of science implies that once scientific results are published, since Republicans might be less inclined to accept reinterpretations of this data. An interpretation from this potential skepticism related to distrusting science. If someone saw initial experimental data, they might not be interested in finding different interpretations or accepting other ideas. So when other results based on reinterpretation of this data were published, some Republicans might have found it harder to accept than other parties. This could lead to a feeling of distrust in the new results.

The next statistically significant question, I6 (p(P(Def)) = 0.03 Fisher's Exact/HBF) asked for a level of agreement to the statement "science is reliable and trustworthy" and there was a statistically significant difference across our three defined classifications of political parties. It is important to note that the PID test, which only tested the Independent and Democratic populations, was also statistically significant (p(P(ID)) = 0.01, Fisher's Exact/HBF), which suggested that this detection was not just a difference between the two dominant political parties. This seemed to align with the University of Chicago study from the literature review that reported an increasing gap between Democrats and Republicans relating to their confidence in science.

I6	I	D	R
Disagree	3%	1%	4%
Neutral	12%	2%	8%
Agree	85%	97%	88%

Table 3.4: Normalized Percentages of P(Def) for I6

The University of Chicago study spoke to the trend that Republicans had less confidence in science and that trend is continued in this study, but with a caveat. All parties including Independents had a majority in agreement with the I6 statement. However, our trend did not show as much of a difference between Independents and Republicans, and Independents are the party with the lowest percentage of agreement.

As for the interpretation, we noticed in Table 3.4 that Democrats are more likely to agree that science is trustworthy. Another study from the PEW research center found that more Democrats than Republicans trust the scientific method to return accurate results [12] and, acknowledging that the scientific method is iterative, our results aligned with their findings.

I9 explored the relationship between education enforcing the concept of change as failure and these results suggested that the individual's political party can be used to predict their answer.

However, the claim was only significant when we compared Independents and Democrats (p(P(ID)) = 0.04, Fisher's Exact/HBF).

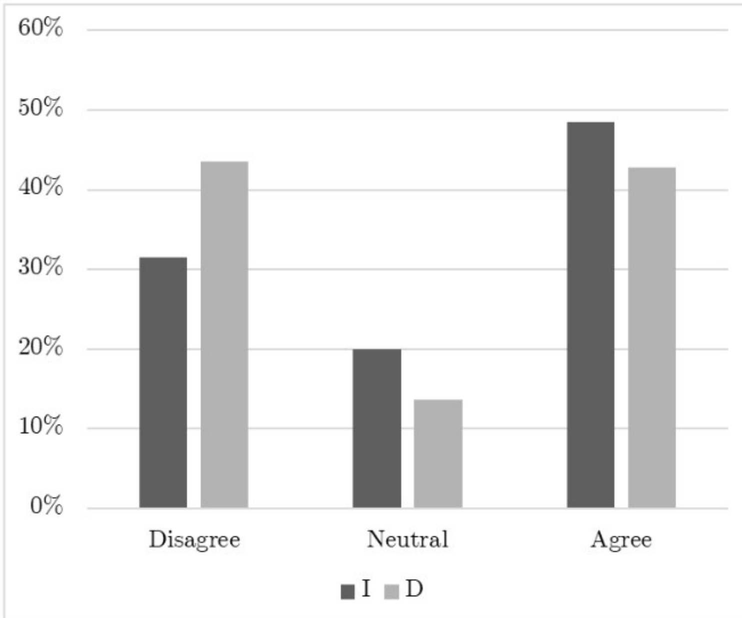

Figure 3.3: Normalized Responses to I9 Broken Down by P(ID)

When we looked at Figure 3.3, we noticed that Democrats were more likely to disagree with the idea that education enforcing change is a sign of failure. There were a few participants who commented on this question and here are some quotes from both Independents and Democrats that expand on a potential trend:

- I/Neutral: "I think that some education frames incorrectness as failure, but good education does not"

- I/Neutral: "How our grading system in the American education system is being ran right now, being incorrect in a lot of students eyes is failure, because we see our grades dropping even if we put in the work and hours of studying."

- D/Disagree: "Good education stresses learning from mistakes and failures. Bad education (which is mostly the case!) is punitive."

From these quotes and others, a possible interpretation was as follows. Democrats who accept the possibility of growing from mistakes and acknowledge that some schools did not allow for this mindset were more likely to disagree while Independents expressing the same idea selected neutral. However, not enough participants used the text box option and mentioned 'good' or 'bad' education for this trend to be considered strong. We still reported these quotes and considered looking into 'good' vs 'bad' education for future work.

I10 was the final likert-style question that is significant (p(P(Def)) = 0.04, Fisher's Exact/HBF). It posited that incorrect science provides credibility to personal ideas and the main difference found was between Republicans and Democrats/Independents. It was clear from Figure 3.4 that Republicans were less likely to disagree with this statement compared to both Democrats or Independents. Democrats, on the other hand, were more likely to disagree than the other two parties.

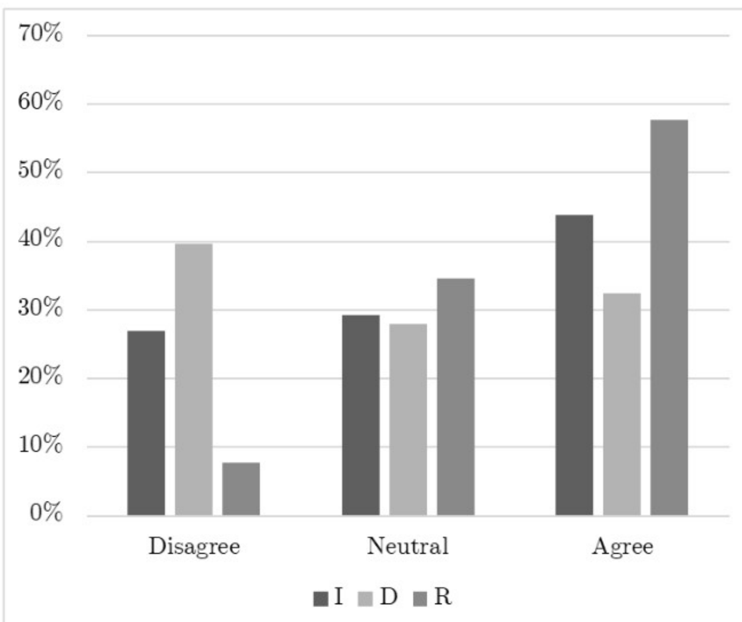

Figure 3.4: Normalized Responses to I10 Broken Down by P(Def)

This item was one of our only items set up to discover potential consequences of not understanding the tentativeness of science or individuals who distrust science. As we saw in Figure 3.4, Republicans were less likely to disagree with the consequence of additional credibility to personal ideas. Potentially, Republicans might be more willing to make claims that are not based on science if they have experienced incorrect science.

Other Significant Predictors

There were a couple other items that were statistically significant with differing social identities. A person's race and ethnicity was related to their belief that scientific theories are subject to ongoing revision (p(RE) = 0.001, Fisher's Exact/HBF).

I4	White	People of Color
Disagree	1%	5%
Neutral	0%	3%
Agree	99%	92%

Table 3.5: Normalized Percentages of RE for I4

From the percentages seen in Table 3.5, people of color are less likely to agree that scientific theories are undergoing consistent revision though they still overwhelmingly agreed. One possible interpretation of this trend acknowledges the history that people of color had with science such as eugenics and previous scientific atrocities (e.g. the Tuskegee syphilis study). Aspects of science were rooted in racial inequities that resulted in myths such as black people have a higher pain tolerance. It was possible that the people of color that disagreed with this statement reflected on this history and their personal experiences which indicated a lack of ongoing revision. However, this is only one of many possible interpretations that we are not well positioned to make based only on quantitative data.

The social factor of gender showed up as significant in I9 and I10 which were also discussed above alongside political identity (pI9 = 0.04 Fisher's Exact/HBF & pI10 = 0.04 Fisher's Exact/HBF). I9 asked about education enforced change as a sign of failure while I10 asked if incorrect science provides credibility to personal ideas. Something important to take note of is the population which includes all gender identities. Only 16 people identified as non-binary or gender non-confirming so it was possible that these results are from disproportionate N values.

From Figure 3.5, we saw that non-binary and gender non-conforming people are less likely to disagree with statements from both I9 and I10 and are more likely to be neutral. They agreed at

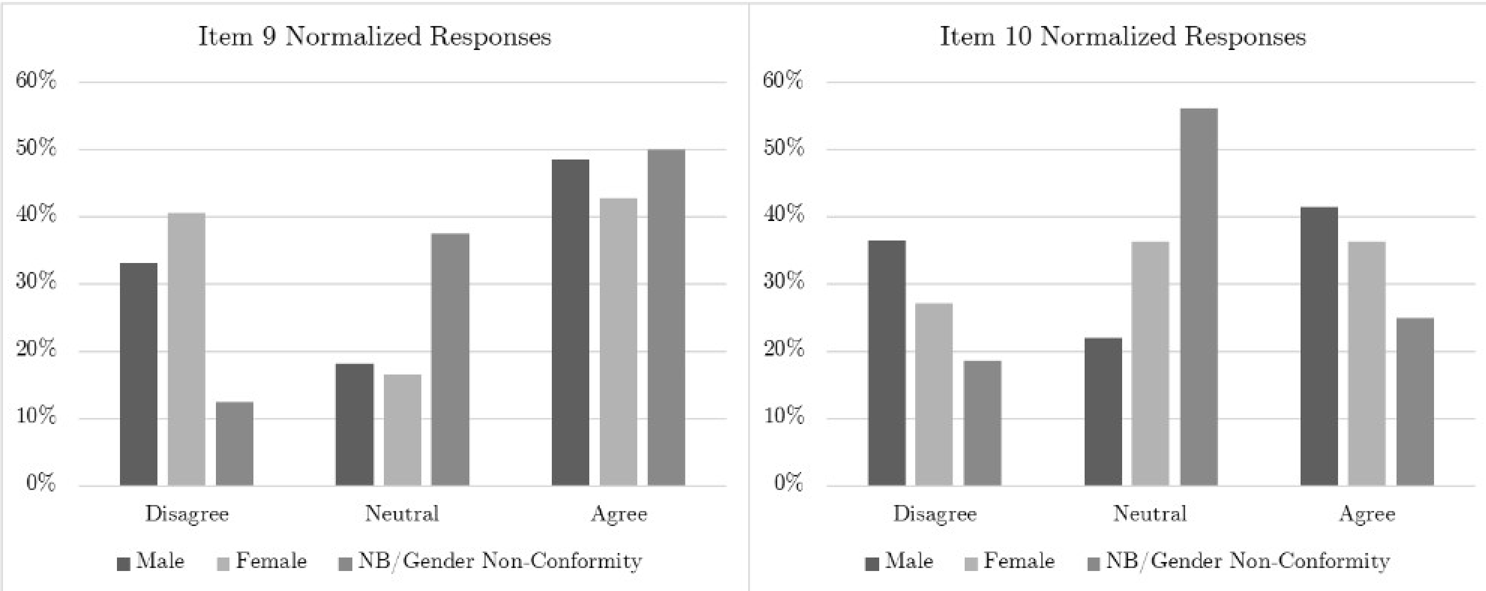

Figure 3.5: Normalized Responses from I9 & I10 with Gender (All) Breakdown

the same rate as men and women in I9 but are least likely to agree in I10. However, due to the small N and multiple conflicting factors and definitions that fell under the non-binary and gender non-confirming label, these results do not support robust interpretation.

Conditional

We broke down the setup for I11 and I12 which helped with the analysis. Both items assumed that someone who chose *agree* or *strongly agree* trust science and their trust was aided either by science remaining static or science changing. Figure 3.6 displayed the breakdown by response to these questions. It is important to note that, since I11 and I12 were conditional items, they were only shown to participants who answered in certain ways on I1. Due to this, I11 was answered by 376 individuals while I12 only had 10 responses (one person misclicked and their data was removed). The other 11 individuals did not see either question because they selected *neither agree or disagree* for I1. So any claims that were made are not as powerful as the previous ones because of the small N value for I12.

We noticed that a majority of people who answered I11 agreed to the concept that the changing nature of science made it more trustworthy. Similar to the other likert-style items, these items had a text box to elaborate on anything else. We selected quotes that explored more about why people agreed or disagreed.

- Disagree: "Science has limitations and is never completely trustworthy. In this sense, because it changes, it shows a history of fallibility and error."

- Agree: "Science is able to adapt to new technology and perspectives, which makes science more inclusive to applications in the real world. Because science changes, it can become more accurate for our world."

- Agree: "I would trust someone who is willing to admit they were wrong than someone willing to die before they admit fault in their work. I view science the same way."

The first quote used the tentativeness of science to prove that science is fallible and therefore, not reliable while the other two quotes acknowledge that changing science either allows it to adapt to the modern world or is more honest. We believe that with the majority of people agreeing, it was acceptable that the changing nature of science made it more trustworthy to CU Boulder students.

As for I12, there was more of an even spread and no one left comments in the text box to

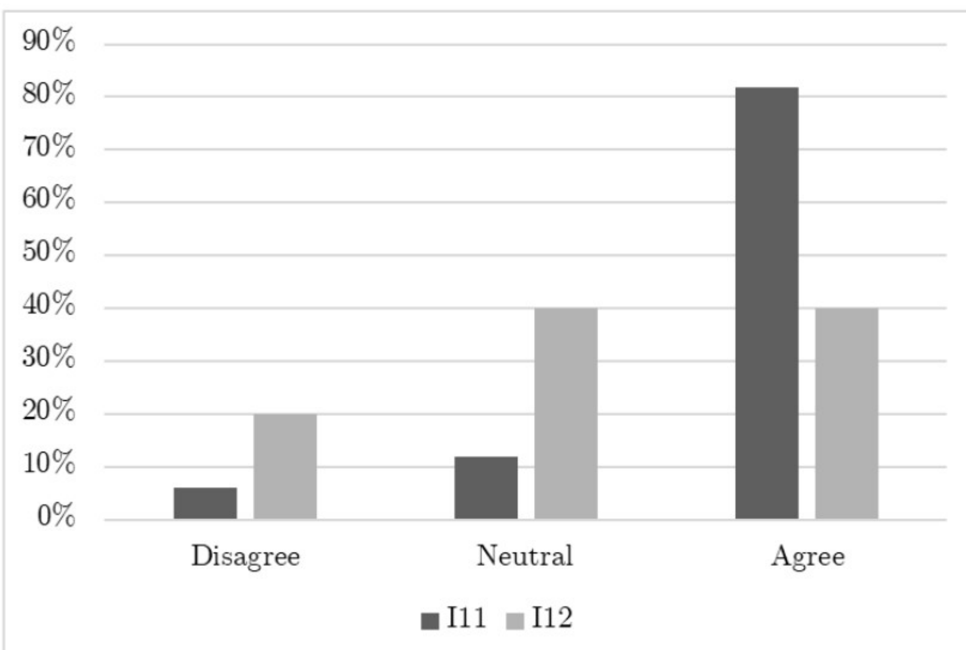

Figure 3.6: Breakdown of Conditional Item Responses

further explain themselves. Due to the small sample size represented by I12, we will not make any interpretations about this data.

3.3 Qualitative Results

In this section, we uncovered results and patterns from the four open-ended questions. We used rubrics and coding in order to use statistical testing to find statistically significant relationships. Open-ended questions lent themselves to descriptive statistics and quotes so we use those to discover trends and support interpretation.

Open-Ended Question 1

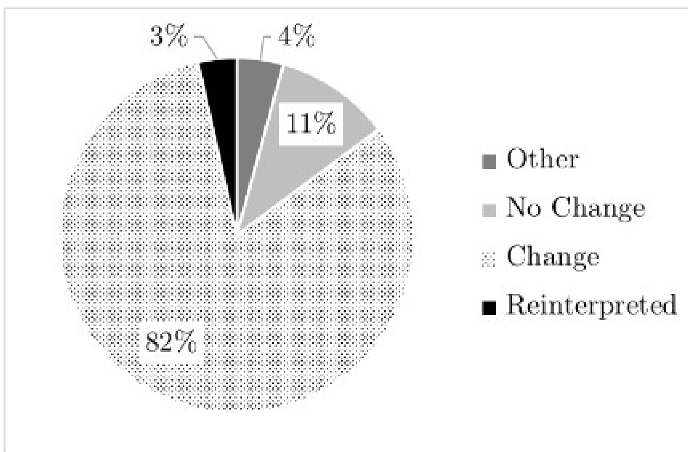

Figure 3.7: Response Breakdown OQ1

As seen in Figure 3.7, the most popular response across all populations was "scientific theories may be changed when experimental techniques improve, or new evidence is produced." We also asked participants to provide examples of science changing or not changing and 91 participants did. We found the most popular examples were the switch from geocentrism to heliocentrism (n=22), the concept of gravity (n=12), the theory of evolution (n=10), the structure of the atom (n=9), climate change (n=9), the concept of relativity (n=9), and the flat earth theory (n=9). COVID was used as an example but rarely (n=5). As a reminder, the email announcement sent to professors for their classes prompted COVID as a motivator.

Most of these examples correlated to a belief in changing science except for two: evolution and gravity. For evolution, two people who used it as an example were coded into the *no change* category, here are their responses.

- "I think scientific theories don't change because to become a theory it has to undergo many experiments or observations and have to be proven more than once to become a theory. An example would be evolution, that theory had many many trials before becoming a theory."

- "I think scientific theories do not change because, in order to be classified as a scientific theory, they go through a long process of research and studying. Once formed, attempts to find evidence to prove or disprove the theory often only add to its credibility. For example, the theory of evolution already had a significant amount of evidence to back it up when Darwin published it. Since then, many have tried to disprove it, but so far, we have only gained support for the theory."

Both of these responses fell under the same belief that, to be considered a theory, there must be a lot of evidence and testing before something is considered a theory. A similar pattern can be seen for gravity. Three non-science responses were coded as *no change* and used gravity as proof. They followed a similar train of thought as the two evolution examples above, if the theory hasn't been disproven so far, it will not change. We believe this is impacted by the textbook definition of a theory. An article from "The Science Teacher" explores the treatment of theory in textbooks and finds that, even in a textbook that includes a section about changing theories, it confuses observations, inferences, and theories together. It also presents scientific information as if it was a permanent truth [16].

The last interesting pattern that emerged is the prevalence of a flat earth theory. The following quote is a prime example of where it shows up:

> "They change based on new present information discovered. WHat was known as a fact a hundred years ago may not hold true with the advancements made today. Such as how it was common belief that the earth was flat, but then as time and technology advanced many came to question and refute this theory."

Every participant who used flat earth as a theory was coding in the *change* category and stated that it was a historical theory disproved by other ideas and there was no uniform social identity group they fit into.

Here are some quotes that revealed another trend; an idea that science doesn't change at a base level. This showed up in both this question and OQ4, which asked about educational influences.

- "Theories never change the base ideas but instead mold to the available data. As new data is collected theories develop, but without new data, theories stagnate." - OQ1

- "Over all it has shown me that science is unique in how most of the fields pertain to facts that are ever changing as new research is conducted then there are other fields that stay pretty stagnant with new breakthroughs but the core knowledge of the field is still the same" - OQ4

The OQ4 individual did address the difference in scientific fields but still held that belief that some science has a 'core knowledge'. We entertained this as a reason why people don't react well to science changing as their definition of 'base science' might include the changing aspect. For example, the science of physics was thought to be completed and then scientists discovered quantum mechanics which shifted our understanding of the whole field.

Open-Ended Question 2 & 3

As stated in the survey design section in methods, we created OQ2 and OQ3 for comparison. The response breakdown is shown in Figure 3.8, and we had quite a few interpretations. We started with a quick evaluation of each item separately.

OQ2 focused on science changing in a way that does not interfere with the participant's life. The breakdown is illustrated in Figure 3.8 with *questioning with sources* as the favorite choice and *questioning* and *accepting* without sources not far behind. Participants voiced several opinions that are not illustrated by the categories, including conflicting ideas about the age of theories relating to their stability and participants questioning what it means to be incorrect.

- "I would ask them why they think that, and to what degree they think it is incorrect. (I believe general relativity is incorrect, and that it's only accurate to about 16 decimal places.)"

- "I would be intrigued, but not necessarily surprised. 100 years ago, scientific theories were completely different, so it's no surprise that ours would evolve too."

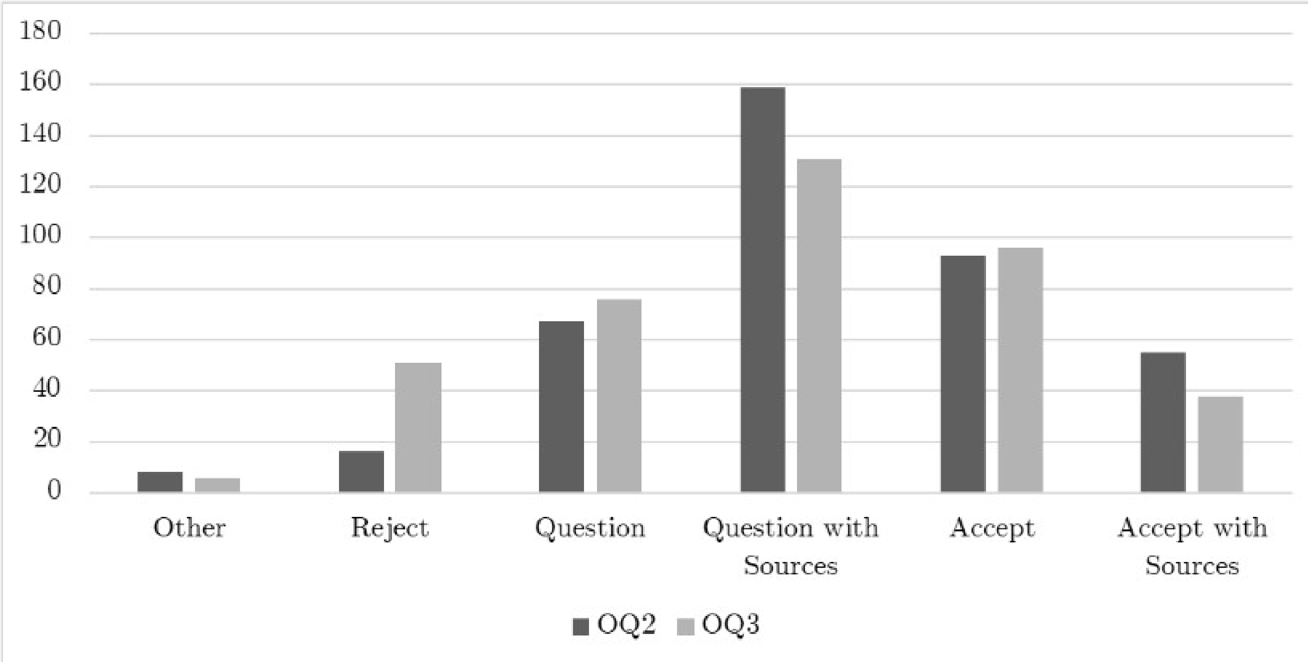

Figure 3.8: Comparison of Response Breakdown of OQ2 & OQ3

The third question portrayed how people react to changing science but there were several comments about how their beliefs would change. On average, people either accepted the science and made no changes to their beliefs or were slower to change their belief than accept the science.

- "Is it based on facts or opinions? If facts, I would probably still ignore them (like, if it was about eating meat being bad)."
- "I guess it depends on the belief since I try to be open minded, but I'm guessing I will express some internalized resistance to the idea."

However, changing beliefs was not the main focus of our study so any interesting responses pertaining to beliefs were noted but not considered important.

An interesting pattern that emerged was the presence of people who said they would not only reject the science, but their trust in the source would be damaged: "I would maybe not like the source as much anymore". This showed up in 9 responses, all rejecting the science. It appears that, for some people, the more personal the changed science, the more it harms trust in not only the science, but also the source reporting it. However, N was too small to extrapolate any robost claims or trends.

With the individual questions reported on, our focus shifted to the comparison. The main difference between OQ2 and OQ3 was the prevalence of the 'reject' response. OQ2 had 16 people who rejected the changing science for a variety of reasons. Some refused to read the paper or and some referred to a numbers game as an explanation. OQ3, on the other hand, had 51 people reject the change. When people are asked to consider their personal beliefs alongside the changing nature of science, over three times more people will reject the science.

In total, 140 answers did not shift over the course of the two questions but only six people rejected OQ2 and OQ3. The most frequent answer in this population was *QWS*. Figure 3.9 broke down the responses on OQ3 from the population who rejected OQ2. While *QWS* was a popular response, it was most likely that someone who rejected OQ2 rejected OQ3 as well.

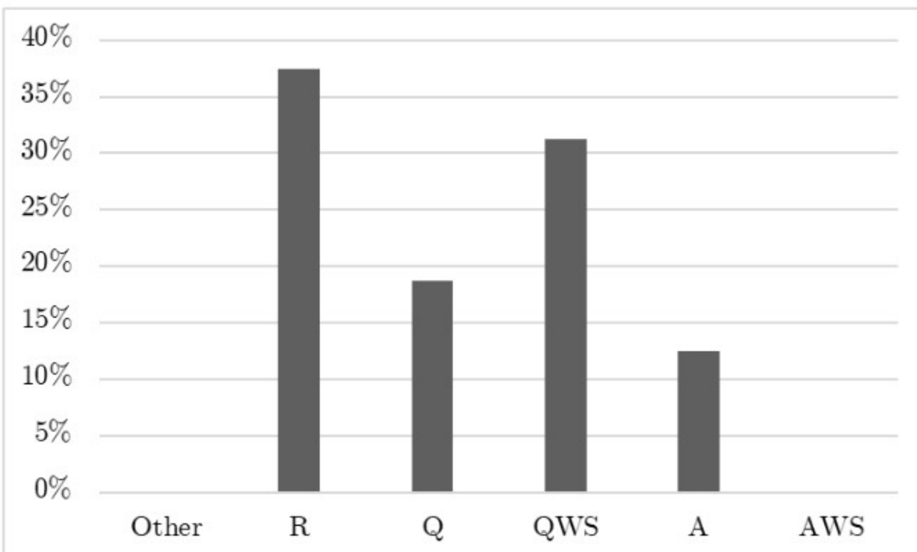

Figure 3.9: Breakdown of OQ3 Responses from OQ2 *Reject* Population

Open-Ended Question 4

The fourth open question was the only open question that had statistically significant p-values along our demographic slices. OQ4 asked participants to comment on how their educational experience impacts their understanding of the tentativeness of science and the breakdown is seen in Figure 3.10. Each participant was only coded once so the values are exclusive as there were only 3 respondents who could have been coded in either *tools* or *changing science* but each of them referenced the changing nature of science more. For example, "My education has given me the knowledge and understanding to take part in and understand these changing natures. Before entering academia, I would not be able to think critically about these things on my own." This person spoke to critical thinking but seemed to use it as a stepping stone to explain their understanding of changing science.

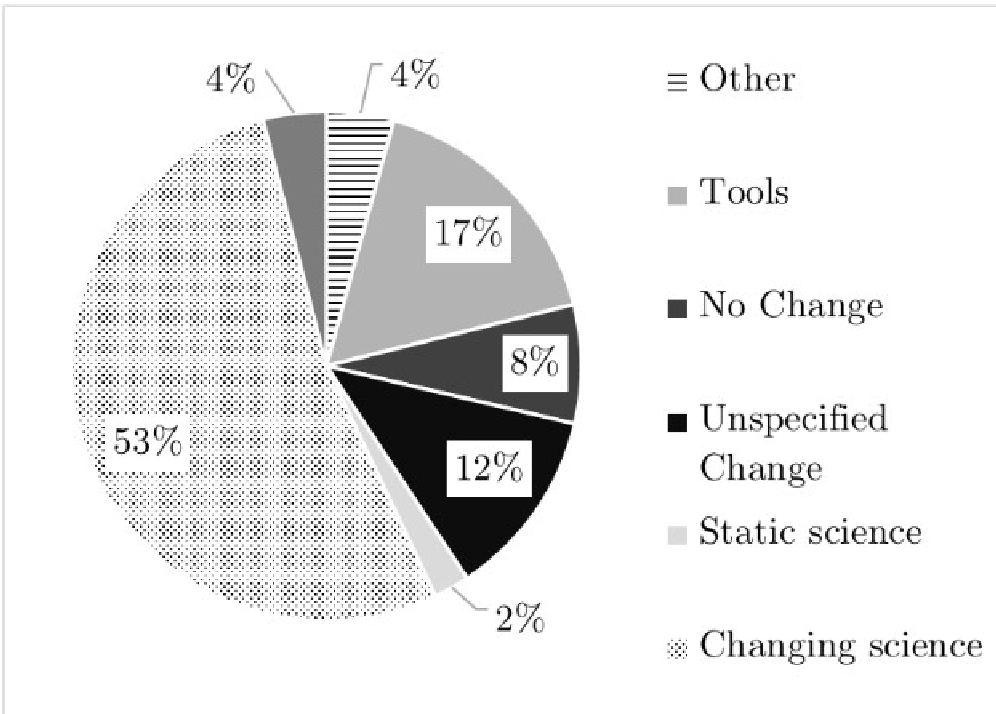

Figure 3.10: Breakdown of OQ4

The most popular result was that participants' educational background led them to believe that science changes but it was not nearly as uniform as OQ1. We were also interested in the intersection between OQ1 and OQ4 as the fourth open-ended question is the only question that addresses the educational background of participants but they asked similar questions. Figure 3.11 broke down the responses to OQ4 from different populations of OQ1.

Out of the 325 responses that we coded as change in OQ1, around half of them stated that

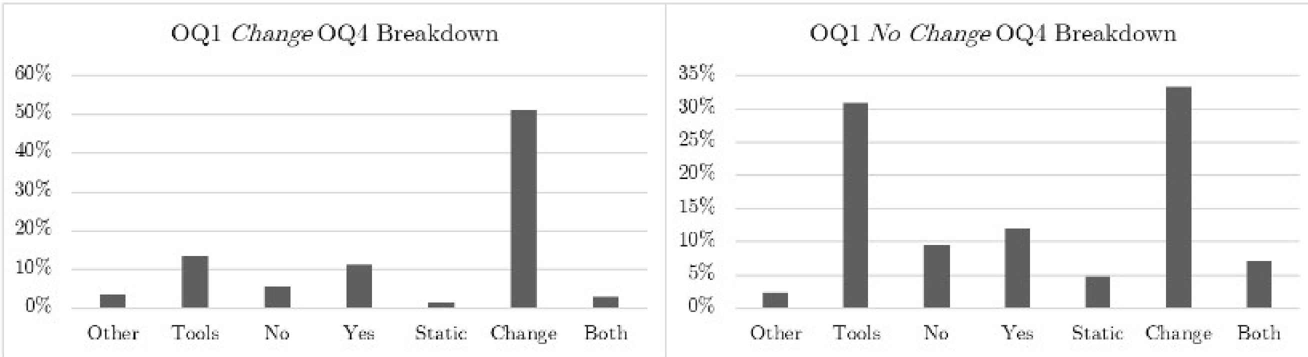

Figure 3.11: Breakdown of OQ4 Responses from Different OQ1 Populations

education played a role in their understanding of the changing nature of science. However, of the 42 responses that we coded as no change in OQ1, a third of them indicated that their education taught them about the tentativeness of science, which made it the most popular answer. We hypothesized that this inconsistency could be due to what is taught in classes or simply because people were inconsistent.

The prevalence of the heliocentric answer helped us understand the inconsistency. If individuals were reflecting on the tentativeness of science they might have been taught that science changed in the past such as the shift from geocentrism to heliocentrism or the discovery of atomic structure. However, they reflected on their life and did not notice any scientific changes.

OQ4 was the last question where political parties were a significant predictor for how people responded. Unlike the Likert-style items in the previous section, both p(Def) and p(All) were important. That means there were relationships between political parties and beliefs about educational impact (p(P(Def)) = 0.02 (Fisher's Exact/HBF) & p(P(All)) = 0.03 (Fisher's Exact/HBF)). OQ4's emergent categories focused on science as a changing or static concept and the tools they learned to better understand science.

Some trends seen in Figure 3.12 were that no Republicans indicated that education gave them tools for understanding science and Democrats were not as likely as other parties to report a lack of educational impact. To understand the trend surrounding the lack of Republicans mentioning tools, we compiled a list of different tools people mentioned:

- "Yes, I think that the more educated you become the better you are at critical thinking and to not believe in everything you see online, but to rather look for evidence to support your ideas." - Other (Socialist)

- "My experience with analyzing data has taught me to investigate holistically and look at all other related factors like sample sizes." - Democrat

- "I understand that scientific studies and numbers can be skewed or messed with to create misleading conclusions." - Democrat

- "I have more experience with reading scientific papers and interpreting them, as well as interpreting media headlines and articles (credible knowledge)." - Democrat

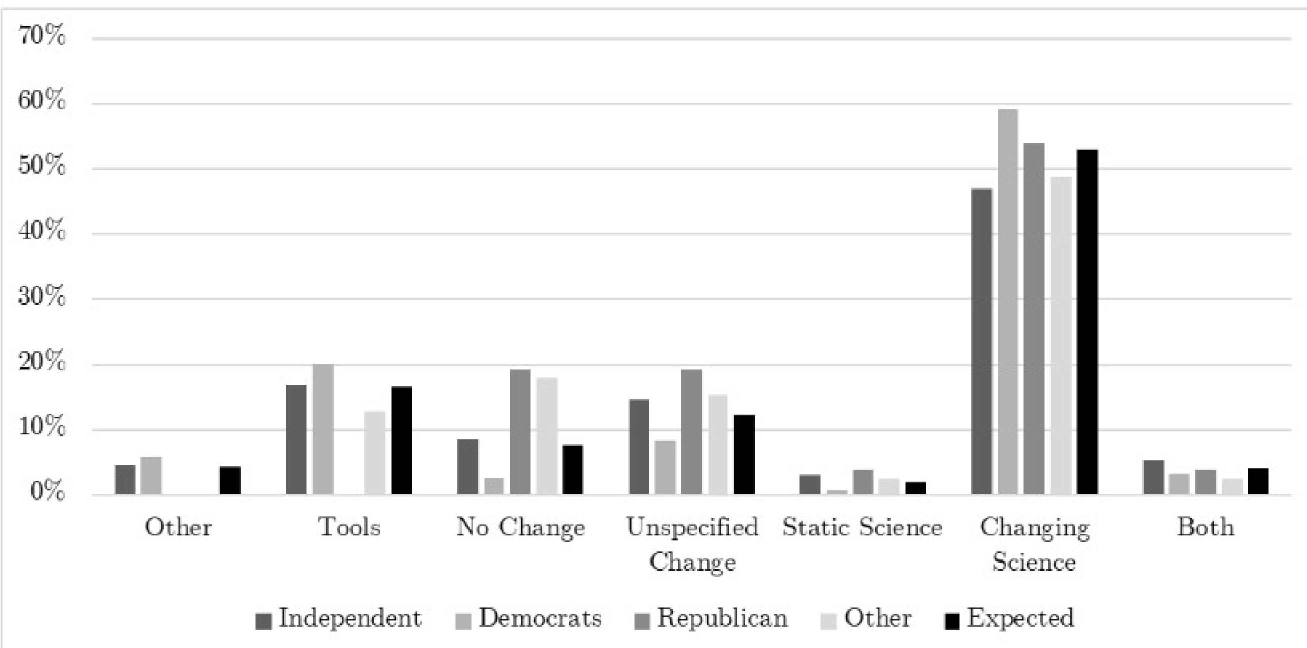

Figure 3.12: OQ4 Response Breakdown by P(All)

While there were tools that mention critical thinking or analyzing data, the most popular tool mentioned was the ability to discern reliable knowledge from false information. A potential reason behind this popular answer might be individuals feeling prompted by the previous open-ended questions which focused on new science being reported from a source.

3.4 Discussion

To summarize the political trends seen above, Republicans were more likely to feel neutral about the reinterpretation of science, Democrats were more likely to trust science, and Republicans were less likely to disagree that incorrect science provides credibility to personal ideas. Before the interpretations are discussed, we drew attention to a trend involving Independents.

In the I3 trends, Independents aligned themselves with Democrats while in Table 3.4 for I6, Independents appeared in sync with Republicans. In I10, Independents don't perfectly match with either party. Research from PEW indicates that, of the 38% of US adults that identify as Independent, 17% lean Democratic and 13% lean Republican [2] which was a mostly equal split. This trend and later data supports the PEW's findings as Independents are not always aligned with one party.

Another theme of the Independents' data not aligned with only one party was in OQ4. When we looked at OQ4's Figure 3.12, the black bar represented the expected value and, for almost every category, the expected value was relatively close to the Independent value. The expected value is not an average, we would need to use interval data to calculate averages and this data was ordinal. But, the expected value represents what the value would be if the variable of interest (in this case political party) had no influence. When the expected values and observed values were similar, like in Figure 3.2, there was no relationship between variables. So our potential interpretation was that people who are aligned with Democrats or Republicans were more swayed by their parties beliefs whereas Independents and people who chose *other* were not.

Another aspect of our findings is their agreement with previous studies. Both the Independents being truly Independent and Democrats reporting feelings of agreement more than other parties support different studies [2] [14]. While there might be other studies that explain the trends we saw in our survey, this agreement was compelling in supporting the validity of this study. We did not run any validity tests on this survey but since we saw results that other validated surveys found it lends strength that individuals were answering the way we expected them to.

Our main takeaway from these results which considered our leading questions was the lack of predictive power that educational experiences held. Our study focused on formal, public, and collegiate experience and found that educational backgrounds are not a statistically significant predictor of individuals' ability to reconcile the tentativeness and trustworthiness of science. This came as a shock. Moving forward, we need to be considerate of the emphasis placed on our current educational system as a single solution. While improving education might not change how people think about the tentativeness and trustworthiness of science, we still encourage advancing formal education. Maybe acknowledging the power polarized politics play in everyday life and creating resources to address that issue is a possible way forward. Or improving informal science education such as labs or field work will build higher levels of trust in science.

Chapter 4

Conclusion

4.1 Limitations

This section covered the limitations of our study. Our main limitations were positionality of the main researcher, the location of the main population, and the limits of surveying and statistics. Some of these limitations can be addressed in future and which we discussed in a later section. Other limitations were mentioned in the demographics section in the methods chapter but will still be included here for completion.

Positionality

We mentioned the identity of the main researcher in the methods section. They are a white, non-binary democrat who is not religious. As the research and interpretations in this study were broken down into different demographics and is potentially influenced by the researchers background and identity. They did strive to approach everything while recognizing their personal lens and make everything as objective as possible.

Data Population

To our understanding, around 99% of the participants are affiliated with the University of Colorado Boulder (CU Boulder) in some way, either as a student or professor. The demographic breakdown of CU Boulder from the Fall 2022 semester was as follows: 45.5%/54.5% women/men and 26.3%/73.7% people of color/white [21]. The percentage of registered voters in Boulder County is as follows: 43%/12%/45% democrats/republicans/independent [4]. This population breakdown can be seen in Figure 4.1 There is no data about religious affiliation or the non-binary population.

While the demographic breakdown of the study's population reflected these patterns well, it was not generalizable to the rest of the country. For example, the 2021 US Census reported 40.7%/59.3% people of color/white population breakdown[22].

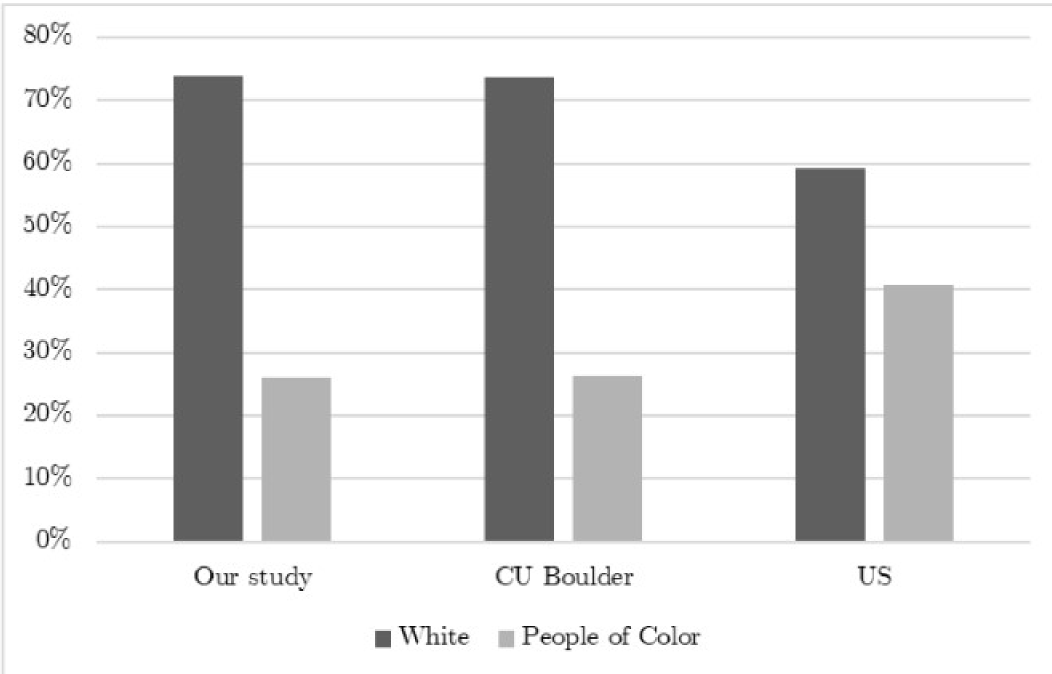

Figure 4.1: Breakdown of Race & Ethnicity Across Different Populations

Data Limitations

There are certain limitations that came with using a survey as our main instrument. Most of the limitations were from the quantitative items while the restrictions from qualitative were due to the survey environment. The survey format did not allow for follow up questions and it did not allow for deeper insight into a participant's thought process. We also set up our survey with an expectation of seeing educational background as a statistically significant relationship with items. So there are no items that are catered towards political parties or other social factors which limits the depth of our claims.

As for issues with quantitative research, relying on only numbers to accurately reflect patterns and trends limited our insight. The structured nature of likert-scale items restricted the participant's answers into one of five categories and left little room for complex answers. Our statistical hypothesis analysis testing is limited as it only reports the relationships and correlation between certain variables, not the reasons. It could not test causality between factors and if it could, the tests could not provide a reason for the relationship. So our interpretations, while backed up by the data from both qualitative and quantitative questions, are simply interpretations, not conclusive results. Due to time constraints we did not validate or test the survey's reliability. Validation and reliability ensures that the survey measures what we assume it measures and the data is replicable which establishes accurate results.

4.2 Summary

We created this summary to provide an overview of our research. It provided information about motivation and previous work surrounding our topic and recaps the methods used to create our survey and analyze the collected data. We also highlighted the key findings from this report.

The motivation behind this study was to explore the connections between education, the tentative nature of science, and trust. How was the tentative nature of science taught in a college-level classroom and did that have any influence on someone's trust in science? We found previous research that studied materials that teachers were given to teach the changing nature of science and instruments that measured the effectiveness of these curricula. We also found studies related to trust and general attitude towards science. The research indicated that Americans had a higher level of trust in science than other nations as well as the power that political parties play in personal confidence in science.

Our survey used a combination of Likert-style items and open-ended questions to investigate participants' understanding of the tentativeness of science and how it impacts their trust. A variety of questions were taken from previously validated instruments and we created the remainder with input from the Physics Education Group at CU Boulder. The survey was distributed in an email to CU Boulder instructors and we received 398 responses. The demographic breakdown represented Boulder's population but limited us in generalizing our findings any farther. We created rubrics to code the responses from open-ended questions using a mixture of a priori and emergent coding. For the Likert-style items, we used statistical hypothesis testing to find any correlations between specific demographics and items. After collapsing the data, we used Fisher's Exact and Holm-Bonferroni to find the statistically significant relationships.

We split the results section into Likert-style results and open-ended results. From Likert-style results, we recognized that nine of our questions had over a 50% agreement rate which limited the power of our claims and indicated that the statements are not predictors of relationships. We were surprised by the lack of correlation between the science and non-science demographic and our questions which is the first key finding. **Formal public college education in science is not a significant predictor of participants' ideas about the tentativeness and trustworthiness of science.** A potential reason why is all college classes treating the topic of changing science in the same manner.

The next key idea comes from an overall trend seen with our significant relationships. **A participants' political party is a significant predictor of their understanding of the changing nature of science and their trust in science.** Four Likert-style questions and one open-ended question correlated with either the difference between Independents and Democrats or Independents, Democrats, and Republicans. The first item, I3 asked about science changing due to reinterpretation of existing evidence and Republicans were more likely to feel neutral about the statement. I6 investigated the belief that science is reliable and trustworthy and Democrats were more likely to agree with the statement. The next item, I9, highlighted a difference between Democrats and Independents with Democrats more likely to disagree with the idea that education correlates change with failure. We found that the trend from I10 indicated Republicans were less likely to disagree with the concept that if science is incorrect, it provides credibility to personal ideas. Results from OQ4 indicated that Democrats were less likely to report that their education had no impact on them and no Republicans cited that education gave them tools to understand science.

The comparison between OQ2 and OQ3 which asked participants about their reaction to changing science and only changed if the science personally affected them or not led to another key finding. **People are more likely to reject changes to science if it affects a personal belief.** We also used horizontal descriptive statistical analysis to understand any themes between the open-ended questions. This led to a potential interpretation that individuals understood that science changes in the past but are unclear about the present tentativeness of science.

Our key takeaway from study was that our current education is insufficient. Our participants were aware of the tentativeness of science and most stated that they trusted science but, as our education system is currently set up, formal college education does not influence individuals ability to reconcile the tentativeness and trustworthiness of science. Building trust in science and understanding of the tentativeness of science might mean improving formal collegiate education but there might be other factors. Maybe improving reconciliation lies in informal education or K-12 education. Or it might be related to political polarization. But if we want people to better understand the tentativeness and trustworthiness of science, we can not rely on our current education.

4.3 Future Work

This section introduces proposals for future work in this vein of study. Most of it requires expanding the data population. We would like to extend the outreach of this survey to different groups across the US to incorporate differing thoughts and potentially generalize the results to a larger population. A larger data set would also allow us to use different methods to find correlations across variables, including a multi-variable correlation test which considers intersecting identities. Intersectional relationships might help us narrow down our interpretations and build stronger reasonings for correlations. A potential statistical hypothesis test we could use to understand relationships among many variables is the ANCOVA (analysis of covariance). However, to use the ANCOVA test, the required N value is in the thousands which is not achievable with our data set.

In future work, we would also like to conduct the survey with students in middle and high schools. One of our reasons for a lack of correlation between responses and participants' background in science was the idea that the tentativeness of science is not treated differently in college level classes. It would be interesting to see if this trend continues in middle and high schools and educational experiences do not make any difference in someone's understanding of the tentativeness and trustworthiness of science. Or if there is a certain age where science education has a big impact but diminishes after a specific year. Or if science classes taken in high school have any impact. And younger students' might not have had time to join a political party. Removing the main predictor from our data set would allow us to investigate other intersections. While formal educational college experiences were not significant in this study, there are lots of potential research areas to explore in other facets of education.

We would also like to measure different aspects of an individual's identity. The WGM survey cited that socioeconomic status played a role in someone's trust in science worldwide so there is potential for an investigation into that predictor in our population. We could also explore the difference between political ideology (liberal/conservatism), registered political parties (democrat/republican), and history of voting. Along those lines, investigating the evolution of identity alongside the tentativeness and trustworthiness of science could provide intriguing results.

It might be beneficial to validate and test the reliability of the survey. As mentioned in the limitations above, a reliable and valid survey returns accurate results. Testing the reliability of the survey could mean testing and retesting or internal consistency. Retesting requires the same participants to repeat the survey after a period of time and internal consistency asks if the survey returns the same answer from different parts that are designed to measure the same thing. The most valuable type of validity for this survey would be internal validity which measures the extent that we can be confident that any relationships found in our survey are not from external factors.

Bibliography

[1] Gerald Albaum. The Likert Scale Revisited. Market Research Society. Journal., 39(2):1–21, March 1997. Publisher: SAGE Publications.

[2] Sara Atske. Political Independents: Who They Are, What They Think, March 2019.

[3] Mesfin Bekalu, Georges Benjamin, Cary Funk, Melanie Green, Lewenstein Bruce, Yeo Sara, U.S. Census Bureau, NSF, and Office of Science and Technology Policy. Science and Technology: Public Perceptions, Awareness, and Information Sources | NSF - National Science Foundation, May 2022.

[4] Boulder County. 2022 Post-Election Data, 2022.

[5] Rodger Bybee. THE NATURE OF SCIENCE IN. 2013.

[6] Jennifer Childress. A Framework for K–12 Science Education | Next Generation Science Standards, July 2011.

[7] Jennifer Childress. APPENDIX H – Understanding the Scientific Enterprise: The Nature of Science in the Next Generation Science Standards, April 2013.

[8] Michael Dunn. Chapter 3: Trust in science and health professionals | Wellcome Global Monitor 2018, 2018.

[9] Michael Dunn. How Covid-19 has increased the world's trust in science, November 2021.

[10] Michael Freeman. Transformation of the Colorado Classical Mechanics and Math Methods Instrument to the Coupled Multiple Response Format. Undergrad Honors, University of Colorado Boulder, Colorado, 2022.

[11] Cary Funk, Meg Hefferon, Brian Kennedy, and Courtney Johnson. Trust and Mistrust in Americans' Views of Scientific Experts, August 2019.

[12] Abigail Geiger. Political Polarization in the American Public, June 2014.

[13] Friederike Hendriks, Dorothe Kienhues, and Rainer Bromme. (PDF) Trust in Science and the Science of Trust, February 2016.

[14] Julia Koblitz. Trust in science is becoming more polarized, survey finds | University of Chicago News, January 2022.

[15] John Laloggia. 6 facts about U.S. political independents.

[16] Lawrence S. Lerner and William J. Bennetta. The Treatment of Theory in Textbooks. The Science Teacher, 55(4):37–41, 1988. Publisher: National Science Teachers Association.

[17] Ling L Liang, Sufen Chen, Chen Xian, and Nafiz Kaya Osman. Student Understanding of Science and Scientific Inquiry (SUSSI): Revision and Further Validation of an Assessment Instrument. Semantic Scholar, 2006.

[18] Michael S. Matell and Jacob Jacoby. Is There an Optimal Number of Alternatives for Likert Scale Items? Study I: Reliability and Validity. Educational and Psychological Measurement, 31(3):657–674, October 1971. Publisher: SAGE Publications Inc.

[19] E McColl, A Jacoby, L Thomas, J Soutter, C Bamford, N Steen, R Thomas, E Harvey, A Garratt, and J Bond. Design and use of questionnaires: a review of best practice applicable to surveys of health service staff and patients. Health Technology Assessment, 5(31), 2001.

[20] William Mccomas and Noushin Nouri. The Nature of Science and the Next Generation Science Standards: Analysis and Critique. Journal of Science Teacher Education, 27, May 2016.

[21] Office of Enrollement. Overall Enrollment Profile, 2022.

[22] USA Facts. US population by year, race, age, ethnicity, & more, March 2023.

[23] Bethany R. Wilcox and H. J. Lewandowski. Students' epistemologies about experimental physics: Validating the Colorado Learning Attitudes about Science Survey for experimental physics. Physical Review Physics Education Research, 12(1):010123, March 2016.

[24] Cody Tyler Williams and David Wÿss Rudge. Effects of Historical Story Telling on Student Understanding of Nature of Science. Science & Education, 28(9-10):1105–1133, December 2019.

Appendix A

All Data Breakdown

		Science		Race		Gender		Religion		Political Identity		
		Science	Non-science	White	Not white	Male	Female	Religious	Not religious	Independent	Democratic	Republican
I1	1	5	6	8	2	4	6	2	7	5	3	1
	0	6	5	8	2	6	5	4	7	1	5	3
	-1	283	93	265	95	171	176	74	285	124	146	22
I2	1	10	6	12	3	6	9	5	10	6	2	2
	0	7	3	3	4	3	4	2	8	2	2	1
	-1	277	95	266	92	172	174	73	281	122	150	23
I3	1	12	4	7	7	3	11	5	9	6	5	0
	0	14	10	18	3	12	11	10	14	6	4	6
	-1	268	90	256	89	166	165	65	276	118	145	20
I4	1	4	4	2	5	3	3	2	5	4	1	1
	0	5	1	1	3	1	5	0	6	2	1	0
	-1	285	99	278	91	177	179	78	288	124	152	25
I5	1	65	19	55	26	35	44	23	55	29	29	5
	0	35	17	32	13	23	24	11	40	14	14	6
	-1	194	68	194	60	123	119	46	204	87	111	15
I6	1	4	7	5	2	3	5	3	5	4	2	1
	0	19	8	15	7	11	12	5	22	15	3	2
	-1	271	89	261	90	167	170	72	272	111	149	23
I7	1	2	5	4	2	5	1	3	2	4	0	1
	0	24	9	19	9	16	15	10	22	10	9	5
	-1	268	90	258	88	160	171	67	275	116	145	20
I8	1	1	2	2	0	1	1	1	1	2	0	0
	0	2	1	1	2	1	2	0	3	2	0	0
	-1	291	101	278	97	179	184	79	295	126	154	26
I9	1	103	40	98	39	60	76	29	104	41	67	6
	0	48	24	50	20	33	31	15	56	26	21	7
	-1	143	40	133	40	88	80	36	139	63	66	13
I10	1	91	33	90	32	66	51	22	100	35	61	2
	0	87	33	80	35	40	68	20	92	38	43	9
	-1	116	38	111	32	75	68	38	107	57	50	15

Table A.1: Likert-Style Data

		Science		Race		Gender		Religion		Political Identity		
		Science	Non-science	White	Not white	Male	Female	Religious	Not religious	Independent	Democratic	Republican
C1	1	14	9	15	5	11	9	8	15	5	8	2
	0	32	13	26	16	16	24	14	31	13	11	6
	-1	237	71	224	74	144	143	52	239	106	127	14
C2	1	2	1	1	1	1	1	0	2	1	1	0
	0	1	3	4	0	3	1	2	2	3	1	0
	-1	2	2	3	1	0	4	0	3	1	1	1
OQ1	O	10	7	14	2	10	7	2	15	9	4	0
	NC	29	13	32	8	20	19	11	30	19	10	5
	C	243	82	225	85	145	153	65	242	99	134	21
	RC	12	2	10	4	6	8	2	12	3	6	0
OQ2	O	5	1	7	1	4	4	2	6	2	2	1
	R	37	14	9	5	5	11	4	12	17	15	4
	Q	60	16	43	23	36	26	18	45	18	30	6
	QWS	98	33	120	33	71	78	28	125	45	56	7
	A	65	32	64	23	39	44	22	68	36	31	6
	AWS	29	8	38	14	26	24	6	43	12	20	2
OQ3	O	5	1	5	1	4	2	0	6	3	2	1
	R	37	14	35	15	15	35	10	41	5	5	1
	Q	60	16	54	16	35	35	16	56	25	23	6
	QWS	98	33	94	35	60	59	24	99	50	68	8
	A	65	32	62	28	49	41	25	69	30	33	8
	AWS	29	8	31	4	18	15	5	28	17	23	2
OQ4	O	13	4	8	8	8	8	5	12	6	9	0
	T	47	20	47	15	24	36	10	56	22	31	0
	NC	19	11	18	8	17	11	10	20	11	4	5
	UC	36	12	33	14	24	21	18	30	19	13	5
	S	6	3	6	2	5	3	1	8	4	1	1
	CS	161	50	159	47	97	100	33	160	61	91	14
	B	12	4	10	5	6	8	3	13	7	5	1

Table A.2: Conditional and Open-Ended Data

Appendix B

Email to Professors

I am currently a senior in physics working on my honors thesis with the physics education research group, mainly Dr. Bethany Wilcox, at CU Boulder. I am conducting research about how people understand the nature of science from the Next Generation Science Standards and how this intersects with people's trust in science. One of the facets of nature of science is that science is subject to revision and change. I am using this survey to measure not only if people believe that science is subject to change but also if people trust science more or less because it changes. Feel free to take the survey yourself if you are curious about the questions it contains.

The target population of the survey are mostly college aged students in Colorado. The survey collects no identifiable information and offers a link to a google survey for their emails if they would like to be entered for the gift card raffle. That ensures that all answers are anonymous. The data itself will be analyzed to find potential patterns and themes but the raw data will not be shared with anyone else.

It would be great if you distribute the following survey to students within your class. It is an online qualtrics questionnaire that should take 5 - 10 minutes. It would be appreciated if you could publish it in canvas announcements and mention it to your students during class. The bit below is meant to act as a copy and paste announcement for you but feel free to make any necessary changes to better fit your class.

Recent world events, including the COVID-19 pandemic have emphasized the importance of understanding how people understand the nature of science and trustworthiness of scientific claims. As part of an effort to better understand students' ideas around this, researchers at CU have developed an anonymous survey to solicit ideas. This survey will close mid Feb 2023 and students who complete the survey will have the option to enter into a drawing for one of five 20 dollar electronic gift cards. Due to the anonymity of the survey, instructors will not be informed about participation. Link is below.

Thank you for your time and consideration. Feel free to contact me if you have any questions or concerns.

Appendix C

Qualitative Rubrics

NR (no response)	NC (no change)	C (change)	R (reinterpreted)
There is no response; they state that they do not know; the response does not address the prompt; OR the response cannot be classified based on the rubric descriptions.	Scientific theories do not change over time if they are based on accurate experiments or facts. OR The response includes misconceptions concerning the nature of science or self-contradicting statements.	Scientific theories may be changed when experimental techniques improve, or new evidence is produced.	Scientific theories may also be changed when existing evidence is reinterpreted.
Examples	I think well accepted theories do not change because before they are accepted, they are tested and peer reviewed so that they are factual and there are no holes in the theory.	I think that they update as new discoveries are made and it prompts reexamination of the current rules.	They change due to new discoveries and reviewing prior data to make determinations.

Figure C.1: OQ1 Rubric

O (other)	R (reject)	Q (question)	QWS (question with sources)	A (accept)	AWS (accept with sources)
There is no response, response does not address the prompt, OR response can not be classified	Rejects the idea of changing science. (due to any reason)	Questions the idea of changing science. (Reading the paper/article does not count as looking into other sources)	Questions the idea of changing sources AND mentions looking into other sources for more info	Accepts the idea of changing science	Accepts the idea of changing science AND mentions looking into other sources for more info
OQ2 Example	I would believe that it is incorrect and be throughly suprised!	I approach it with caution, but if there is an explanation I'm interested in hearing it.	I would react by doing some research of my own and gathering information from other sources to inform myself in order to understand why that source made that assertion.	I wouldn't be surprised as it is only a theory and with new evidence and technology, it is very possible to disprove them.	I will consider the theory and the findings that support it before I agree with it.
OQ3 Example	I would not believe them as much if it is something I believe in.	Ask why. There has to be a reasoning for that core belief being scientifically incorrect and evidence that backs up that reasoning.	I would ask them about where they were receiving this info from and then research that myself.	I would accept and move on.	Again, if I could find multiple reputable sources that agreed with those statements, I would consider changing my opinion or beliefs on what I previously thought was true.

Figure C.2: OQ2 & OQ3 Rubric

O (other)	NC (no change)	UC (unspecified change)	T (tools)	SS (static science)	CS (changing science)	B (both)
There is no response, response does not address the prompt, OR response can not be classified	States that educational experiences had no effect on their understanding	States that educational experiences had an affect, but it does not fit into the nature or science OR is unspecified	States that educational experiences have provided them with tools for understanding science	States that educational experiences taught them science does not change	States that educational experiences taught them science changes	States that educational experiences taught them both that science does not change and changes
Examples	Not at all, I've had a relatively constant understanding of how I should think about things and come to my beliefs	I think my educational experience is the only thing that has impacted by view of the changing/static nature of survival.	My education has allowed me to thinking more critically about claims and the changing of science.	Through my education, I've learned a lot more about the static nature if science than changing because a lot of the ore elements of science have seemed to remain the same throughout my experience.	My educational experience has definitely shown me that science is always changing and evolving as we find better research methods, new problems, and new solutions.	My educational experience has made me think that some equations and theories are basically facts and other equations and theories are circumstantial.

Figure C.3: OQ4 Rubric

Appendix D

Qualtrics Survey

Intro and consent

Permission to Take Part in a Human Research Study:

Title of research study: Investigating beliefs about the nature of science
IRB Protocol Number: 22-0500
Investigator: Bethany Wilcox

Purpose of the study
The purpose of this study is to investigate how individuals think of and reason about the nature of science. As emphasized by the recent messaging around the COVID-19 pandemic, how people interpret and make sense of science and scientific claims has risen to a topic of particular interest. You are being invited to participate in an anonymous survey in which you will be asked to respond to questions asking you about your ideas about the nature of science. Participation is entirely voluntary, and we expect it to

- I consent
- I do not consent

Thank you for your time. You may close the survey window.

Please confirm that you are 18 years of age or older.

- Yes, I am 18 years of age or older
- No, I am not 18 years of age

Open Questions

With examples, explain why you think scientific theories do not change OR how (in what way) scientific theories change:

If a reputable scientific source stated that a well-known theory (i.e. general relativity or big bang theory) is incorrect, how would you react to the finding?

If a source you trust stated that a personal core belief (e.g., eating meat is bad for the planet, aliens exist) is scientifically incorrect, how would you react to the finding?

How has your educational experience impacted your opinion regarding the changing/static nature of science?

Likert Scale - NOS

For the follow questions, please rate your level of agreement.

	Strongly Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
Science is subject to change	○	○	○	○	○

	Strongly Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
It is **possible** that scientific theories may be completely replaced by new theories in light of new evidence	○	○	○	○	○
It is **possible** scientific theories may be changed because scientists reinterpret existing observations	○	○	○	○	○
It is **possible** scientific theories are subject to ongoing testing and revision	○	○	○	○	○
It is **possible** scientific theories based on accurate experimentation will not be changed	○	○	○	○	○

If you would like to expand on any questions above, please feel free to do so here.

Likert Scale - Trust

For the follow questions, please rate your level of agreement.

	Strongly Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
Science is reliable and trustworthy	○	○	○	○	○
Science, in general, makes life better for the average person	○	○	○	○	○
Science is an important method for discovering our natural world	○	○	○	○	○
Education enforces that being incorrect	○	○	○	○	○

	Strongly Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
and asked to change is a sign of failure					
The fact that science is sometimes incorrect provides credibility to personal ideas	O	O	O	O	O

If you would like to expand on any questions above, please feel free to do so here.

Conditional.1

For the follow question, please rate your level of agreement.

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
Science is more trustworthy **because** it changes.	○	○	○	○	○

If you would like to expand on any questions above, please feel free to do so here.

Conditional.2

For the follow question, please rate your level of agreement.

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
Science is more trustworthy **because** it does not change.	○	○	○	○	○

If you would like to expand on any questions above, please feel free to do so here.

Demographics

Demographic Information

We are trying to reach out to a broad range of participants to ensure collection of diverse perspectives. If you are comfortable, please share your information below.

With which of the following do you identify?

- ○ Man
- ○ Woman
- ○ Non-binary/Gender non-conforming
- ○ Not listed: ______
- ○ Prefer not to answer

With which of the following do you identify? (Check all that apply)

- ☐ Asian
- ☐ Black/African American
- ☐ Caucasian
- ☐ Hispanic/Latinx
- ☐ Native American
- ☐ Pacific Islander
- ☐ Not listed: ______
- ☐ Prefer not to answer

With which of the following do you identify?

- ○ Independent
- ○ Democratic
- ○ Republican

- Other ____________
- Prefer not to answer

If you consider yourself a member of a religion or religious organization, please let us know below.

____

Have you ever recieved a degree in a scientific field?

- Yes
- No

Are you currently working on a science degree?

- Yes
- No
- Other ____________

Block 7

What is the highest level degree you completed?

- Associate degree
- Bachelors degree
- Masters degree
- Doctorate
- Other ______

Block 8

What is the highest level degree you completed?

- High School Graduate
- GED or equivalent
- Some college, no degree
- Associate degree
- Bachelors degree
- Masters degree
- Doctorate
- Other ______