## 05.26.

* More technical discussion
* Define metrics
* Link recommendations to what data we have or what data we would do to test
* Inculde 2 paragraphs about the actual treatment, include link to the treatment document
  + Show all the work we did
* Discuss who did what

## 05.23.

* Tactic results addition graph
  + Include in presentation
  + Run regression with that outcome with all the controls to reduce variation and get a more precise estimate
* Presentation in general
  + Focus on main results
  + Include other stuff in the appendix
  + Include results with controls included (key demographics, randomizations)
  + Additional versions with fewer/ other subsets of controls need to go into the appendix

## 05.19.

* Tactic results additions
  + Look at the true questions (Ross)
    - 1 if true and no tactic is selected (this is not manipulative selected)
    - 1 if misleading and at least one of 3 covered tactic is selected
    - 0 otherwise
    - Add them up for all 6 questions
    - → this will be one score
    - → look at the delta post-pre
  + Difference in selecting fake tactic → should go down, or at least not up (Ross)
  + Difference in number of tactics selected by treatment vs control (Ross)
    - Total number across all questions
    - # of times each tactic is selected across 6 questions
* Reduce variance in general (KG)
  + Control delta ~ covariates → pick covariates
  + Reduction for the sharing Q
* Less rigorous (KG)
  + Control for all three attention score
  + All attention checks passed in HTE
* Change results with new multiple hypothesis correction (KG)

## 05.11.

* Why are people sharing? How can we elicit what’s the gut reaction (initial urge to share)?
* How is the treatment affecting the initellectual vs emotional part of the brain (relate to the sharing Q)
* What we could have done differently?
* Hypothesis correction under FDR for motivating future research

## 04.27.

* Order of demographic and post (or very end?)
  + Start with demographic and take out what we get from the pre test (easier to defend and shorter)
* Email during their business hours. Don’t be too pushy. Not too many questions at the same time or otherwise try scheduling an appointment. Polite and not too much burden.
* Write up the power calculations, assumption for pre and post
  + How much reduced variance for control?
    - Pre post design literature or make re
    - Randomly assign control questions to pre and post
* Don’t ask too many questions during presentation, explain what we are doing and how we got to it
* TODO:
  + Finalize pre/post questions 6 questions:
    - 2 True
    - 3 False (1 of each tactic)
    - 1 Combined (2 tactics mislead, graph that doesn’t mislead)
  + Edit demographic section based on comments and discussion
  + Do correct power calculations based on what we discussed
  + Write up analysis plan nice and clear
  + Ask Tom:
    - Power calculation approach and assumptions
    - ..
  + Discuss later:
    - Which 2 outcome questions
    - Exact design of outcome questions
    - ..

## 04.25.

* How do we want to measure - identifying misinfo
* Covariates - cut down on the correlated ones
* Tasks
  + Finalize tactic messages (Réka)
  + Control group course - literature review (Ben)
  + Outcomes survey test finalize & streamline (Quentin)
  + Power and motivation smaller control (Ross)
  + HTE & ML explanation (Kush)
  + Slides (slightly later)
  + Qualtrics (later)
  + Meeting with Susan (Tuesday)(Réka to email her - DONE)

## Experiment Design Discussion with Susan

* + Goal: understanding whether our tactics course helps participants identify misleading(manipulative) posts + reduce sharing/engagement behavior
  + 3 options we are considering:
    - **2 arms: 1 control, 1 treatment course**
      * **Advantages:**
        + **Higher power for HTE analysis since more treatment units**
        + **Partner hypothesizes small treatment effects for any possible way of spending the 5 minutes of intervention → we should do our best effort to have an effect: splitting the sample or weakening the treatment message might not be the best**
      * **Disadvantages:**
        + **Cannot disentangle more specific causes such as which specific course or delivery caused improvement. (however, we do have multiple outcomes of interest that could be split up for separate analysis - ability to identify specific tactics; overall ability to identify manipulative content; overall behavior/engagement with misinfo)**
    - 3 arms: 1 control, 1 treatment course (text focused), 1 treatment course (interaction/delivery enhanced)
      * Advantages:
        + Can answer 2nd question of delivery method effect size
      * Disadvantages
        + Based on our lit reviews and conversation with the partner we have a strong belief in that the interactive method is the most effective one and the one that they would want to use.
        + Lose power on HTE
    - 4 arms: 1 control, 3 treatment courses (each tactic) (maybe do a combined one still and 5 groups total? e.g. 40% combined (in case effects are small), 15% each into each single tactic treatment (in case there happen to be big effects), 15% control)
      * Advantages
        + Can disentangle specific tactics that were effective
      * Disadvantages
        + Lose a lot of power (likely not enough participants)

For treatment estimation

For HTE

* + - * + (Might not be the most effective use of the sample - the tactics are relatively different so combining the results with prior knowledge it can be possible to provide suggestive evidence for each of the tactics even if we only have the combined treatment e.g. if the graph recognition improved then it’s likely from the graph treatment and not from the anecdotes, even if there is just combined treatment - this of course relies on prior knowledge and is not as strong as our precide treatment estimates overall))
    - Modified version of the above: 4 groups randomly picking with 2 of the 3 tactics to include
      * Advantages:
        + Stronger treatment than just 1 message and still able to identify individual effects of the treatments (?)
      * Disadvantages:
        + Same as above, although treatment weakness might be less of a concern relative to just one of the three messages?
  + Control group
    - Does our logic for having a smaller control group make sense?
      * We’ll have a pre-post difference so we expect very small variance for the control groups → reasonable to have a smaller group size ?
    - Ideas for how to make control group task neutral since we have pre-post test design

## 

## Summary of 04.21. Meeting

* Outcomes
  + Good to focus on teaching ability to **recognize tactics**
  + **Link it to specific behaviors**: sharing intent, talk about, google-ing, action/buying
  + Want it to be **scalable, relevant to anybody** - health topics better than over polarized covid or partisan misinfo stuff
    - Ok to use their examples, it is hard to come up with stuff (though we may use other examples too)
    - We can ask her for more examples on bad visuals
* Treatments
  + Ok to focus on ~3 specific tactics or elements
  + Ok to **focus on specific elements** and do our best in **delivering effectively**
  + Focus on **less tested, less obvious** tactic elements
  + Reference/examples:
    - E.g. there are [26 different ways](https://twitter.com/johnfocook/status/1335628473533083649) of manipulating, we will show you 3 less discussed ones
    - Other inspirations: [cranky uncle](https://crankyuncle.com/game/), <https://www.goviralgame.com/> (emoitions/fear, fake experts, conspiracy/puppets; how misleading do you find headlines, overlap in pre-post)
  + Ideas
    - Fake experts
      * Easy to test for/have good outcome questions
    - **Graphs**
      * Less actionable suggestions in the graphs
      * Could work well with better infographics
      * <https://venngage.com/blog/misleading-graphs/>
      * <https://www.youtube.com/watch?v=E91bGT9BjYk>
    - New ideas
      * Ad hominems (attack person)
      * **Anecdotes / cherry picking**
      * False choices
      * Slippery slopes
      * Strawmen (weak version)
      * *Logical fallacies - would need to be specific (e.g.* ***false comparison)***
      * ***Blowfish/red herring***
    - <https://crankyuncle.com/a-history-of-flicc-the-5-techniques-of-science-denial/>
* Delivery
  + Didactive is frustrating
  + Find ways to communicate that others are doing this too (social element/connection, you’re not alone)
  + **Interactive delivery sounds good**, we can do what we think will work best/be most effective
  + Phrase content so that we can reach everyone: **talk about manipulation** instead of misinformation or fake news as these two have been polarized and weaponized
* Subgroups
  + Hypothesized important ones
    - Partisanship
    - Education level
    - Health focus, vaccinated, wears masks
    - Self-identified misinfo susceptibility
  + Not strong hypothesis on stuff working differently for different people
  + Goal: get as much info as we reasonably can and **do post HTE analysis**
* Effect size:
  + Hard to push people
  + **Expecting small effects**, any effect would be helpful *(→ one or two groups max??, do everything we can to increase power, incl. pre and post ??)*
  + Anything that can make people be less susceptible to manipulative communication in a scalable, widely relatable way would be very helpful
* Outstanding tasks
  + Pick tactics to focus on with corresponding messages and outcomes [Quentin, Ben, Reka(bit)]
    - How to do different treatment groups? (power?)
    - E.g. pick out the three from the 26, narrow down and be focused
    - From original course or something else? She wanted new stuff so maybe 2 new and 1 from the pilot (e.g. misleading graphs, emotions)
    - Need the specific outcome headlines - representative of tactics
  + Decide on additional outcome questions (behaviors?) (how many? power?) [Quentin, Ben, Reka(bit)]
    - Which outcomes do we want?
    - Which outcomes / behaviors (actions) do we want to affect?
    - How to make the outcomes less noisy / how exactly are we measuring it
  + Specific ways to deliver interactively [Reka]
    - Keep in mind the timing/length
    - Not spend too long on this
  + What analysis do we want to do? [Kush, Ross]
    - What hypotheses are we testing?
    - List of covariates for HTE and stratification
      * Pre-analysis plan: how many HTE groups are reasonable (power?)
      * Exact list of HTEs
      * Ways of handling the outcomes - e.g. which ones are we grouping and which ones are separate
  + Anything else?

## Motivation and hypotheses

**Research Questions** (from packet)

* Which components of the course content are most effective at making respondents better at identifying misinformation and less willing to share misinformation posts on social media?
* What pedagogical approaches work best? (explore if giving people information is more or less effective compared to various forms of quizzes designed for engagement)
* Do different interventions work better for different subgroups? How can we discover the different subgroups that may respond heterogeneously to our treatments?
* **DELIVERABLE:** The team will submit a pre-analysis plan including a summary of the precise research questions, treatments, and hypotheses

What hypotheses are we testing?

* Specific content elements?
* Specific pedagogical methods?
* What outcome do we care about? [“better at identifying misinformation and less willing to share misinformation posts on social media” ]
* Do different interventions work better for different subgroups?

→ Review relevant literature

→ Analyze pilot data to inform the above

→ Talk to partner

### Literature review

#### Methodology review

* Hutchins, S. S., Brown, C., Mayberry, R., & Sollecito, W. (2015). Value of a small control group for estimating intervention effectiveness: results from simulations of immunization effectiveness studies. *Journal of comparative effectiveness research*, *4*(3), 227-238.
  + <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4905819/>
    - Simulation study to examine the tradeoff of having a smaller control group compared to intervention group
    - In their simulations, the mean intervention effect and effect sizes were equivalent regardless of control group size
    - “These results…indicate that studies with smaller control groups can generated valid and accurate evidence.”
* Wason, J., Magirr, D., Law, M., & Jaki, T. (2016). Some recommendations for multi-arm multi-stage trials. *Statistical methods in medical research*, *25*(2), 716-727.
  + <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4843088/>
    - Examines multi-arm, multi-stage trials
    - “In a traditional RCT in which the endpoint measured for both the control and experimental treatments have the same variance, the optimal allocation between arms, in terms of maximising the power, is 1:1
    - However, when there are multiple experimental arms all being compared against the control arm, the optimal allocation is no longer 1:1.
    - The optimal allocation to the control group has been shown to be approximately √*𝐾* patients allocated to the control group for everyone one patient allocated to a given experimental treatment”
* Wason, J., Stecher, L., & Mander, A. P. (2014). Correcting for multiple-testing in multi-arm trials: is it necessary and is it done?. *Trials*, *15*(1), 1-7.
  + <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4177585/>
    - “A sizeable proportion of published multi-arm trials do not correct for multiple-testing. Clearer guidance about whether multiple-testing correction is needed for multi-arm trials that test separate treatments against a common control group is required.”
* <https://blogs.worldbank.org/impactevaluations/when-should-you-assign-more-units-study-arm?CID=WBW_AL_BlogNotification_EN_EXT>
  + “When the outcome is continuous and you’re after powering a test for the difference of two means, this is straightforward: you want to allocate treatments in proportion to the standard deviations in the two groups.”

#### Control Treatment Literature Review

* <https://www.tandfonline.com/doi/abs/10.1080/00220973.1990.10806551>
  + “Studies were clustered into control procedures that included explicit manipulation of attentional elements (Attn + ), explicit description of the placebo activity plus implied reference to attention (Act+ ), and exclusive reference to the activity (Act-Only)... Compared with the other two procedures, Act+ controls were employed more often in interpersonal rather than academic studies in which subjects knew they were participating in an experiment, had an expectancy for performance change, and the placebo activity resembled the treatment task.”
* <https://journals.sagepub.com/doi/pdf/10.1177/1745691613491271>
  + “Active control groups are superior to “no-contact” controls, only when the active control group has the same expectation of improvement as the experimental group can we attribute differential improvements to the potency of the treatment.”
* <https://library.oapen.org/bitstream/handle/20.500.12657/46921/9781000179033.pdf?sequence=1#page=160>
  + Control group asked to solve a “neutral word puzzle”
* <https://www.researchgate.net/profile/Sander-Van-Der-Linden/publication/349494557_Countering_Misinformation_and_Fake_News_Through_Inoculation_and_Prebunking/links/60338c744585158939bf0ad8/Countering-Misinformation-and-Fake-News-Through-Inoculation-and-Prebunking.pdf>
  + Participants engage in content about freezer burn
* Macula Degeneration

#### Innoculation literature review

*Example:*

* *Citation*
  + *Link*
  + *Key points:*
    - *Points*
    - *Points*
  + *Relevant for us:*
    - *Promising or ineffective content elements?*
    - *Promising or ineffective delivery methods?*
    - *Promising heterogeneities?*
    - *Promising outcomes and/or way of measuring them?*
    - *Anything else?*

*List of papers reviewed (put citation here to check it’s not done already*

1. *Roozenbeek, J., van der Linden, S. Fake news game confers psychological resistance against online misinformation. Palgrave Commun 5, 65 (2019). https://doi.org/10.1057/s41599-019-0279-9*
2. *Guess, A. M., Lerner, M., Lyons, B., Montgomery, J. M., Nyhan, B., Reifler, J., & Sircar, N. (2020). A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. Proceedings of the National Academy of Sciences, 117(27), 15536–15545.* [*https://doi.org/10.1073/pnas.1920498117*](https://doi.org/10.1073/pnas.1920498117)
3. *Cook, J., Lewandowsky, S., & Ecker, U. (2017). Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. PLoS ONE, 12(5): e0175799. https://doi.org/10.1371/journal.pone.0175799*
4. *Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. Nature, 592(7855), 590–595.* [*https://doi.org/10.1038/s41586-021-03344-2*](https://doi.org/10.1038/s41586-021-03344-2)
5. Ross, R. M., Rand, D. G., & Pennycook, G. (2019, November 13). Beyond “fake news”: Analytic thinking and the detection of false and hyperpartisan news headlines. https://doi.org/10.31234/osf.io/cgsx6
6. Pennycook, Gordon and Rand, David G., Who Falls for Fake News? The Roles of Bullshit Receptivity, Overclaiming, Familiarity, and Analytic Thinking (March 22, 2019). Pennycook, G. & Rand, D. G. Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. Journal of Personality, Forthcoming, Available at SSRN: <https://ssrn.com/abstract=3023545> or [http://dx.doi.org/10.2139/ssrn.3023545](https://dx.doi.org/10.2139/ssrn.3023545)
7. Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior Exposure Increases Perceived Accuracy of Fake News. Journal of Experimental Psychology: General. https://doi.org/10.1037/xge0000465
8. Chen, C. X., Pennycook, G., & Rand, D. G. (2021, July 9). What Makes News Sharable on Social Media?. <https://doi.org/10.31234/osf.io/gzqcd>
9. Roozenbeek, J., Maertens, R., McClanahan, W.P.M., van der Linden, S. (2020). Disentangling Item and Testing Effects in Inoculation Research on Online Misinformation: Solomon Revisited. Psychological and Educational Measurement 81(2), pp.1-23. Doi: 10.1177/0013164420940378
10. Basol M, Roozenbeek J, van der Linden S. Good News about Bad News: Gamified Inoculation Boosts Confidence and Cognitive Immunity Against Fake News. Journal of Cognition. 2020;3(1):2. DOI: <http://doi.org/10.5334/joc.91>
11. Bruder, M., Haffke, P., Neave, N., Nouripanah, N., & Imhoff, R. (2013). Measuring individual differences in generic beliefs in conspiracy theories across cultures: conspiracy mentality questionnaire. *Frontiers in psychology*, *4*, 225. <https://doi.org/10.3389/fpsyg.2013.00225>
12. Maertens, R., Götz, F. M., Schneider, C. R., Roozenbeek, J., Kerr, J. R., Stieger, S., … Linden, S. (2021, July 6). The Misinformation Susceptibility Test (MIST): A psychometrically validated measure of news veracity discernment. <https://doi.org/10.31234/osf.io/gk68h>

**Digital literacy interventions:**

* Guess, A. M., Lerner, M., Lyons, B., Montgomery, J. M., Nyhan, B., Reifler, J., & Sircar, N. (2020). A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. *Proceedings of the National Academy of Sciences*, *117*(27), 15536–15545. <https://doi.org/10.1073/pnas.1920498117>
  + Relationship between digital media literacy and people’s ability to distinguish between low- and high-quality news online
  + United States and India
  + Short, scalable interventions could be effective in fighting misinformation around the world
  + Improved discernment between mainstream and false news headlines in the US by 26.5% and in India by 17.5%
  + Discernment remained measurable several weeks later in the US but not India
  + No effects among respondents in rural area of northern India where rates of social media use are much lower (this one was a face to face version of the intervention)
  + Relevant for us:
    - They used Facebook’s “Tips to Spot False News,” which were developed in collaboration with the nonprofit First Draft
      * E.g. “[b]e skeptical of headlines,” warning that “If shocking claims in the headline sound unbelievable, they probably are.”
      * <https://www.facebook.com/help/188118808357379>
      * Information was simply just shared
      * Appeared effective!
      * Respondents evaluated 16 total articles: 4 mainstream news articles that were congenial to Democrats (2 from low-prominence sources and 2 from high-prominence sources), 4 mainstream news articles that were congenial to Republicans (2 from low-prominence sources and 2 from high-prominence sources), 2 pro-Democrat false news articles, 2 pro-Republican false news articles, 2 pro-Democrat hyperpartisan sources, and 2 pro-Republican hyperpartisan sources.
      * This type of intervention more effective for older, college graduates, interested in politics, politically knowledgeable, Republican identifiers, more polarized in their feelings toward the two political parties, scored lower in conspiracy predispositions and their feelings toward Donald Trump
* Badrinathan, S. (2021). Educative Interventions to Combat Misinformation: Evidence from a Field Experiment in India. *American Political Science Review*, 1–17. <https://doi.org/10.1017/S0003055421000459>
  + Field experiment in India
  + Hour-long in-person media literacy training
  + Treatment did not significantly increase respondents’ ability to identify misinformation on average
  + Respondents who support the ruling party became significantly less able to identify pro-attidudinal stories
  + Motivated reasoning
* McGrew, S., Smith, M., Breakstone, J., Ortega, T., & Wineburg, S. (2019). Improving university students’ web savvy: An intervention study. *British Journal of Educational Psychology*, *89*(3), 485–500. <https://doi.org/10.1111/bjep.12279>
  + Study of students
  + In-person treatment
  + Students in the treatment group were significantly more likely than students in the control group to have shown gains from pretest to posttest
* Moore, R. C., & Hancock, J. T. (2022). A digital media literacy intervention for older adults improves resilience to fake news. *Scientific Reports*, *12*(1), 1-9. <https://www.nature.com/articles/s41598-022-08437-0>
  + Older adults are especially susceptible to fake news online–likely due to media literacy
  + 1-hour self-directed series of interactive modules for identifying misinformation online
  + Older adults in the treatment condition significantly improved their likelihood of discerning fake from true news from 64% pre-intervention to 85% post-intervention
    - Control group did not improve (55% to 57%)
  + Treated older adults were also more likely to employ strategies for identifying misinformation online compared to pre-intevention and control group
* *Roozenbeek, J., van der Linden, S. Fake news game confers psychological resistance against online misinformation. Palgrave Commun 5, 65 (2019). https://doi.org/10.1057/s41599-019-0279-9*
  + Convenience sample of people playing the Bad News Game and consenting to participate in a study
  + In the game, players take on the role of a fake news producer and learn to master six documented techniques commonly used in the production of misinformation: polarisation, invoking emotions, spreading conspiracy theories, trolling people online, deflecting blame, and impersonating fake accounts.
  + Large sample of 14000+, very tight confidence intervals
  + Relevant for us:
    - The interactive game appears effective
    - Participants learn the 6 most common tactics of misinformed by using them
    - Outcome: reliability of tweets: 2 control, impersonation, conspiracy, discrediting, and polarization
    - Hetrogeneities: they don’t find any significant by demographics: political leaning, age, gender, education, cognitive reflection
    - Do find heterogeneity: prior susceptibility: they split the sample by the median reliability score before the intervention and the effect is bigger for the more susceptible
* Roozenbeek, J., Maertens, R., McClanahan, W.P.M., van der Linden, S. (2020). Disentangling Item and Testing Effects in Inoculation Research on Online Misinformation: Solomon Revisited. Psychological and Educational Measurement 81(2), pp.1-23. Doi: 10.1177/0013164420940378
  + <https://journals.sagepub.com/doi/pdf/10.1177/0013164420940378>
  + Inoculation interventions are somewhat influenced by item effects, and not by testing effects
    - Item effects: Research in this area has often looked at pre- and post-intervention difference scores for the same items, which may imply that any observed effects are specific to the survey items themselves
    - Testing effects: it is possible that using a pretest influences the outcome variable of interest, or that the pretest may interact with the intervention
  + Relevant for us:
    - Suggests that using a pretest in our survey could be effective
    - We need to take care in selecting our outcome questions

**Other types of interventions:**

* Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, *592*(7855), 590–595. <https://doi.org/10.1038/s41586-021-03344-2>
  + The veracity of headlines has little effect on sharing intentions, despite having a large effect on judgments of accuracy
  + Sharing =/= belief
  + Subtly shifting attention to accuracy increases the quality of news that people subsequently share
  + People often share misinformation because their attention is focused on factors other than accuracy–and therefore they fail to implement a strongly held preference for accurate sharing
  + Relevant for us:
    - Partisan alignment was found to be a much stronger predictor of sharing than veracity
    - Outcome: how likely to share news headline on FB (balanced by partisanship and accuracy)
    - Treatment: participants were asked to rate the accuracy of a single non-partisan news headline at the outset → Effective!
    - the most obviously inaccurate headlines are the ones that the accuracy salience treatment most effectively discourages people from sharing.
* Porter, E., & Wood, T. J. (2021). The global effectiveness of fact-checking: Evidence from simultaneous experiments in Argentina, Nigeria, South Africa, and the United Kingdom. *Proceedings of the National Academy of Sciences*, *118*(37), e2104235118. <https://doi.org/10.1073/pnas.2104235118>
  + Experiment in 4 countries
    - Argentina, Nigeria, South Africa, United Kingdom
  + Fact-checks reduced false beliefs in all countries, with most effects detectable more than 2 weeks later with little variation across countries
  + Fact-checking can serve as atool in the fight against misinformation
* Basol, M., Roozenbeek, J., Berriche, M., Uenal, F., McClanahan, W. P., & Linden, S. van der. (2021). Towards psychological herd immunity: Cross-cultural evidence for two prebunking interventions against COVID-19 misinformation. *Big Data & Society*, *8*(1), 20539517211013868. <https://doi.org/10.1177/20539517211013868>
  + Prebunking (inoculation) intervention
  + A novel five-minute browser game
  + Game increases the perceived manipulativeness of misinformation about COVID-19
  + Game improves people’s attitudinal certainty (confidence) in their ability to spot misinformation
  + Game reduces self-reported willingness to share misinformation with others
  + First two effects remain significant for at least one week after gameplay
* *Cook, J., Lewandowsky, S., & Ecker, U. (2017). Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. PLoS ONE, 12(5): e0175799.* [*https://doi.org/10.1371/journal.pone.0175799*](https://doi.org/10.1371/journal.pone.0175799)
  + Focus on consensus on climate change and on anthropogenic global warming (AGW)
  + False-balance media coverage (giving contrarian views equal voice with climate scientists) lowered per- ceived consensus overall, although the effect was greater among free-market supporters.
  + Misinformation that confuses people about the level of scientific agreement regarding anthropogenic global warming (AGW) had a polarizing effect, with free-market supporters reducing their acceptance of AGW and those with low free-market support increasing their acceptance of AGW.
  + Inoculating messages that (1) explain the flawed argumentation technique used in the misinformation or that (2) highlight the scientific consensus on climate change were effective in neutralizing those adverse effects of misinformation.
  + Relevant for us:
    - Only 1000 sample and 4 treatment groups + control is too many for statistically signifcant results
    - Most susceptible (free market supporters) best candidates for intervention
    - Interventions: inoculation (explaining how tactic works), texts pre-debunking misinformation, combination, none, only misinformation
    - Online survey with simple texts
    - Outcomes: AGW acceptance, free-market support, trust in climate scientists, trust in contrarian scientists, attribution of long-term climate trends to human activity, perceived consensus, and mitigative climate policy support

Behaviors/Techniques

* Ross, R. M., Rand, D. G., & Pennycook, G. (2019, November 13). Beyond “fake news”: Analytic thinking and the detection of false and hyperpartisan news headlines. https://doi.org/10.31234/osf.io/cgsx6
  + Measured analytical thinking using CRT questions (tendency to slow down and think through questions instead of going with gut instinct).
  + Not an experiment, measured association of misinfo task performance + willingness to share against analytical thinking measure.
  + Reasoning typically helps people differentiate between low and high quality political news, rather than facilitate belief in misleading content.
  + Analytic thinking was not generally associated with increased willingness to share hyperpartisan or false headlines.
  + Would we want to consider measuring how long it takes for people to think through each question?
  + The study asked the following question: “What is the likelihood that the above headline is true (7 point scale)?”. I feel like this is a better question to ask that combines both the confidence and the true/false question in one and is also a more accurate question since they are not doing evidence searching.
* Pennycook, Gordon and Rand, David G., Who Falls for Fake News? The Roles of Bullshit Receptivity, Overclaiming, Familiarity, and Analytic Thinking (March 22, 2019). Pennycook, G. & Rand, D. G. Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. Journal of Personality, Forthcoming, Available at SSRN: <https://ssrn.com/abstract=3023545> or [http://dx.doi.org/10.2139/ssrn.3023545](https://dx.doi.org/10.2139/ssrn.3023545)
  + belief in fake news may be driven, to some extent, by a general tendency to be overly accepting of weak claims. This tendency, which we refer to as reflexive open-mindedness, may be partly responsible for the prevalence of epistemically suspect beliefs writ large
* Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior Exposure Increases Perceived Accuracy of Fake News. Journal of Experimental Psychology: General. https://doi.org/10.1037/xge0000465
  + familiarity with the headlines correlates positively with perceived accuracy of fake and real news
  + even a single exposure increases subsequent perceptions of accuracy, both within the same session and after a week. Moreover, this “illusory truth effect” for fake-news headlines occurs despite a low level of overall believability and even when the stories are labeled as contested by fact checkers or are inconsistent with the reader’s political ideology
  + We should ask “Are you familiar with the above headline (have your seen or heard about it before)? (with three response options “yes”, “unsure” and “no”)”
  + Also from an ethical perspective, we should offer a debrief question at the end that explains which of the articles were misinfo so that we do not negatively influence the participants.
* Chen, C. X., Pennycook, G., & Rand, D. G. (2021, July 9). What Makes News Sharable on Social Media?. https://doi.org/10.31234/osf.io/gzqcd
  + We find that sharing is positively predicted by two separate factors. One factor does involve the headline’s perceived accuracy, as well as its familiarity. The second, however, involves the headline’s **perceived importance and emotional evocativeness**. This second factor is negatively associated with the headline’s objective veracity, and less decision weight is put on the second factor by subjects with more cognitive reflection and political knowledge, and by subjects who are less politically conservative.
  + We can consider adding a perceived importance/severity question and/or emotional evocativeness question to our test. This can help see if the combined emotion test helps people identify the emotional questions and reduce sharing of those questions.

#### Literature summary

* Familiarity concept → more trusting of what’s familiar
* Needing to slow down → better judgment afterwards

| Paper | Content element | Delivery method | Heterogeneities | Outcomes | Other/notes |
| --- | --- | --- | --- | --- | --- |
| 1. | 6 common misinformation methods explained and used | Interactive game | Highest pre-usceptibility most affected. No other heterogeneities | Rating reliability of tweets | Effective |
| 2. | Facebook’s “Tips to Spot False News,” which were developed in collaboration with the nonprofit First Draft | Simple text | more effective for older, college graduates, interested in politics, politically knowledgeable, Republican identifiers, more polarized in their feelings toward the two political parties, scored lower in conspiracy predispositions and their feelings toward Donald Trump | Evaluate headlines | Effective, First Draft was involved |
| 3. | inoculation (explaining how tactic works), texts pre-debunking misinformation, combination, none, only misinformation | Simple text | Most susceptible (free market supporters) best candidates for intervention | Survey constructs | 1000 subjects and 5 groups not enough statistical significance |
| 4. | Asked to rate accuracy of a single element | Interactive |  | how likely to share news headline on FB (balanced by partisanship and accuracy) | At baseline Partisan alignment was found to be a much stronger predictor of sharing than veracity |
| 7 | Prior Exposure Increases Perceived Accuracy of Fake News | Pre-survey filter |  | Even a single exposure increases subsequent perceptions of accuracy |  |
| 8 | What Makes News Sharable on Social Media | N/A | People with more cognitive reflection, political knowledge, and less politically conservative were less affected by emotional evocativeness. | Sharing is predicted by accuracy, perceived importance and emotional evocativeness of content. |  |
| 9 | Choosing pre and post treatment outcome variables | Bad News (game) | N/A | Inoculation interventions are somewhat influenced by item effects, and not by testing effects | Justifies testing pre-treatment, indicates choice of outcome question in survey impacts measured treatment effect |
| 10 | Uses the same pre and post outcome questions | Bad News (game) | None based on respondent demographics/characteristics; some found based on in-game types | Judging reliability of fake “tweets” |  |
| 11 | Conspiracy susceptibility index | Survey questions | N/A | N/A | Could be useful covariates |
| 12 | Misinformation susceptibility index | Survey questions | N/A | N/A | Could be useful covariates |

### Pilot analysis

#### Covariates included in the pilot

* Age
* Residency
* Gender
* Sexual orientation
* Ethnicity
* Education
* Income
* Religion
* Political leaning
* Trust
  + General
  + Trump, Biden, Fauci, scientists
  + Various news sources
  + News on various social media sites
* News source
* Social media use
  + Frequency
  + Post
  + Goal for posting
  + Share personal info
  + Reacting
  + Sharing
  + Following
  + Angry reaction (and frequency)
  + Unfriend/block
  + Reporting
* Search for political/social news online
* Checks online news source is trustworthy
* Ability to judge whether news is true or not
* Frequency of thinking about online news accuracy
* Agree/disagree:
  + False news and information is a problem for society
  + Human caused climate change is a problem for society
  + COVID 19 is a problem for society

#### Pilot Analysis

Promising covariates to look at in the pilot:

* Susceptibility for misinformation: Trump supporter, white, strong political interests, older, education
  + Very underpowered analysis cause of lack of data so no statistical significance
  + Order of importance based on heurestic importance of plots - Fauci trust (only one with HTE in pilot), ideology (pol interest \* party), Scientist Trust, Education (3 split), Trust Trump, Age, Post on social media, Race,
  + All of them showed potential for HTE, but for time constraints can limit covariates
* ML to let the data speak
  + Promising results on pilot data, should definitely be applied for full experiment
  + Only binary treatment possible, so done for control v treatment for all treatments
    - Combined quiz had very small p-value for it’s group (split on political interest independent on party although even after removing it we got good results)
* Strong correlations in order of magnitude-
  + Trust in Fauci with ideology (party affiliation \* strength of affiliation), trust in trump, trust in scientists
  + Trust in scientists with ideology, trust in Trump
  + Trust in Trump with ideology
  + Post on social media and social media use
  + Age with white or not
  + Post on social media with age
* Important for outcome measuring propensity to share misinformation in order - Political Interest (independent of party), Scientist Trust, Fauci Trust, Race, Age
* What covariates will be important for stratifying?
* Pilot analysis summary
  + Ideology, Trump, Age, Trust - consider for stratifying/targeting
  + Tactic quiz more dense more effective
  + Combined quiz lower effectiveness
  + Timing tactic and duration - less accurate
  + Longer people took - more accurate

## Treatments

Goal: Come up with a proposed targeted course to use in our 15 min survey.

* Courses available: <https://docs.google.com/document/d/1xTCTgb9do5Y_5p3SdATcHGL1PFVzG9QQYWOwtHZkprk/edit>
* What components of these courses do we care about?
* Are there examples that can be used to better convey the knowledge? (eg better quizzes?)

Tactics: <https://twitter.com/johnfocook/status/1335628473533083649/photo/1>

Pilot Treatments

* Tactics
* Tactics + Emotions
* Quiz Variations

Specific treatment elements/options/considerations

* Topics and location / use covariates for targeted treatment?
* Short snippet and effects?
* Very polarizing stuff - avoid
* More interactive the better, shorten/less text and more little quiz questions
* Emotions course appears less effective
* Pilot combined confusing/too much
* Focus on fewer takeaways/key points
* Adding in headers
* Less is more / focus on specific points and deliver them clearly and effectively?

In the pilot, tactics followed by combined quiz were the most effective overall for recognizing misinformation.(main outcome)

For a subsidiary outcome testing if participants will share misinformation on social media, tactic quiz, followed by tactics and combined quiz (almost equal) were most effective.

Components

Additional Potential Treatments

* Examples focused course?

## Covariates

* See pilot analysis
  + Covariate groups:
    - Income, Reflect, CheckSource, SearchPoliticalNews, Education\_level\_2, pol\_interest, age\_subgroup, ReactSocialMedia, ShareSocialMedia
    - FauciTrust, White, Ideology, CovidThreat, TrumpTrust, Ideology, ClimateThreat, FakeNewsThreat, GeneralTrust
    - BlockUserSocialMedia, AngrySocialMedia, FollowIssueSocialMe
* **Order of importance for HTE :** Income (4 split), Fauci trust, Reflect, CheckSource, SearchPoliticalNews, ideology, Scientist Trust, Education (3 split), Trust Trump, Age (4 split), Post on social media, Race
* Use of covariates:
  + What to stratify on - e.g. Misinformation susceptibility, conspiracy mindedness
  + What to include - Correlation between covariates
  + What to target on ? - Going all in on one covariates and sample only certain people (?), presecreen (?)

## Outcomes

Considerations

* Add debrief?
* How strongly do they feel about it?
* Are they looking ok - which outcomes do we care the most about? → ask them
* Sources for specific questions: pilot, other literature papers, ask them, might need to find more relevant examples and good fake graphs

Questions for Partner

* Are there examples that are less obvious topics of misinfo? Vaccine misinfo seems too easy to identify.
* Are there examples of better looking graphs that are misinfo?

### Survey Test

Proposed Test (DON’T OPEN YET): <https://docs.google.com/presentation/d/1GK1GH2zmvdPdErncjL50juUftGFu47HIeOmniMSK74U/edit#slide=id.g108fca4174c_1_0>

* Examples to select: <https://docs.google.com/presentation/d/1OB4-QFeHRR1ET1IS7duMIRckWEudKll2n8odks4nkZg/edit>
* Which examples should we select that are suitable for testing? Should they be aligned to the material taught in the course?
* How do we select a pre-test and post-test that are comparable?

Dogfooding

* Post test seems very slightly harder, though the weighted confidence versions are relatively correct.
* Truths are hard, at least 1 person was wrong for each of the truth questions except for the pre-test graph
  + Apparently aloe vera does help with burns
* 2 most missed misinfo
  + Carrots don’t cure diabetes
  + Autism not associated with vaccines (need to read carefully)

Time limit:

* Demographics: 3 minutes?
* Treatment: 5 min
* Leftover, 7 min
  + How long per question? 30 sec? - It really depends on the length
    - For simple questions: <https://verstaresearch.com/newsletters/how-to-estimate-the-length-of-a-survey/>
    - For longer text: word count and average reading speed: 250 words / minute

Pre Post Test vs Control - see experiment design section

### Outcome Questions

Ideally we want to see how the course helps users identify misinfo. Given the complexity of misinfo and the limited exposure to the treatment, we likely can only measure a proximate outcome of whether users become more wary of misleading content and take time to think/fact check before sharing/spreading potentially false information.

* Which outcome questions do we want to ask?
  + **Misinfo:** What is the likelihood that the above headline is true (6 point scale)?
  + **Sharing:** If you were to see the above article on social media, how likely would you be to share it?
  + **Familiarity:** Are you familiar with the above headline (have you seen or heard about it before)? (with three response options “yes”, “unsure” and “no”) → this one seems more like a pre question?
  + Additional Potential Questions
    - Fact Check: Willingness to fact check (how do we ask this in a way that is reflective of true response?)
    - Emotions: “How exciting is this headline?” “How worrying is this headline?”
    - Severity: “Assuming the headline is entirely accurate, how important would this news be?”
    - What would you like to do next
      * Fact check
      * Read more
      * Share it
      * Show it to someone
      * Report
      * Nothing
      * Other
* How do we ask these questions? Will order matter?
  + Sharing before misinfo? If we do vice versa, we are prompting them to think if it is misinfo before sharing, which literature has shown to decrease willingness to share. If we are testing intervention of the course, we might want to ask sharing first? → good point. It comes down to what First Draft cares about
  + Misinfo before fact check?
    - Thinking about fact checking first might prompt users to believe the content has misinfo.
  + Willingness to share misinfo (how do we compared to normal content, like if someone just doesn’t share things in general)
  + Likeliness to fact check
  + agree with the claim
  + trust the information
  + severity of the claim

## Questions to ask the partners

* To clarify:
  + “The organization First Draft News has created the “inoculation against misinformation” course that contains different kinds of content. Our first question is to identify Which components of the course content are most effective at making respondents better at identifying misinformation and less willing to share misinformation posts on social media?” → Many options. We also know that many other similar research projects have been conducted e.g. Guess et al. What elements do they care the most about?
  + How about delivery elements? (Based on what we see in the literature, more engaging approaches look more promising)
  + Do they want to focus on any particular subgroups? E.g. most susceptible?
  + What is the most important outcome they care about? (E.g. wanting to share, identifying misinformation)
  + What is the minimum effect that would be useful? (For power calculations)
* Our findings
* Ideas
* Analyze the pilot data and form hypothesis tests that we can discuss with them
* What additional courses/variations are available that you would be interested in testing? Anything higher priority than what is already in the pilot?

## Experiment design

Considerations

* Generalizability?
* Reading comprehension versus ability to spot misinformation
* “You will receive project-specific tutorials that give you guidance on the format and contents of the pre-analysis plan. You should ensure that your pre-analysis plan is precise enough for the teaching team to implement your experiment as designed.”
* Pre+Post Treat vs Control-only Design
  + Advantages of pretest
    - If we suspect that the population of users that drop out of the treatment (don’t want to read through all the information) would be different from those that stay, then we would want a pretest design.
    - Helps reduce between person variability since we match each person to themselves. This can help us have greater power.
    - Many papers in the literature seems to be doing this
  + Disadvantage
    - Longer survey due to initial test
    - Will having the initial test cause people to be able to guess what we’re trying to test and thus artificially inflate their performance?
    - Our pre and post tests need to be comparable.

Covariates

* What pretreatment/demographic questions to ask?
  + Do we need to stratify?
  + For the purpose of HTE
* Additional potential covariates
  + Location (Urban vs Suburban vs Rural)

Treatments

* What treatments to include?
  + Components of the treatments
  + Type of treatments (pedagogical approaches)

Outcomes

* What outcome do we want to show changes to?
* Ability to answer test questions correctly (may correlate with misinfo identification for the specific tactic being tested, but not as generalizable to all misinfo)

Overall

* How many data points do we need to collect? What is our budget?
* 3 levels of Qs - which components, what pedagogy, which groups to target

## Analysis

* multiple hypothesis testing vs ML split for finding HTE

### Outcome variables

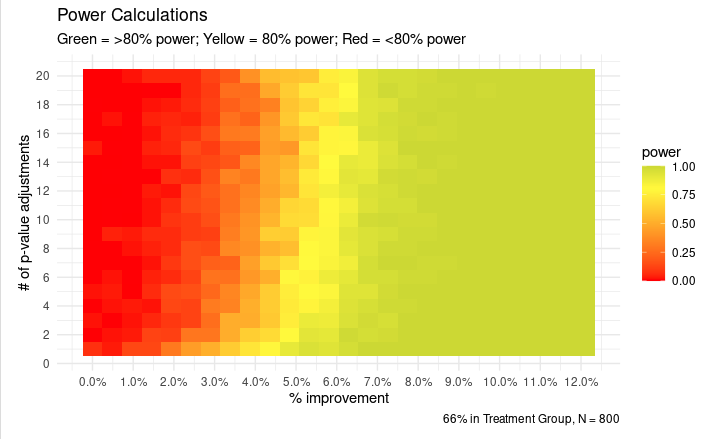
* (1 test) Ability to identify manipulative content (1-6 scale, score based on that and whether misleading or not) (Average over 6 questions)
  + (Secondary outcome, binarized version, code 1-3 vs 4-6 as correct answer or not and average over these)
* (3 tests) Ability to identify specific tactics correctly
  + Binary variable about each tactic correctly identified or not (Average over 6 of these questions)
  + 3 such variables for Graphs, Anecdotes, Comparisons (note: the graph outcome might be the average of multiple sub graph questions)
* (2 tests) Desired behavior
  + Likelihood of sharing on a scale of 1-6 (score based on 1-6 scale and whether misleading or not) (Average over 6 questions)
    - Weight this by average propensity to share at baseline
  + Same as above for another behavior, reporting
* HTE (2-3-4?)
  + Focus on the main outcome of correctly identifying
    - HTEs:
      * Susceptibility at baseline (use a split on the median so have half and half of the sample (?))
      * Ideology
      * Income
      * Age, race (or combined these with above, or drop them)
      * ML HTE (doesn’t count for power analysis)

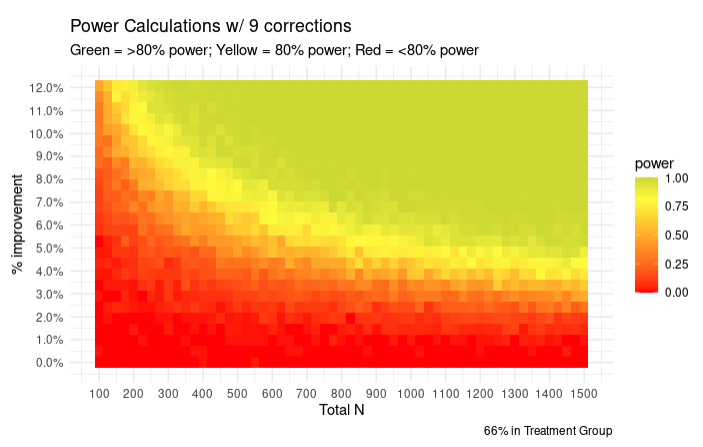
### Power Analysis

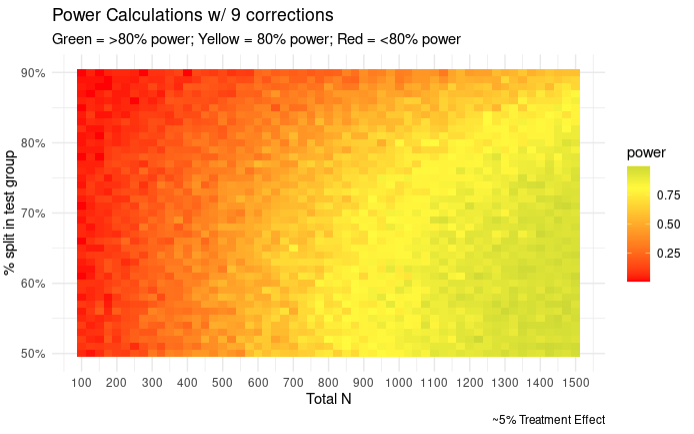
* benjamini hochberg correction for significance level
* (1000-1500 samples)
* Split of treatment and control (**N800, 9 corrections, split vs % effect**)



* HTEs (mention how many HTE’s we can test) (**N800 and split fixed; % effect vs number of corrections**)





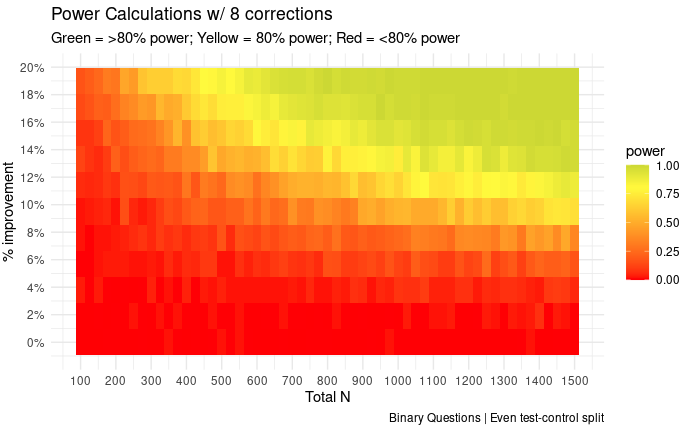


N = 800

Split

Variable effect

9 corrections



TODO:

* Use an appropriate but less conservative approach than Bonferroni (see in the R file given to us or based on literature - in this case include the source.)
* Power calculation given the 6 tests we have + taking into account the (2?) HTEs as appropriate (these should be less penalizing?)
  + Explain the rationale for the approach (with citation) if using something other than the method given to us
  + Create the appropriate graphs to motivate number of tests we can do
    - Use the pilot for control and treatment variance
      * Control:
        + Within people change by picking 5 of the outcome questions randomly for each individual to be the pseudo pre and the pseudo post outcome questions. Make sure to pick half of the true and hald of the false questions as pre and post for each person.
        + Calculate the average difference between the 5 pre and 5 post questions for each person → find the variance
      * Treatment:
        + Between people approach
        + Compare similar people who are in control versus in the different treatments (e.g. propensity score matching, match each control person to one in each treatment arms). Take the difference between the average answers in each of the treatments and in the control, for each control person. Find average difference across the 10 questions for each control person. → find the variance
        + Alternative approach to do it on the group level but I’m not sure if that’s an appropriate comparison with the control approach?
      * Which variables to use from the pilot:
        + For the main outcome, use the confidence weighted misinfo or not pilot variable
        + For the binary ones, use the binary version of the misinfo or not pilot variables
        + For the behavior question, use the sharing pilot variable
      * Control treatment split ratio
        + Follow the suggestion of this: <https://blogs.worldbank.org/impactevaluations/when-should-you-assign-more-units-study-arm?CID=WBW_AL_BlogNotification_EN_EXT> (seems like out best bet given that control should have significantly lower variance)
    - Given the above, know the alpha, beta, and variances, create graphs with N and minimum detectable effects
    - Calculate expected effect based on the pilot, for reference.

### HTE & Machine learning approach

Appendix

**Heuristic analysis** on plots for the main outcome-

**Importance -**

Fauci trust (only one with HTE in pilot), ideology (pol interest \* party, 0.06 - 0.01)

, Scientist Trust (0.19 - 0.11) Education (3 split) (0.18- 0.09), Trust Trump (0.24 - 0.04), Age (0.43 - 0.13), Race (0.28 - 0.7),

**Age**:

2 way split did not work well.

4 way split:

<30 Tactic quiz worked best (tactics did not)

30-45 tactics quiz and Tactic (roughy same) worked well

45-65 combined quiz worked well (rest decreased score)

For >65 Tactics & Combined worked best

Should include age (potential for HTE)

Older people had better scores (does this make sense or should we change outcome?)

**Race:** white v non-white

Non-white Tactic quiz and combined quiz worked best

White Tactics worked best

Whites had higher score (and also more improvement)

Should include race (potential for HTE)

**Education:** lower than or higher than bachelors

Lower combined quiz and tactics worked best

Higher combined and tactics worked best

Higher score for higher than bachelors

High p-value

Should include education

**Ideology** (Political interest with party): club with party of interest for better results

**Trump Trust**

Trust trump - Tactics worked best (improvement for all though)

Do not trust Trump - little improvement max in combined quiz

Higher scores for trusting Trump, although more improvements for those who do not

Should include

**Scientist Trust**

Trust Scientist - Combined quiz worked best (very little improvement)

Do not trust Scientist - improvement for all max in Tactics

Close to significant for multiple treatments (tactics, +quiz)

Higher scores for not trusting Scientist, although more improvements for those who do not

Should include

* Mention unequal split between control and treatment (exact %)
* Research Q