Methodology for Studying the Influence of Facebook Ads on Meat Reduction

Draft 4.1.3.0 | Last updated 4 January 2015

0. Introduction

Animal Charity Evaluators (previously Effective Animal Activism) suggests that one of the most cost-effective interventions for helping animals is to fund Facebook advertisements that link people to videos that attempt to convince people to reduce or eliminate eating animal products. The ultimate goal is to reduce demand for meat and thereby reducing nonhuman animal suffering in factory farms (see also Brian Tomasik's "Donating toward Efficient Online Veg Ads"). The case for desiring to reduce animal suffering has been made elsewhere and will not be made here.

The case for advertisements was originally made on the basis of two studies that suggest a cost-effectiveness of $0.03 to $36.52 to reduce a year of nonhuman animal suffering in a factory farm, after several considerations are taken into account, though later research (such as ACE's leafleting study and their humane education study) has called somewhat into question these estimates and re-calculation is needed.

Existing studies appear to suffer from a few problems, most notably of which is insufficient sample size. However, other problems include (a) really high non-response bias, (b) a lack of a usable control group, (c) difficulty with nutrition questions, and (d) concerns of response bias and general social desirability bias.

It's important to address these concerns before deciding to potentially funnel hundreds of thousands more dollars of effective altruist money into this cause. Recognizing this, Jeff Kaufman called for improving research into the impact of ads on diet choice. Jason Ketola saw this proposal and wanted to make it into a reality. He got in touch with Peter Hurford who was also interested in working on it. Together, Peter created the first draft

of this methodology document and passed it around for comments while Jason researched the technological feasibility.

Around the same time, Nick Cooney, working for Mercy for Animals, was interested in designing a rigorous study to test MFA's use of Facebook advertisements. He saw Peter's methodology and decided to use it for his experiment. He later contacted Peter and Jason and teamed up with them to do one joint study. They contracted a reputable third-party analysis firm, Edge Research, to do the analysis and consult on the methodology. Ads will be distributed using MFA's existing ads and platform.

Thanks to several generous supporters, the study will take place in early 2015.

1. Table of Contents

2. Objectives

- **Main Objective:** Design and launch a study that reliably assesses the cost-effectiveness of funding pro-vegetarian advertisements on Facebook that direct individuals to landing pages playing particular videos, without falling to the pitfalls of previous studies.

- Use a study that contains a control group and a treatment group with a large enough sample size to detect potentially the small ad effect size currently suggested by research (0.5% to 3% conversion rate, see “Conversions per Pamphlet” in “How Much Does it Cost to buy a Vegetarian” for an overview) at 80% statistical power or higher (see section 4.1 for details).

- Employ reliable diet questions that mitigate commonly observed pitfalls as much as possible (see section 4).

- Avoid more complex tests, like comparing treatments, as it’s good to start out simple, and the sample size needed for just assessing one treatment is already large.

- Be generally statistically rigorous and reliable, avoiding common problems that allow statistics to go to the dark side.

3. Methodology

3.1. Methodology in Brief

This will be a field study where participants are unknowingly and anonymously recruited via a Facebook ad that they willingly click (see section 3.2). Once the ad is clicked the participant will be automatically sorted into either a treatment group where they view a

website with an “anti-meat” message intended to influence diet (see section 3.3) or a control group where they view a website with a control vidieo not intended or expected to influence diet choices -- currently up in the air between a video advocating donating to eliminate neglected tropical diseases or a video advocating reducing world population (see section 3.4).

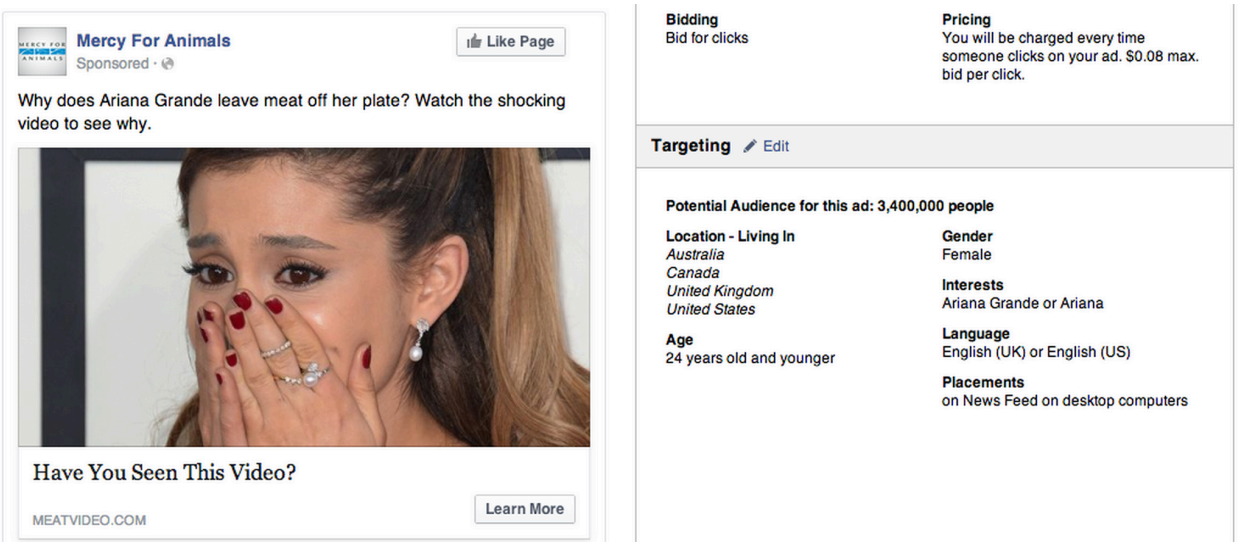
Visiting these websites will place a cookie on the user’s computer that can be used to “retarget” them via Facebook Ad Exchange and Google’s Ad Network for participation (see section 3.6) in a follow-up survey three months later (see section 3.8). All the user will see is a Facebook ad offering them a chance at a prize (yet to be determined) for completing a survey (see section 3.7). (To be clear, initial advertising is only on Facebook, but follow-up retargeting is on both Facebook and Google ads.)

There will be no connection between this survey opportunity and the previous vegetarian ad they clicked on three months ago. This will allow us to deliver a survey to both groups without tipping them off that it’s vegetarian related, thereby reducing bias (see section 3.8). To be clear, both groups will see identical initial facebook ads and both groups will get identical follow up surveys after identical recontact time horizons.

This assessment will then be compared across the treatment and control group to figure out the influence of the Facebook ads and associated landing page videos on diet change. The amount of reduction in the treatment group relative to the control group, if any, will then be compared to the cost of procuring additional Facebook ads to assess cost-effectiveness (see section 3.9).

3.2. The Facebook Ad

In order to get people to watch our videos, we have to get people to click our ads. Therefore, our ads have to be appealing. Nick Cooney has already run >$50K in online ads, using in-page ads (as opposed to sidebar ads) that look like this:



Ad prices are chosen via an auction and vary depending on cost per mille (cost for the ad to appear one thousand times) or cost per click (cost for each time the ad is clicked) and one can use either method for their ad payment.

Since we will be running our survey within existing MFA ad campaigns, we will be able to use the ad network that MFA has already set-up. This means that we increase our external validity with regard to testing MFA's set-up (since it is identical) and we get to have our ad campaign run by experienced campaigners.

We will also be using MFA's targeting for their ads, showing ads only to women aged 13-24, which Nick Cooney says are the most likely to consider becoming vegetarian (see "Well-Planned Facebook Ads: The Most Cost-Effective Way To Create New Vegans, Vegetarians and Meat Reducers?").

3.3. The Treatment Webpage

We will be using "Meat Video" (http://www.meatvideo.com/) as our treatment page, though possibly on a different domain. The page will be modified to include the retargeting cookie.

3.4. The Control Webpage

We will set up a site that is identical to our treatment page in layout, with the only difference being the content will be swapped with a control video not meant to influence diet choice.

However, since we will be showing this control page to thousands of people, we still should take an opportunity to do good. Therefore, we will show END7's video about reducing neglected tropical diseases and point participants to a donation page specifically set up for our campaign (so we can track our donation impact).

3.5. The Control-Treatment Randomization

Navigating between the treatment and control page is complex. In order to resolve this, we will use A-B testing software Optimizely, which takes a user and places them into either a treatment page or control page.

There are several things that could go wrong, which I present here FAQ style.

Q: *What happens if the user returns to the page again, such as by clicking the ad a second time? Won't it be jarring to see a treatment page and then later see a control page? And won't that mess up analysis?*

A: This is a critical problem that we need addressed. Luckily, Optimizely "freezes" that page in place (via cookies) so that if the user returns to the page again they will have a consistent treatment or control experience, so this problem is resolved.

Q: *Won't it be jarring for a user to click an ad about, say, why Ariana Grande leaves meat off her plate, and then end up with a control video about, say, NTDs?*

A: Yes, this is jarring. However, the problem does not appear to be one we can address, so we'll just have to leave it be. It does not seem like this would pose that much of a problem, and the cost of addressing the problem is higher than the benefits of mitigating it.

Q: *What happens if the user shares the site on social media?*

A: Further users that visit the site will themselves be sorted into either the treatment or control group, and they themselves will then become eligible for the retargeting ad offering the survey. This is an important part of the testing, as the further reach of the pages is part of the impact that MFA hopes to have.

Q: *What happens if the user clicks on a share, notices it goes to the control page, and ends up confused? Will this be jarring? Will this risk exposing the study if the two users talk?*

A: As far as we can tell, this kind of situation may occur. We are not yet sure how to resolve this problem.

3.6. The Retargeting

Whenever a user enters the site, they will have a cookie placed on their computer that notes whether they were in the treatment group or the control group. The Facebook Ad Exchange can then be used to send ads to all people with one of those two cookies, and the contents of the cookie can be noted by the ensuing survey software.

Implementing the retargeting will be done by Edge Research, using MFA's existing ad platform.

3.7. The Survey Ad

Sometime after the study, we will post ads advertising a survey in exchange for some prize (prize and language of the ad are yet to be determined). This ad will have no noticeable link to MFA, vegetarianism, or the fact that the participant had clicked on an MFA ad three months ago. Clicking the ad will take the user to a survey.

The original ads will be open for three months, after which there will be a three month retargeting window to collect follow-up surveys. This means we will survey users somewhere between one and six months after the original intervention.

Unfortunately, it currently looks like we will be unable to capture the retargeting distance (one month vs. six months vs. somewhere in between) for each user, so we will not be able to use that in our analysis.

All data collection will be handled by Edge Research.

3.8. The Follow-up Survey

3.8.1. Initial Thoughts

The survey is still being figured out and the design of the survey will be finished in partnership with Edge Research. Preliminary pilots of survey questions will also take place on Amazon's Mechanical Turk platform in early 2015, using other third-party research partners (other than Edge Research). What follows are the preliminary

thoughts on the survey design, informed a little bit by consultation with Edge and some MTurk piloting, though more work is needed.

It's important to be very careful with this survey, as participants can be fickle. Overall, we face a trade-off between (a) making the survey detailed, so as to collect the most useful information and (b) making the survey simple, so as to not fatigue participants, reducing response rate.

We originally expected disguising the survey would be worth the costs, but our preliminary A-B test on MTurk showed this not to be the case. Therefore, we're going to go with a small survey with only minimal disguising (the questions themselves look about diet, but the survey ad is still unconnected to diet choice in anyway, and nothing is clearly about vegetarianism).

3.8.2. Survey Draft

Currently, I propose a survey that would look something like this:

1.) In the past two days, how many servings have you had of the following?

	0	1	2	3	4	5	6	7	8+
Dairy (cheese, milk, yogurt, etc.)									
Chicken (fried chicken, in soup, grilled chicken, etc.)									
Turkey (turkey dinner, turkey sandwich, in soup, etc.)									
Fish and Seafood (tuna, shrimp, crab, etc.)									
Pork (ham, pork chops, ribs, etc.)									
Beef (steak, meatballs, in tacos, etc.)									
Eggs (omelet, in salad, in baked goods, etc.)									

2.) What is your age? (drop-down)

3.) What is your gender? (drop-down)

3.8.2. Survey Draft

This survey primarily relies upon a Food Frequency Questionnaire, which seems to be the most efficient way to assess both meat reduction and meat elimination. Assessing meat reduction is important, as MFA expects most of their impact is in reducing meat consumption as opposed to inspiring people to outright eliminate it.

The survey also has distraction questions within the FFQ (fruit, vegetables, soft drinks, and nuts) so as to make the survey appear less to focus primarily on diet. There also is another parallel FFQ-like matrix that asks about other habits. Together with demographic information, this survey is quick and can be passed off as a general market research survey rather than an attempt to assess diet change.

3.9. The Analysis

The Analysis will be handled by Edge Research. They will produce initial cross-tabular findings, as well as an anonymized raw data in an SPSS file or CSV file. This data will be made public.

Later, Edge Research will complete “a final data report including (a) an outline of the research methodology and rationale, (b) high level findings and takeaways, and a (c) drill downs on specific areas and audiences”.

It’s currently unclear what precise methodology Edge Research will use to analyze the data, but the expectation is that they would use a Chi-Square test to compare the food frequency questionnaires between both the treatment and control groups, looking both for meat reduction and elimination.

3.10. Debriefing and Research Ethics

Some might point out that this survey has more questionable research ethics, since great length is taken to prevent participants from realizing they’re in a study, and participants are being actively manipulated through graphic video to change their behavior.

From a utilitarian calculus, it seems like the good of reducing meat intake *greatly* outweighs any harms of graphic material or studying on people without their immediate consent. Indeed, participants might even end up being thankful for being shaken to

greater moral awareness. Harms to participants are brief in duration and intensity, and not too out of line with what a participant would experience in everyday life. The graphic content is needed in order to inspire the desired behavior change, and the deception around the study is needed in order to achieve reliable results.

However, there's no need to pursue vegetarianism at any cost. Instead, we should aim to debrief participants about the nature of the study upon its completion. We can also use this opportunity to follow-up with participants to provide support for maintaining any diet change they may have made.

4. Previous Problems and How They Are Addressed

4.1. Small Sample Size

The solution to this problem should be simple -- increase the sample size.

4.1.1. How many have we had?

But looking back on past studies, how have we done so far?

Sample Size of Previous Studies

Name	Date	Treatment	Control
Cooney's FB Study	Fall 2011	104	Did not use
THL's First Leaflet Study	Fall 2012	"nearly 500"	Did not use
ACE's leafleting study	Fall 2013	123	23 (control leaflet), 477 (no leaflet)
ACE's Humane Education Study	Fall 2013	169	60
THL's Second Leaflet Study	Fall 2013	524	45

This is a concern because if our prior expectation is that people will become vegetarian 0.1% to 3% of the time, any actual effect with these sample sizes could be absorbed in statistical noise. The control groups are also *far* too small to work.

4.1.2. How many would we need?

But how large of a sample would we need? To figure out what sample size we can use, we can use an "a priori sample size calculator for the two group chi-squared test" (since a two group chi-squared test will be our method of analysis, though David Chudzicki suggests avoiding power calculations due to concerns mentioned in this paper). For this, we need a desired power level (chance of detecting the effect given that the effect exists), a probability level (our threshold p-value), and an estimate of the proportion of people who become vegetarians (or vegans, or reduce meat, etc.) in the control group and treatment group.
t
Any desired sample size is going to be a trade-off between wanting more statistical power and wanting to spend less money collecting the sample. An additional consideration is that the desired sample size is the number of people we need to take the follow-up survey, which might be a smaller proportion of the total number of people that see the ads, due to inefficiencies in our conversion process.

We asked some people on Facebook what minimum difference level they cared about, thus allowing us to decide what bounds we needed to be able to have enough statistical power to find within our study. We got back 0.5% (3 people), 1% (4 people), and 2% (2 people). Depending on the expected rate of vegetarians among the control group, we'll need to recruit somewhere between 3.2K and 63.2K survey respondents.

Here's a table with some estimations with the classic probability level = 0.05 and desired power level = 0.8, using two different estimates for the control group. "Control %" refers to the baseline number of vegetarians found in the control group, "Treatment %" refers to the number of vegetarians found in the treatment group. "Difference" (treatment - control) refers to the amount of vegetarians that could therefore be attributed to the treatment. "Sample size needed" refers to the amount of total survey completions that we would need in order to reliably assess whether the given difference is statistically significant, or just random chance.

Control %	**Treatment %**	**Difference**	**Sample Size Needed (treatment + control)**
3%	5%	+2%	3210
3%	4.5%	+1.5%	5298
3%	4%	+1%	11000
3%	3.5%	+0.5%	40294
5%	7%	+2%	4624

5%	6.5%	+1.5%	7826
5%	6%	+1%	16718
5%	5.5%	+0.5%	63282

This means that if we expect the effect to be small, we'd need to recruit easily 10-100x more people than have been in previous studies.

4.1.3. Considering response rates

Note that if we need 7826 people in our survey sample and expect only 10% of the people who click the ad to answer the survey, we'll need to fund and get nearly 100K people to click on the ads. If we need 63K survey respondents, we might need 630K ad clicks or more! So we'll be drawing from a large pool of participants.

4.2. Non-Response Bias

Past studies had large issues with non-response bias. The past Facebook survey recruited people into the follow-up survey by offering a 1 in 10 chance of winning movie tickets through either (a) another Facebook ad or (b) through email.

While there is a concern that fewer people might see the second ad or choose to click on it, the real problem is that the Facebook ad was only shown to people who "liked" the vegetarian video page on Facebook and the email was only sent to people who provided their address in exchange for requesting a free "veg kit". This means the responders are skewed towards those who liked the site enough to interact with it and show some sort of approval (liking or entering an email), thereby skewing the survey sample to those who are more likely to be pro-vegetarian. The same is true of the leafleting study, though less so, which only collected data from people who could remember receiving a leaflet.

Ideally, there should be no such difference between the population that responds to the survey and the population that does not. The method we use, the retargeting cookie, will display the survey to all people who clicked on the original ad -- regardless of their further interest -- therefore sampling from the entire population we're interested in.

It's possible that our sample could still be biased if, for example, people more willing to go vegetarian were also more willing to click on ads that offer rewards for taking surveys. However, the ad offering the survey should bear no relation to the ad advertising vegetarianism, so we regard this possibility as unlikely.

4.3. Insufficient Control Group

It's cool to collect information on the impact of the advertisement and current data points to potentially large effects -- 2% to 3%. But is it possible that people are going vegetarian anyway, without the effect of the leaflet? Perhaps our survey is just capturing a general natural shift? This may be a good thing on its own (as long as there's no equally corresponding shift from vegetarians back into meat eating), but it still means we're wasting our money on advertisements. Therefore, we need a control group.

Since February when we recently announced the need for a control group, several studies have been performed that included a control group. This is a good step forward, but the control groups still have not been large enough to be useful (see "4.1. Small Sample Size").

4.4. Response Bias

Response bias is an issue very different from non-response bias -- people respond, but consciously or unconsciously respond in a way to please the researchers. That is, if the survey is run by a pro-vegetarian organization, people may be more likely to claim they are vegetarian, even when they're not.

To mitigate this problem, we aim to make no apparent connection between the original video and the later survey (see section 3.7) and the survey itself will be disguised with distractor questions so as to minimize the chance a participant thinks the purpose of the study is assessing one's vegetarianism (see section 3.8).

4.5. Social Desirability Bias

Social desirability bias is a related concern where people "lie" about being vegetarian not because they think it's expected of them by the pro-vegetarian researchers, but because society generally views it as virtuous to be vegetarian and people consciously or subconsciously want to identify as virtuous. This is a bit difficult to work with.

One idea is to include the Marlowe-Crowne Social Desirability Scale and tracking how it correlates with declared vegetarianism, though we currently think this would make the survey too long. Another idea is to make a post-hoc adjustment to the vegetarian rate based on external research. A third idea is to assume the social desirability will be equal in both the treatment and control groups, so we shouldn't be concerned.

We're not currently sure how to handle this yet.

4.6. Bad Nutrition Questions

People are really bad at understanding what it means to be vegetarian, which is how you end up with self-proclaimed vegetarians who have eaten meat just the previous day. Likewise, there is controversy over whether the meat category includes things like fish or chicken. We need better nutrition questions that can handle some of these difficulties without making the survey too complex or long.

We plan on solving this by using a Food Frequency Questionnaire, which is pretty standard in the literature and is supported as a best practice by Animal Charity Evaluators.

5. Funding

5.1. Initial Thoughts on Funding Needs

The costs of the study are very hard to estimate as the desired sample size is very much in the air and retargeting costs are rather variable. Here's our best guess:

- We need to get at minimum 3.2K people to take the survey to have any reasonable hope of finding an effect. Ideally, we'd want say 16K people or more.
- 3.2-16K survey takes requires, at minimum, 3.2-16K ad clicks, which, at $0.20ea, involves $640-$3.2K.
- We'd also have to offer a prize, which varies depending on what we offer. Maybe this would cost $0.20ea more per person who takes the survey, driving up costs to $1.3-6.4K.
- We'd also need to have everyone who takes the survey click on the original ad for either the treatment or control group, so that's another $0.20ea, for a new total of $1.9-9.6K.
- We need domain names for the video and for the survey. That's $22. Pretty cheap compared to the rest.

- Of course, the above is still pure fantasy, as it would require 100% of people who get sorted into a treatment or control group to also end up not only clicking the ad to take the survey, but end up completing enough of the survey to be useful. The actual number of survey completions per ad click is a very crucial number that we don't have yet.

- Our costs can then be modelled as **[($0.20+p) + $0.20/r]s + 22 + t**, where **p** is the cost of the prize (survey incentive) per survey completion, **r** is the percent of people who view the initial ad to go on to complete the survey, **s** is our desired sample size, and **t** is the value of the time spent on this study.

- The best we can do is pilot test to find out what **r** would be. But we could guess. If **r** is 0.02, then our total costs (assuming **p** is $0.20 and not counting **t**) would be $33.3 - $166.4K. Ouch, that would be expensive! If we got a response rate of 10% instead, our costs would go down to $7.7-38.4K, which is more manageable.

- This gives an interesting optimization problem between **p** and **r** -- to the degree that a larger prize would draw in a higher response rate, it might be worth offering something better. Though I could imagine that too high of a prize might scare people off as being "too good to be true". Unfortunately, I don't know if we would get all that many opportunities to test this that much.

5.2. Final Budget

We ended up budgeting to get a sample size of 6K (3K in the treatment and 3K in the control) because this sample size seems sufficient to provide enough statistical power to make our study reliably detect small effects (see section 4.1.2). More importantly, the underlying ad spend we need to make to recruit 6K participants at expected response rates is at the upper threshold of what MFA's current ad network can handle without loosening it's targeting (and expected cost-effectiveness).

Item	**Cost**
Fees for Edge Research	**$9.5K**
MFA ads that lead to treatment	166K ads @ $0.17ea = **$28.2K**
MFA ads that lead to control	166K ads @ $0.17ea = **$28.2K**
Survey ads for both groups[1]	6.6K ads @ $0.45ea = **$3K**
Survey incentives	**$3K**

TOTAL: $71.9K

[1]: This cost assumes a 2% response rate (that is, 2% of those in either the treatment or the control group will later click on the ad to take the survey) and a 90% through rate (that is, 90% of those that click on the ad will complete the survey). If these two numbers are higher, we will adjust our spending accordingly, aiming to maximize the number of successful survey completions.

5.3. Total Funding Procured

To date, we have acquired $42,700 from various donors to use specifically on this survey project. We are also using $28K of MFA's existing advertising spending that would have already been spent on treatment ads, so we have $70,700 in total to spend on this survey.

This is enough that we're comfortable going forward with the study as-is and are therefore not widely advertising a need for funding, but we would like to raise an additional $1.2K to scale up the study to as big as we think it can be.

6. Resources

6.1 Resources Consulted

- Peter Hurford's "How Much Does it Cost to Buy a Vegetarian?"
- Brian Tomasik's "Donating Toward Efficient Online Veg Ads"
- Nick Cooney's "Well-Planned Facebook Ads: The Most Cost-Effective Way To Create New Vegans, Vegetarians and Meat Reducers?"
- Nick Cooney's Facebook Ad Survey
- Jon Bockman's Methodology for a Video Outreach Study
- Joey and Xio's Methodology for a Leafleting Study
- Peter Hurford's "A Re-analysis of ACE's Leaflet Study"
- Jason Ketola's "Facebook Adroll Experiments"

6.2 Other Relevant Discussions

- Jeff Kaufman's vegetarian advocacy survey proposal
- Jason Ketola's initial Google Doc and associated conversations

- Peter and Jason's 30 Dec Skype Conversation and associated Facebook discussion
- FB discussion of Joey and Xio's leafleting study
- Peter and Jason's 5 Jan Skype Conversation and associated Facebook discussion
- Vegetarian Advocacy Study Meeting, 7 Sep 2014
- Vegetarian Advocacy Study Meeting, 12 Aug 2014
- Zachary Groff's Comments and Discussion