

Santtu Kauppila

**Unraveling the Reason of Participation in Crowdturfing
Campaigns using Neutralization Theory**



UNIVERSITY OF JYVÄSKYLÄ
FACULTY OF INFORMATION TECHNOLOGY
2020

ABSTRACT

Kauppila, Santtu

Unraveling the Reasons of Participation in Crowdturfing Campaigns using Neutralization Theory

Jyväskylä: University of Jyväskylä, 2020, 70 pp.

Information Systems, Master's Thesis

Supervisor(s): Soliman, Wael

Disinformation is thought to be ubiquitous in digital space, especially now that a booming industry is beginning to evolve from previously honest crowdsourcing enterprises. These crowdturfing projects sell astroturfed workers to act out whatever disinformation their customer demands of them. Our analysis on the topic spanned three research questions: "What drives these ordinary people interested in crowdsourcing projects to join digital disinformation campaigns?" and "What are the dimensions in which an individual worker can perform digital disinformation acts?" To explore these questions, we brought in neutralization theory in order to better explore the phenomenon and imposed another question alongside it in: "To what extent can neutralization theory be used to justify participating in digital disinformation acts?" In this Master's Thesis we set out to study the possibilities with vignette-based interviews where we asked each of the participants to judge each 'excuse' of a neutralization technique for how likely they believed it to occur. From these discussions we determine the perceived plausibility of each neutralization technique through various angles and excuses. From this, we hypothesise that Bryant et al.'s additions of appeal to good character and victimisation, in particular, are too extreme or detached for digital disinformation contexts, failing to produce satisfactory results. In relation to this, we find that the more selfish an interviewee determined the motive of a participant to be, the better the overall plausibility of that particular excuse. It was generally thought that participation in digital disinformation campaigns was a purely selfish exercise. Finally, and maybe most importantly, through the discussions we find that many of the interviewees thought that these digital disinformation campaigns were both legal in the eyes of legislation and allowed by the social media platforms. This may stem from a simple lack of knowledge, which is likely a key reason as to why ordinary workers participate in crowdturfing. In addition, to this, nearly all of the interviewees spoke with contemptible pessimism about digital discourse and online media. What ramifications this has for digital spaces as a whole, remains to be seen.

Keywords: Digital Disinformation, Neutralization Theory, Crowdturfing

TIIVISTELMÄ

Kauppila, Santtu

Unraveling the Reasons of Participation in Crowdturfing Campaigns using Neutralization Theory

Jyväskylä: University of Jyväskylä, 2020, 70 pp.

Information Systems, Master's Thesis

Supervisor(s): Soliman, Wael

Disinformaation nähdään olevan ubiikki digitaalisissa ympäristöissä, etenkin nyt, uuden alan kehittyessä entisesti rehellisistä talkoistamis (crowdsourcing) hankkeista. Nämä crowdturfing kampanjat myyvät astroturfattuja työntekijöitä tekemään mitä tahansa disinformaatiota asiakas haluaakaan. Analyysimme aiheesta tehtiin kolmen tutkimuskysymyksen ympärille: "Miten tavalliset talkoistamisesta kiinnostuneet työntekijät liittyvät digitaalisiin disinformaatio kampanjoihin?" ja "Mitkä ovat ne ulottuvuudet missä tavallinen työntekijä voi suorittaa digitaalisen disinformaation tekoja?" Näitä kysymyksiä tutkiaksemme, toimme mukaan neutralisaatio teorian, jotta pystymme paremmin tutkimaan ilmiötä, ja kolmanne kysymyksemme: "Mihin pisteeseen asti neutralisaatio teoria pystyy selittämään digitaalisen disinformaation tekojen suorittamista?" Tässä Maisterin Teesissä tutkimme aihetta vignette pohjaisilla haastatteluilla, missä kysyimme kuinka uskottavana jokainen haastateltava piti jokaista neutralisaatio tekniikkaa. Kysymyksenä oli kuinka todennäköisenä he pitivät jokaisen neutralisaatio tekniikan käyttöä kontekstissa. Nämä haastattelut antavat meidän nähdä kuinka todennäköisinä ja realistisina kukin neutralisaatio tekniikka nähtiin. Testasimme jokaista neutralisaation tekniikkaa useasta näkökulmasta. Näistä keskusteluista voimme hypotesoida että Bryant et al.'n lisäykset 'hyvään luonteeseen vetoaminen' ja 'viktimitaatio' ovat liian kärjistettyjä ja ulkopuolisia digitaalisiin disinformaatio konteksteihin. Tähän liittyen, näemme myös kuinka mitä itsekkäämpänä haastateltava näki työläisen motiiviin, sitä realistisempänä tämä piti tätä motiivia. Haastateltavat näkivät digitaalisiin disinformaatio kampanjoihin liittymisen lähes täysin itsekkäänä toimintona. Lopuksi, ja ehkä tärkeimpänä, olemme hämmentyneitä siitä määrästä haastateltavista ketkä pitivät näitä kampanjoita laillisina ja sallittuina sosiaalisen media alustojen silmissä. Tämä saattaa johtua pelkästään tietämättömyydestä, mikä voi olla yksi avaintekijöistä miksi normaalit työntekijät liittyvät crowdturfaamiseen. Tämän lisäksi lähes kaikki haastateltavat puhuivat digitaalisesta kanssakäymisestä ja nettimediasta halveksuvalla pessimismillä. Mikä on tämän lopullinen seuraamus digitaaliselle maailmalle, jää nähtäväksi.

Keywords: Digitaalinen Disinformaatio, Neutralisaation Teoria, Crowdturfing

TABLE OF CONTENTS

ABSTRACT

TIIVISTELMÄ

1	INTRODUCTION.....	6
	1.1. Definitions	7
2	LITERATURE REVIEW.....	9
	2.1 Methods of Digital Disinformation	9
	2.1.1. Publishing	9
	2.1.2. Astroturfing	11
	2.1.3. Metrics Manipulation	11
	2.2 Crowdturfing	12
3	THEORETICAL FOUNDATION.....	14
	3.1 MEECES	15
	3.1.1. Money	15
	3.1.2. Ego	16
	3.1.3. Entertainment	16
	3.1.4. Cause	17
	3.1.5. Entrance	18
	3.1.6. Status	18
	3.2 Neutralization Theory	20
	3.3 Analysis Framework	23
4	METHOD.....	25
5	RESULTS.....	30
	5.1 Denial of Responsibility	32
	5.1.1. One more participant doesn't affect anything	32
	5.1.2. I didn't realize this wasn't allowed	33
	5.1.3. Summary on denial of responsibility	34
	5.2 Denial of Injury	35
	5.2.1. It's not a big deal	35
	5.2.2. It's digital so it doesn't matter	36
	5.2.3. Nobody who could be hurt, was hurt	36
	5.2.4. Summary on denial of injury	37
	5.3 Denial of the Victim	38
	5.3.1. What I did affected nobody	38
	5.3.2. Nobody believes what they see in the internet anyway	39
	5.3.3. Summary on denial of the victim	39

5.4	Condemnation of the Condemners	40	
5.4.1.	It's the consumer's fault for being fooled	40	
5.4.2.	Moral relativism	41	
5.4.3.	It's the rules/morality that's wrong	42	
5.4.4.	Summary on condemnation of the condemners	42	
5.5	Appealing to Higher Loyalties	43	
5.5.1.	I couldn't say 'no' to Jessica	43	
5.5.2.	This is the future of advertising	45	
5.5.3.	I thought it would be interesting	45	
5.5.4.	I thought it would be fun	47	
5.5.5.	I did it for the money	47	
5.5.6.	Summary on appealing to higher loyalties	48	
5.6	Appealing to Good Character	49	
5.6.1.	I can't believe you would doubt me, your friend	49	
5.6.2.	I'm entitled to profit from my social media presence	50	
5.6.3.	Summary on appealing to good character	51	
5.7	Victimisation	52	
5.7.1.	I did it for the wronged workers	52	
5.7.2.	The rival company is doing the same	53	
5.7.3.	Summary on victimisation	54	
6	DISCUSSION.....		55
6.1	Theoretical implications	55	
6.1.1.	Neutralization theory	55	
6.1.2.	MEECES	58	
6.2	Practical implications	61	
6.3	Future research directions	63	
6.4	Limitations	64	
7	CONCLUSIONS.....		66
8	REFERENCES.....		67
9	APPENDICES.....		73
	Appendix 1: English Vignette	73	
	Appendix 2: Finnish Vignette, Suomenkielinen vinjetti	75	
	Appendices 3-4: Results for Denial of Responsibility	77-78	
	Appendices 5-7: Results for Denial of Injury	78-79	
	Appendices 8-9: Results for Denial of the Victim	79-80	
	Appendices 10-12: Results for Condemnation of the Condemners	80-81	
	Appendices 13-17: Results for Appealing to Higher Loyalties	82-84	
	Appendices 18-19: Results for Appealing to Good Character	84-85	
	Appendices 20-21: Results for Victimisation	85-86	

1 INTRODUCTION

In the beginning of the internet, digital media promised to usher in a new age of information and connectivity. However, as years go by and we slowly come to realize the true nature of this beast that's been released on our civilizations and social discourse, that we may very well be in the new age of disinformation.

Using Edelman's trust barometer from 2019 we find ourselves in a clearly paradoxical situation. Edelman describes the difference between informed and uninformed people partly as their difference in consuming media and news. Then, when we see their individual trust levels of media and news, we find that informed individuals only trust media at 54% (2018) and 58% (2019) while the uninformed only have a trust of 44% (2018) and 47% (2019) with Forbes (Patricia Barnes, 2020) reporting that 33% have no trust in media with another 27% reporting that they only have little trust. Knight-Gallup (2018) finds similar results with 73% percent of Americans surveyed stating that the spread of disinformation is a major problem of current day media. How can we then have an informed populace of media consumers, defined by their trust and consumption of seemingly media filled with disinformation, and the uninformed mass, characterised by their lack of consuming media, which they regard as untrustworthy? The carts are before the horses. And despite all this, news consumption has increased by 22 pts between 2018 and 2019 (Edelman, 2019). One thing is clear at least: distrust in media is a profoundly prevalent issue many find themselves confounded by.

What got us into this situation? One theory is that social media platforms, in their quest to attain more and more profits, found a way to "Our Algorithms exploit the human brain's attraction to divisiveness." (Jeff Horwitz, 2020) The leaked Facebook internal presentation continued with "If left unchecked, more and more divisive content in an effort to gain user attention & increase time on the platform." Facebook, at the time, shelved the research and ignored its findings.

With social media platforms inadvertently having found this 'attraction

to divisiveness' and having built their platforms to facilitate and enhance, and profit from it, it was no wonder that illicit profiteers would rise up in an attempt to profit from it as well. First as a means of governments to subvert each other (Chen 2011, Pham 2013), then as a means of companies competing against each other (Chan 2012, BBC 2013). It's been reported (Wang et al., 2012) that 90% of all activity on the two of China's largest crowdsourcing sites, Zhubajie and Sandaha, were related to crowdturfing.

Here, then, is where we find our research questions:

What drives these ordinary people interested in crowdsourcing projects to join digital disinformation campaigns?

What are the dimensions in which an individual worker can perform digital disinformation acts?

To what extent can neutralization theory be used to justify participating in digital disinformation acts?

1.1 Definitions

The European Commission Communication on Tackling Online Disinformation (2018b) has them define digital disinformation as "verifiably false or misleading information that is created, presented and disseminated for economic gain or to intentionally deceive the public, and in any event to cause public harm." The European Commission specifies that the definition doesn't include, for example, mistakes, jokes, satire or parody, partisan slant, advertising or illegal content. This definition unfortunately leaves us on a very subjective basis when it comes to disinformation, seeing how information can become disinformation based on the intent of the publisher, the viewpoint of the consumer, and/or the legality of the content.

The EU High Level Expert Group (2018) similarly writes: "[digital disinformation] includes all forms of false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit". Again, we find the clauses 'for harm' and 'for profit', leading us to believe these to be necessary considerations when it comes to the description of digital disinformation motives.

Attaching intent to the classification is something Wardle & Derakshan (2017) also distinguish with three levels of harm: (1) the intent of the disinformation is harmless – for example a joke or otherwise, that ends up being taken out of context, (2) the intent of the disinformation is to cause harm, (3) malinformation, where honest information is used in deceptive ways – "lies, damned lies, and statistics". Within this classification, the common concept of 'clickbaiting' fall mostly under classification 1, as the intent of the disinformation is not to produce harm but generate revenue. As such, solely

harmful disinformation publishing is difficult to find and even more difficult to prove as digital information, notoriously, obscures intent and so long as money is the primary objective, one can't claim it as harmful.

Similarly, the final category is difficult to classify items into as, for example as seen with 2016 Democratic National Committee email leaks published by Wikileaks in 2016, what's seen as harmful for one side is seen as necessary publication of truth to the other. Thus, again, we find that it is important to attach only objective articles of disinformation into these categories, and not subjective opinions based on the nature of the content. The full extent of these motives is proposed in a later chapter.

However, having to rely on the integrity of experience on the behalf of the consumer is bound to lead to false positives in the identification of digital disinformation. Sites relying on users to report disinformation may find the user experience to be unreliable in identifying news from advertising. Amazeen & Wojdysky (2018) find that less than 1 in 4 adults can tell a native advertisement from published news content. This number is likely to be even less for content, which is purposefully created to seem as though it was natural honest content, such as in the case of digital disinformation campaigns.

Because of these factors, we feel it is important to not rely on subjective interpretations, or at least include more categories to the classification of digital disinformation than simply monetary and malicious. In fact, many categories of digital disinformation rely solely on the good will and understanding of the consumer, such as jokes and political slanting. One can easily argue that so long as the consumer doesn't realize the joke is a joke (i.e. Eats the onion) the information within it, to them, is producing a disinformative effect. Of course, not much can be done if the consumer willingly ignores the 'parody' label many social media platforms now utilize (Jordan Valinski, 2014).

These issues are similarly raised by Don Fallis (2015) where he finds that many definitions of disinformation either exclude or include topics and acts that are not disinformation. For the purposes of this paper and furthermore, we propose the use of his definition, whereby disinformation is misleading information that has the function of misleading its readers, and digital disinformation being disinformation acted in digital space or through digital means.

2 LITERATURE REVIEW

In the literature review section we intend to answer the question of 'What are the digital disinformation acts an individual worker is capable of?' We do this by reviewing the existing literature.

2.1 Methods of Digital Disinformation

As the field is broad and each of the methods require an inspection on themselves, the contents of this literature review will be divided into sections of the classifications of methods presented within this paper.

EU-Joint research centre (2018) attempted to define the terms of fake news and false news, and other types of digital disinformation and found that there was no consensus on the terminology or spread. However, from their descriptions we can extract some of the methods these digital disinformations are distributed in. When looked together with NATO StratCom COE (Sebastian Bay et al. 2018) we can hypothesize an initial grouping for these methods of digital disinformation as publishing, astroturfing, and metrics manipulation.

2.1.1 Publishing

Publishing is the act of presenting a topic that is purposefully falsified, in either content or contextualization (Sebastian Bay et al. 2018. p. 14-15). All facets of fake news are included under disinformation publishing (misleading headlines, disinformative content, misrepresentation, etc.), but acts can be classified into this category even if they weren't performed under a name and brand of a news publisher, so long as the perpetrator is the one bringing forth the topic.

In some cases publishing acts create a discussion forum in of itself that's for the purposes of discussing the published material. The disinformation act of

publishing then can be further amplified with astroturfing, which is explored in the following chapter.

The prevalence of this form of digital disinformation has surprisingly little research. What doesn't help is the social media platform's seeming unwillingness to cooperate with third party entities. For example, in Information Operations and Facebook (J. Weedon et al., 2017), while Facebook claimed that malicious actors accounted to less than 0.1% of civic traffic, no actual data was provided to support this claim. Of course, accurate numbers will be hard to find, even by the platforms themselves, seeing how the purpose of digital disinformation is to not be visible to the general public, lest nobody be influenced by it.

The eternal arms race of social media and the actors attempting to manipulate its users for their own malicious purposes is mostly fought outside the public eye, making any estimation by people uninitiated a wild guess at best and estimations by experts a shot in the dark. Allcott et al., (2017) estimated that social media users saw between one to three stories from known fake sources during the last month leading up to the 2016 election.

The impact of publishing based disinformation is an equally sparsely studied subject. While it is undeniable that malicious actors are present in social media, it is unclear as to how effective they are in their attempts. As previously alluded to, the existence of malicious influencers is not proof of their impact in any given topic. The same concern was found by J. Lazer et al., (2018, p. 1095): *"However, knowing how many individuals encountered or shared a piece of fake news is not the same as knowing how many people read or were affected by it."* In fact, the effect might very well be minimal. A study on the effect of political advertising (Joshua L. Kalla & David E. Broockman, 2017) makes the argument that the effect may very well be zero or at most extremely minimal. However, Emily Thorson (2016) finds that even a marginal exposure to negative political disinformation leads to persisting effects, even if this disinformation is later corrected. Continuing, Soroush Vosoughi et al. (2018) find that lies spread faster than truth, with political disinformation spreading much faster than any other category, and that this spread can be mostly attributed to human distributors. They theorize that this is because the false information is seen as novel and that this surprise leads to more people spreading the fake news as opposed to truthful information or correction.

It is important to note that most of the available studies in regards to disinformation are done in the field of politics. As we will soon see, this is only a single facet of a single dimension we've identified for digital disinformation motives.

2.1.2 Astroturfing

Astroturfing is the act of masquerading with the general populace to seemingly present honest and legitimate grassroots opinions and thoughts, through the usage of written or image material (Adam Bienkov, 2012). Examples of astroturfing are black propaganda and paid reviews. Commenting on one's own work, or any work you are personally connected to, without disclosing the connection is also an attempt at disinformation. Of course, not every example given is as nefarious, but the degree of harm is not relevant for the purposes of identifying the method itself.

The effects of astroturfing on consumer behaviour are well recorded through mere-exposure effect if nothing else. Goetzinger (1968) found that simply exposing people to a topic, even a nonsensical one, like a trash bag in his experiments, enough times, the perception of that object would naturally steer towards positivity and acceptance. This combines well with Zhuang's findings (2014), where he remarks that while online products with manipulated reviews are more suspicious than not, even if the consumer doesn't outright know of this manipulation, and that the existence of this suspicion lowers their buying intent, there is no difference in the actual concrete purchase intent between manipulated and legitimate products of the same rating. The existence of manipulation does not affect consumer's intent on purchasing, leading us to believe that astroturfing is a very capable tool for manipulating consumer behaviour. This is further reinforced by Zajonc's later experiment (2001) where he concluded that the exposure doesn't need to be conscious for it to be effective. Combined with Zhuang's findings, we are led to believe that while consciously the consumers were deterred by the suspicion, the unconscious effects of mere-exposure effect made this resistance ineffective in avoiding deceptive purchases.

2.1.3 Metrics manipulation

Metrics manipulations are attempts at inflating both the visible and hidden metrics found in various social sites. These include inflating likes, views, page visitors, shares, and the number of reviews (Sebastian Bay et al. 2018, p. 9), among others, such as algorithmic manipulation (Bradshaw, S. 2019). Jenn Chen (2020) explains the three most important metrics to track in social media to be 1) likes, comments, retweets, etc., or the visible metrics, 2) engagement rate, which is tracking the ratios of the quality of engagement of to a post (commenting is higher quality than sharing which is higher quality than liking, which is higher quality than impression, etc.) and 3) account mentions, which are how much your posts and account are mentioned by third parties in their discussions.

While some of the activities can be classified within astroturfing, how we

define a clear division between the two is the usage of writing. Metric manipulation activities are purely numerical increases to statistics, while astroturfing activities always include written text. As such, while visible metrics do include commenting, so long as this commenting is in any way meaningful, i.e. not simply spammed nonsense, it will be listed under astroturfing, rather than metrics manipulation. Often this category is the one involving bots, but platforms are very keen in purging numbers impacted by manufactured means, making this a very inefficient way of creating traffic.

2.2 Crowdturfing

Since our discussion is based on digital disinformation campaigns, it is pertinent we discuss the commercialized version of digital disinformation.

Crowdturfing is a field of commerce, where disinformation intents are delegated to a general working populace who willingly participate in the process, receiving monetary incentives (crowdsourced astroturfing). Mark Leiser (2016) discusses the topic under the synonym of cyberturfing, remarking on its effectiveness and spread, noting that 68% of users regard online reviews of hotels and travel operators as more important than other sources of information. With this calculation alone, he estimates that, going by the Competition and Market Authority's (2017) numbers, that an estimated £23 billion could be affected by online reviews alone.

Rinta-Kahila & Soliman (2017) explain the network of participants to have five dimensions.

- 1) Customer: The initiator with a goal, which they hope to achieve through digital disinformation means.
- 2) Agent: A paid recipient tasked with finding the workers who can complete the goal. Can either be an individual company with their own workers or an ISP site, who find their workers through the internet.
- 3) Worker: The eventual actor of the digital disinformation act. They are tasked to perform operations under the guidance of the Agents in order to complete the goal of the Customer. These workers are often called social bots due to their task instructions effectively functioning as a sort of programming for their behaviour, but this term, as it stands, has multiple meanings.
- 4) Target platform: The social media site where the digital disinformation act takes place. Unaware of the digital

disinformation act taking place.

- 5) Common user: Sincere internet users who are made to believe the workers are similarly sincere users, leading them to be swayed by the disinformation as they think it to be an honest opinion by a similarly honest user.

In this network, the customer, agent and worker are the perpetrators while the platform and users are the unaware victims of their disinformation acts.

Crowdturfing campaigns can operate through legitimate crowdsourcing sites, masquerading as legitimate marketing operations (Wang et al., 2012).

Workers participating in these campaigns can find themselves utilizing any of the previous three methods: publishing, astroturfing and/or metrics manipulations. These workers don't often utilize crowdurfing as a method themselves, which means that while crowdurfing is a prominent method of digital disinformation, it is not an activity workers participating in digital disinformation utilize themselves; the workers are simply the tools of crowdurfing, initiated by the customer.

It is often the case, where the workers are replaced with true bots - an algorithm that mimics human behaviour to the best ability of their programmer. The problem many of the Agents find, however, is that Platform operators are very well versed in detecting and removing bots from their platform (Ratkiewicz et al. 2011), leading to the usage of human workers in the crowdurfing campaigns as illustrated by Rinta-Kahila & Soliman. The winning teams of the DARPA twitter bot competition came to the same conclusion, in that purely machine based algorithms were far too easily detected to be useful (Subrahmanian et al. 2016). The rate at which bots and humans converse with each other is disproportionate as well. While both quote each other with similar rates, bots are far more likely to converse with other bots as are humans to humans (A. Bessi & E. Ferrara 2016), further devaluing their influence. As such, bots, while cheap and numerous to use, have little impact on the actual discourse conducted and are very unlikely to produce positive results.

3 THEORETICAL FOUNDATION

In order to determine the reasons for why individual workers participate in digital disinformation campaigns, we need to discuss the possible motives as to why disinformation is spread in the first place.

For the purposes of clarifying and simplifying the motives, we are looking at the primary motive, as, unsurprisingly, the true purpose of disinformation acts can be hard to determine. Covert operations, such as Russia's involvement in the MH17 plane crash (Golovchenko et al., 2018), strive to obfuscate the true purpose of the campaign, leading any surface level analytical deduction astray. For the time being, we do not concern ourselves with this point of contention as it becomes impossible to define anything if we suspect every finding we make.

Similarly, it is important to make the distinction between categories and not distill the perceived motive too far. One can argue that many motives could in the end, be distilled into 'wish for more power'. As a category, however, this lacks nuance and is essentially useless. Attaining power over something is clearly a motive for certain digital disinformation campaigns and in these cases we will list them as such, but should the motive on the surface appear as political, monetary or otherwise, it will be more beneficial and informative to list the motive as such.

3.1 MEECES

To explore these motives, we're utilizing HoneyNet and Max Kilger's MEECES motivators for hackers (2004). These are: money, ego, entertainment, cause, entrance and status. While not perfect as a framework, it should provide us a base to discuss our findings and the possibility of adapting it for individual digital disinformation workers.

The key problem with MEECES is that it is meant for the individuals alone. As such, it is, at places, inapplicable to companies or groups initiating digital disinformation campaigns. As this study doesn't stretch beyond the individual worker, it is important to note that the framework used here may not be applicable to third party initiators. If there was a study based on them, we may need to utilize a different list of motives.

Later, we'll find that many of the categories of MEECES fit very well into neutralization theory. We'll discuss this topic more under the neutralization theory section.

Furthermore, Soliman et al. (2019) found that selfish intrinsic and extrinsic motivations drive workers into joining crowdsourcing campaigns and projects. In the adoption phase, these motivations are commonly monetary and/or curiosity. Through usage and participation in these campaigns, the intrinsic motivation diminishes, eventually leaving only extrinsic motivations. Finally, demotivational factors remove all motivations, reducing the workers to a state of amotivation and driving them out of the campaign. Thus, it is important to view each motive as possibly intrinsic or extrinsic motivator, or even possibly both or either.

3.1.1 Money

The simplest category within motives is, as mentioned before, money or monetary reasons. These are disinformation actions taken in the attempt to directly gain monetary benefits. As mentioned before, if an action is not listed within this category, that doesn't mean a motive among others can't be monetary. In fact, most actions humans take overall are at some level motivated by power or money, and some would even equate the two. This category is dedicated to actions specifically to gain money, usually by swindling the recipient. These include manufactured product reviews, phishing and other similar scams. It is important to note that all tertiary disinformation operatives, for example crowdturfing agents, are more than likely motivated by the payments received (Soleiman et al., 2019).

3.1.2 Ego

Ego is the thought on the behalf of the worker that they are beyond the rules that would lead their actions to be punished in the first place. Workers motivated by ego want to see themselves capable of rising to the challenge provided by the third party customer and the campaign they've set up. Max Kilger further explains that this may derive from the desire to see established functions used in a completely distorted means.

Ego as a whole is an intrinsic motivator, in that it depends solely on the personality of the worker and whether they believe themselves beyond the rules and/or interested in breaking established systems.

Another way ego can manifest as a motive is curiosity. Soliman et al. (2019) explained this as the interest in the process and possibilities of crowdsourcing. We see no reason why such curiosity couldn't manifest in digital disinformation and, especially, crowdturfing contexts.

Thus, while curiosity can come in many forms it mostly boils down to an honest attempt at trying to understand the mechanisms and methods of disinformation through personal participation. There are many ways this knowledge can eventually be used for, with common usages including research, reporting, countermeasures, and/or development.

As some of the possible curiosity motivations are slanted towards honest information gathering, it is easy to think that people participating with this motive believe themselves to be excused for their participation. Meaning, that their personal curiosity outweighs the harm they cause through their participation. This is also why this subset of motives is set under ego, rather than entertainment, which we will discuss next.

3.1.3 Entertainment

Entertainment motives find their reasoning for disinformation from the enjoyment of seeing the reactions to the disinformation act itself. Max Kilger described this with "You'll still see a hacker break into a system, trash it up and sit back, and watch the system administrator scurry around trying to save it." While this description is for hackers, the original framework of MEECES, we see no reason why or how this couldn't be applied to disinformation spreaders as well.

What makes the attempt to categorize motives is often nebulous exactly because certain elements of disinformation actions can be enjoyed by a certain crowd. This means that even if a disinformation topic or discussion was being widely spread, it may very well be the case that it is spread by like minded people who, once they get the joke, they begin spreading it and join in on the 'campaign'. Of course, this doesn't remove the possibility that the campaign

was originally initiated by a third party customer taking advantage of this possibility of true grassroots spread. Of course, this doesn't mean that the initiation of this 'joke' or campaign wasn't made by a third party customer, but its spread may be due to people unaffiliated with this initiation. The entertainment stems from simply knowing they're spreading disinformation and that some people may believe it or from seeing the responses and reactions to the blatant disinformation they're spreading.

On the other hand, entertainment motives may be the reason why a worker joins in on a disinformation campaign in the first place. If the perception is that the campaign is fun or the activities taken within it are interesting or if the worker is curious about digital disinformation, they may join a campaign as well. This may allow digital disinformation campaigns to gamify their campaign in such a way that it is more fun to act in.

Motives are classified as entertainment when the worker's goal is simply to provide entertainment to themselves and/or others, who are 'in the joke', from the gullible and/or irritated people believing whatever they're spreading.

3.1.4 Cause

Cause as a motivator is when the worker believes they are working for a higher cause and that this cause is worthwhile enough that it justifies the use of disinformation as a means of achieving it. This can further manifest in two different ways, either as trying to positively enhance your own side's optics or in the attempts to bring down all the opposing forces.

The attempt to change somebody's opinion by positive reinforcement can either be political, brand related, scientific, medical or nutritional, among other things. This classification is defined by the disinformation agent's desire to change or reinforce somebody's opinion on a topic towards the better. When this act is classified as disinformation, is when it is accomplished by misleading and deceptive means, such as falsification or misrepresentation. Again, it must be stated that more often than not, the primary motivation will be monetary in the end, but should the intent on the surface to simply sway individual or public opinion, it can be classified within this category as long as there's no direct pathway or goal to monetization of opinions.

The opposite of this category is when the worker is motivated to dissuade and hurt the opinion and perception of certain topics. This is achieved by providing deceitful or manufactured information with a negative spin for the purposes of deceiving the reader into disliking the topic. Contrary to common belief, this is much harder than is often laid out in the media, as people are notoriously convinced of their own beliefs and unwilling to change opinions (Nikolov et al., 2015). Regardless of this safety mechanism of the human psyche, these negative efforts are very common and can often be

attributed to individual actors as, unlike positively swaying opinions, these actions only need an 'enemy' to push down, rather than a clear 'ally' to prop up, making it much more difficult to pin down a clear 'culprit' of the campaign, not that the existence of a campaign is evidence of the 'culprit's' involvement in it (i.e. Black propaganda).

While these two topics fall under the same category in MEECES, it may be necessary to consider them as separate categories for future purposes. While they are, in essence, functioning towards the same goal of 'promoting a cause', there are vastly different ways of achieving either but not both. Similarly, there may be large differences in the neutralisations used and accepted for either category, and the personalities of workers that are perceived to use either of the above categories.

Cause can be classified as intrinsic or extrinsic, since, while the cause definitely is always an extrinsic goal, it is usually through intrinsic belief into this cause that drives people into performing malicious acts in service of it, as the goal itself can't force anybody to take these actions.

3.1.5 Entrance

Entrance as a motivation is the bragging rights one can pronounce from having successfully completed disinformation acts. As the MEECES theory originated from hacker community, this motivator was when a hack was completed in order to gain entrance to certain groups or forums, where these topics were discussed and planned.

In the context of digital disinformation this works in a similar fashion, although we haven't been able to locate any forums that require one to produce disinformation, individually or as part of a third party campaign, in order to enter, there are forums where such disinformation acts are planned and discussed at the very least, for example 4chan (Storyful.com, 2018).

While as a category this is quite weak in application to digital disinformation itself or to this study, its existence as a possible motivator can't be denied.

3.1.6 Status

Status is the other side of entrance, where, if entrance was the initiation, status is the motivator that has the perpetrators attempting more prestigious and difficult goals and objectives. Status is then derived from the successful accomplishments and bragging rights to those achievements.

How this manifests as a disinformation agent is when these forums and portals plan for more and more elaborate disinformation campaigns, seeing

what they can get away with and what they can get people to believe. For example, when 4chan users convinced iPhone users that their phones were chargeable by microwaves (Lillian Radulova, 2014), then that their iPhones were waterproof (Fernando Alfonso III, 2020), and that they could drill a microphone jack into their phones (Ellie Flynn, 2016).

However, since disinformation acts such as these are illegal, it is difficult or impossible to properly attain status from one's achievements in these fields. In fact, it may be the case, since entrance isn't required or demanded for such sites, both of these motives should be put under entertainment or ego, rather than their own category. Regardless, as neither entrance nor status are well applicable to disinformation, they do not make an appearance during the study nor in any discussion related to it.

Another way status can manifest is in order to achieve personal recognition by utilizing disinformation to promote oneself. This is either done by astroturfing as positive responses to yourself by yourself or by utilizing crowdturfing services. Soliman & Rinta-Kahila (2018) found that, it may very well be the case that the individual users purchasing this service are being fooled themselves with marketing phrases such as "*everybody does it*" and "*kickstart your career*". In addition to this claim of 'leveling the playing field', ordinary vanity is also included for people who simply don't care about rationalizing their perceived misdeeds.

Overall, MEECES provides us with a good foundational base to examine digital disinformation motives for. While the final two motivations, entrance and status, are, at this point, a dubious inclusion, it is clear that, if it was possible to create a community for disinformation or around disinformation, that these categories would exist in some capacity at least. While status could be revised into what Soliman & Rinta-Kahila (2018) found, it isn't quite the original intention of MEECES's status. Thus, we will leave this discussion for another paper, since these motivations do not quite fit into the constraints of the study we are conducting and, thus, are hard to produce data for or against.

3.2 Neutralization Theory

Neutralization theory was originally conceived by Matza and Sykes in 1957. It describes the methods criminals rationalize their actions to themselves and others. It is based on four observations.

1. There are clearly people and/or groups a person is not willing to victimize.
2. Perpetrators express guilt over their illegal actions.
3. Perpetrators are not immune to conformity.
4. Perpetrators often respect and admire law-abiding individuals.

From these observations Matza and Sykes believed that there was something that was binding these perpetrators to law-abiding actions or, at least, making it so that they themselves perceived that their actions were justified. From this observation, Matza and Sykes theorized the first five neutralization techniques.

- Denial of responsibility is to pass the blame onto another cause. Either the criminal was forced to commit the act or their situation didn't allow an alternative. A common phrase associated with denial of responsibility is the words "It is not my fault."
- Denial of injury is the attempt to downplay the significance of the act. That, in fact, their activity did no perceivable harm and was innocuous. A common phrase associated with it: "It's not a big deal."
- Denial of the victim is the attempt to blame the victim's character as the cause for their actions, claiming they deserved whatever was done to them. A common phrase: "They had it coming."
- Condemnation of the condemners is initially classified as the claim that the victim is in fact the one who is spiteful and in the wrong for pursuing action. How this manifests is as the combination of the previous two condemnations, that 1) there was no victim and 2) there was no harm, and it is only their malicious intent that claims there is both. In short, condemnation of the condemners is to deny the injury and victim while attaching a malicious motive upon the condemner.
- Appealing to higher loyalties is a claim that their action was motivated by a greater good which supersedes the morality/law of the current time. This can be anything from helping a friend to fighting a perceived evil. A common phrase: "It's the rules that are wrong."

While these five were the initial neutralization techniques proposed by Matza and Sykes, the theory has been added many times over the years since, but, instead of testing out all of them, we will be adding only two of them which were proposed by Bryant et al. (2018). These two are appealing to good character and victimisation.

- Appealing to good character is an attempt to show the impossibility of the crime itself as they themselves are of such admirable character. Another method this can manifest is a sort of claim to a counterbalance where their past good deeds allowed them to act this way today. A common phrase: "It was just one time."
- Victimisation is an attempt to justify their actions because of a previous injustice towards them or somebody who they relate to in any way (religious, ideological, racial, nationality, ethnical, etc.). A sort of opposite to the previous technique, except in that victimisation can be claimed on the behalf of somebody else or after a long period of time. A common phrase: "I did it for him/her/them."

The reason why we are testing Bryant et al.'s neutralization techniques alongside Matza & Sykes' is due to their recency. We felt that we had enough room within the interviews to include additional questions and the thought Bryant et al.'s 2018 suggestions of appealing to good character and victimisation could use some additional testing. In addition to this, victimisation matches well with cause as a motive from MEECES, giving us a very distinct line of testing the validity of MEECES as a motivational framework.

While criticism has been raised, especially in regards to the point of time when the neutralization occurs, as neutralization occurring after the criminal activity can simply be an excuse come up on the spot and the apparent guilt the theory was built upon is, in fact, a ploy to garner sympathy from the accusers because they were caught doing the criminal activity. Regardless, this doesn't change the fact that neutralization was attempted. Sincerity of the neutralization, in our opinion, is outside the scope of this study, only the classification of the attempted neutralization technique.

Similarly, findings by Hindelang (1970, 1974) and Agnew (1994), whose studies in juvenile crimes found problems in the premise of neutralization theory. The assumption of good and wholesome character functions for the vast majority of individuals, but people growing in unfortunate circumstances may result in them growing a character that idolizes or is indifferent to violent and unlawful behaviour. One could make the assumption that people with criminal behaviour would hold the vast majority of these sorts of people. In the cases of Hindelang's and Agnew's studies, the juvenile criminals held no remorse towards their actions and, as such, had no need to neutralize them in any way.

Agnew, however, did find that those with the most approving or indifferent attitudes to violence grew to become that way after their acts, leading him to believe some neutralization had taken place at some point in their lives, but it became second nature after subsequent exposures to violence. Jim Mitchell & Richard A. Dodder (1983) found statistically significant correlation between neutralization techniques and delinquency, while Austin (1977) and Minor (1981) found weak relationships between violence and neutralization techniques.

The material on neutralization theory at the moment leads one to believe that it can, at least, be used to deduce the reasonings behind small scale disinformation acts, ones more closer to delinquency than outright subversion between countries. Neutralization theory has already been used to produce statistically significant results when it comes to delinquency (Mitch & Dodder 1983, Minor 1988).

Neutralization theory has already been utilized in digital contexts to relative success. Hinduja (2007) and Wilhelm (2020) both found that neutralization theory could be used to explain the cognitive processes of people's participation in online piracy. Higgins (2008) when studying music piracy further elaborated that "victimless" crimes having higher potential for neutralized crimes could be the result of ignorance and lack of education, postulating that the solution to piracy may be there. Marcum et al. (2011) add to this by studying piracy through neutralization theory, self-control theory and differential association theory, concluding that differential association theory together with neutralization theory could explain the behaviour. While we won't explore these two theories any further, utilizing them for digital disinformation contexts could provide further insight.

Much of the cross-section of digital contexts and neutralization theory relates to digital piracy, which, at this point, is a well supported application of neutralization theory. In regards to other digital applications, Zhang & Leidner (2018) studied cyberbullying with neutralization theory with the premise that the perception of a) invisibility, b) asynchronicity, c) anonymity and d) publicity of digital spaces contribute to the lower threshold of cyberbullying to bullying. We believe these four aspects to be a factor in digital disinformation contexts as well. Combining with Higgin's (2008) find where "victimless" crimes were more easily neutralized, digital disinformation may be the perfect storm of plausible deniability for participating workers.

The material on neutralization theory's applications in regards to large scale events, such as crowdturfing campaigns, is still to be found and this study will certainly not answer that question definitively as the vignettes presented are all hand crafted and the people answering are only theorizing and deducing. Despite this, we believe there to be adequate justifications to utilizing neutralization theory in digital contexts, especially in regards to "victimless" crimes, such as targetless disinformation.

3.3 Analysis Framework

To recap everything we've examined so far: Rinta-Kahila & Soliman (2017) laid out a framework, in which crowdturfing activities operate in. Here, a customer initiates a campaign intent and creates a plan, which they forward to agents, who are employees within their company or ISP sites, such as fiver, that manage such activities. Then, workers utilizing these sites, mostly for legitimate reasons, find these campaigns, usually framed and explained in such a way that makes them seem legitimate, joining them for one reason or another. These workers then act according to the guidelines set by the campaign and gather monetary or other rewards for their actions. These actions by individual workers are limited to publishing content, astroturfing as regular honest users or simple metrics manipulation. Crowdturfing as a method is limited to customers and, while not impossible, it is improbable for workers to utilize crowdturfing themselves to actualize their own role as a worker of another crowdturfing campaign.

To actually study the participation of these workers, we find ourselves with a problem. Seeing how these activities are criminal, it is unreasonable to locate such participants for this study. In addition, any attempt at directly asking interviewees about whether they would participate in criminal activities goes against research ethics as well as being unlikely to produce meaningful results, seeing how they are devoid of proper contexts that drive workers into participation. Similarly, since the topic is disinformation, which is a form of "victimless" covert crime, studying the phenomenon through case studies is nigh impossible.

Because of this degree of criminality a different angle is necessary. We believe this angle to be neutralization theory, which doesn't make an attempt to find the true motive of the crime, but, instead, focuses on how the perpetrators attempt to justify their crimes to a third party. This way, we don't need to figure out true motives for the workers and we can, rather, gather information from reactions to these justifications.

However, since, neutralization theory makes no claims about the true motives of the perpetrators, we believe it to be necessary to fit the results in some manner of motivational framework to better examine the results. We aren't attempting to create one based on our findings since we aren't receiving any direct information from first-hand experiences and rely mostly on the perceptions of the motives. Here, we use MEECES to theorize the reasons for participation of these participating workers. This allows us to examine our results in a theoretical framework and better evaluate their validity and applicability. We will also be able to see whether the two are usable in conjunction as well as determine whether MEECES, as a whole, is viable as a motivational framework in digital disinformation contexts, although this result

will not be conclusive.

Our framework then for this study is the network of crowdturfing participants (Rinta-Kahila & Soleiman's, 2017), where the worker performs methods of digital disinformation, explained by NATO StratCom COE (2018) and EU-joint research center (2018) in their reports on the topic. To explore the motives of these individuals we are utilizing MEECES (Max Kilger, 2004) as a basis and utilizing neutralization theory (Matza & Sykes 1957, Bryant et al., 2018) to properly explore them for a safe and ethical research context.

4 METHOD

The general framework of a digital disinformation act, as explained by Rinta-Kahila & Soliman (2017), involves the cooperation of a customer, agent and workers, who act their campaigns on the platform upon the users utilizing it. While some deviations are possible, such as internal disinformation divisions within organizations or countries, we believe this model to be a good fit as the general all-purpose model. For the purposes of our vignettes, we will be focusing on the workers as IT students are a good fit for performing these acts, making them the perfect category to base a study on.

From this framework we have two reasonable options when it comes to studying digital disinformation acts, these being: the user or the worker. These are both individuals and are both accessible in large quantities within the university; the users directly and workers plausibly, seeing how university students were the primary demographic of workers, for example in Samsung vs. HTC (BBC, 2013). While there are topics that could be studied from the perspective of the user, since they don't have any direct contact with the disinformation act itself, their viewpoint is very limited to their own experiences. As such, the viewpoint for this study was decided to be the worker in Rinta-Kahila & Soliman's framework.

Seeing how it was decided that we're exploring this topic from the point of view of a petty criminal, it was key to consider methodology. As explored in the 3.3 analysis framework, finding people who've participated in crowdturfing campaigns was unreasonable. Even if such an individual would be found, they would also have to confess to the crime they've committed, and, despite vowing to strictly academic professionalism, I don't think this was a viable option. Similar results would be found from surveys and if the questionnaire would be expanded to large populations of the internet, while this would provide a degree of anonymity required for such questions, there would be no way to confirm any of the data received by this way. Similarly, as the topic of interest is covert digital disinformation, studying it through case studies would be reports and law cases at best between the customers and the victims of these campaigns. Observation is even more impossible.

To solve this problem, we resort to vignettes methodology. Barter and Renolds (1999) describe there to be three main purposes of vignettes: 1) explore actions in context, 2) clarify judgements, and 3) allow for less personal exploration of dangerous and sensitive topics. The purposes one and three especially provide us with good solutions to the problems explained before. Vignettes are a tool of qualitative research where the participants are provided a story describing actions in a context, after which these participants are questioned about their impressions on the story. While there exists a multitude of ways to present vignettes ranging from interactive hyper-linked presentations to videos (Jenkins et al., 2010), we determined that there was no need to complicate this process. With this in mind, the vignette we crafted was a text document, which the interviewees read before the interviews. Preferably, this vignette would be read right before the interview itself, to maintain first, fresh impressions, but in two cases this wasn't achieved. Face-to-face interviews were decided over a survey based on the same vignettes as simple yes-or-no answers would prove unsatisfactory and, we believe, that additional information can be stemmed from the discussions that arose around these interviews.

In total, 9 people were interviewed for this study. They are listed in the following table in the order their interviews were conducted in.

Interviewee	Age	Language	Sex	Field of Study	Interview Length
F1	25	English	Female	Physics	45 minutes
F2	22	English	Female	Physics	34 minutes
F3	19	Finnish	Female	Literature	22 minuets
M1	28	Finnish	Male	Business	19 minutes
M2	29	Finnish	Male	IS	22 minutes
M3	28	Finnish	Male	Physics	30 minutes
M4	28	Finnish	Male	IS	28 minutes
F4	26	Finnish	Female	Education	18 minutes
M5	22	English	Male	IT	23 minutes

Table 1 Interviewees and their information.

There were many attempts to find interviewees unaffiliated to the researcher, but no truly such individuals were found for this study. Many courses within the university of Jyväskylä offered cooperation, but since participation was voluntary, no students sent a single notification of interest. This was further exacerbated due to covid-19 pandemic closing down many avenues of finding

participants through venues, such as the physical university location. As such, the participants are either friends and acquaintances of the researcher (1F, 4F, 1M, 2M, 3M, 4M) or friends or acquaintances of these people (2F, 3F, 5M). At most, the degree of separation is one person.

While the interviewees are not completely unaffiliated, they are still of the likely age group, which the workers are, as well as reporting that they used social media regularly. With this in mind, we believe the selection of interviewees, while not perfect, to be more than capable of fulfilling our research purposes.

For two of the interviewees (3M, 3F) the data is slightly impacted due to time between first reading of the vignette and the interview.. These are due to miscommunication where 3M didn't realize the interview required a functioning microphone and due to a sudden event for 3F. Both of these events lead to a rescheduling, with 3M being interviewed 5 days from the initial reading and 3F after 3 hours.

In addition, some interviews were conducted purely through voice and had no physical face-to-face contact. This was done to adhere to covid-19 guidelines or distance. 1F, 1M, 3M were done face-to-face, while the rest were through voice. What eventually are the effects of these occurrences is unclear, but their existence is noted regardless.

The vignette is formulated around the Samsung vs. HTC online smear campaign in 2013 (BBC, 2013) from the perspective of a typical participant in the worker-role explained in Rinta-Kahila & Soliman (2017) dimensions of crowdturfing, an IT-student. The sex of the student was made to match the interviewee in an attempt to increase their empathy towards them. This was achieved by changing the name of the worker character: Paul in the case of a male interviewee and Jessica in the case of a female interviewee. This character is introduced to the campaign by a friend of theirs, named whichever was the other name left. Henceforth, Paul refers to the worker character and Jessica refers to this friend who initially introduced them to this campaign.

The Vignette follows Paul through a crowdturfing campaign while he performs a multitude of activities listed in 2.1 Methods of Digital Disinformation. At the same time, the vignette gives plausibility to each of the motives of MEECES, to which we hope to utilize the seven neutralization acts in the interview portion. The story culminates in Paul being forced to explain themselves to another unnamed friend of theirs who disagrees with what they have done. For the purposes of the vignettes, Paul's motive is left mostly nebulous and up for interpretation, as what we are interested in are how the workers neutralize their actions, rather than what their motives were, as we have no way of accurately determining that and plainly stating it in Paul's case would bias the results to whatever we wanted from the vignette. We believe that by alluding to many motives, the interviewees can consider each, in turn, whenever it becomes relevant to the neutralization technique at hand.

The vignettes can be found in the appendices section. Appendix 1 for the English version and appendix 2 for the Finnish version.

Once the interviewee has read the vignette they are first asked as to what would be the excuse given by the worker character for their acts, after which a discussion is held, wherein they are asked for the plausibility of each neutralization technique. If the interviewee had already come up with a neutralization technique as their primary reason, it is not asked a second time during the interview portion. The order in which the questions were asked wasn't strict and, instead, questions were asked as they became topical, especially if the interviewee raised a comment or question that indicated something of the sort.

Example dialogue from the interviews:

Interviewer: So, what do you think Jessica (Paul's character as the interviewed is female) would use as her excuse?

F2: Why should she make an excuse? I mean, they're friends with this person, right?

Well, this friend is ready to cut ties because of it, so she has to reason it somehow.

Umm... Well, I mean, if I was the friend of this Jessica, I wouldn't mind it at all. It was her job and she did it and that's that.

But you're not the friend. You're the... fly on the wall seeing this event happen.

Oh, alright... Okay... In that case... I guess Jessica would say that it was her job and she didn't think anything was wrong with it.

Neutralization techniques were brought up through the guiding questions that are listed alongside each technique in section 4. These are also the ways these comments were identified from these interviews. The answer to the initial question, as seen above, were marked as a plausible acceptance of that neutralization. In follow-up discussions, whenever a question was brought up, the interviewees would be prodded long enough that they gave a yes or no answer to how plausible they thought of Paul using that particular neutralization theory, simplified as excuses for the interviewees.

In addition, if an interviewee brought up an interesting excuse that wasn't in the ones originally thought as one of the questions, it was added into the questions in order to get more viewpoints and thoughts on the topic from additional interviewees. It is pointed out in that excuse's specific section whether it was added into the questions over the course of the interviews and at which point this happened.

For the Results section, we utilize inductive reasoning to determine the implications of each set of answers to a neutralization technique. We go through each neutralization technique in the order listed before in section 3.2. Within each technique, we first list the specific question asked from the interviewees. Then, we explain the overall reception of the excuse by the interviewees, the specifics of which can be found in the appendix section. After these, we delve deeper into the thoughts and opinions of the interviewees along with any reactions to lead-up questions as well as any further findings and interesting topics that might have stemmed from the interviews.

In addition, we determine the plausibility of any given neutralization technique by comparing the amount of interviewees considering the excuse plausible as opposed to implausible. Anything with two or fewer plausible answers is deemed highly implausible, while anything with two or fewer implausible answers is deemed highly plausible. Such polarisation, we believe, is sufficient enough to determine the significance of a result, despite the relatively small number of interviews conducted. Anything else, beside these two judgements, is deemed plausible although if the answers lean heavily to the implausible, we make a note of this and discuss it in the section.

Exceptions were done if the number of answers was particularly low. If the number of answers to a question is six or fewer, we require one or fewer plausible or implausible answers to deem it significant enough to claim that topic as very plausible or implausible.

Anything that's within one answer of achieving very plausible or implausible status is marked as likely or unlikely. This, in essence, is the same as plausible but with a higher degree of certainty about it.

5 RESULTS

Digital platform usage varied greatly between the participants as did their prior knowledge of the topic, this being digital disinformation and the existence of crowdturfing campaigns. No knowledge of neutralization theory was required nor was it ever mentioned outright during the interviews.

We will now go through each neutralization technique and list our findings within them. Since some of the interviews were done in languages other than English, some of the quotes listed are translated and not literal. Furthermore, it is important to reiterate that the neutralization techniques were not gone through in the same order for each interviewee and, thusly, the example lines you are about to see are not from the same point of the interview. This likely makes one-to-one comparisons between interviewees difficult, but in order to maintain fluid, dynamic conversations, we felt it important to not restrict the conversation to specific order of topics. In addition, the example line may not be the reason that establishes plausibility. Instead, it is the most interesting input of the interviewee on any given neutralization technique. You can find the complete list of these in the appendix section.

Before we do that, however, here is the overall plausibility of each question asked in this study and the plausibility we've derived from the interviews. Significant finds are in bold. Conclusions with an asterisk in them mark results that are inconclusive because of analysis done afterwards. Refer to the respective sections for these reasons as to why these questions were thought to be inconclusive.

Excuse	Y to N Ratio	Conclusion
One more participant doesn't affect anything	8 to 1	Very Plausible
I didn't realize this wasn't allowed	4 to 5	Plausible

Table 2 Denial of Responsibility

It's not a big deal	6 to 3	Plausible and likely
It's digital so it doesn't matter	4 to 2	Plausible
Nobody who could be hurt was hurt	4 to 0	Very plausible

Table 3 Denial of Injury

I didn't affect anything	6 to 3	Plausible and likely
Nobody believes digital media anyway	5 to 4	Plausible

Table 4 Denial of the Victim

It's the consumer's fault for being fooled	5 to 4	Plausible
Social media monopoly is why we think this is wrong in the first place	3 to 6	Inconclusive*
It's the rules that are wrong	1 to 8	Very implausible

Table 5 Condemnation of the Condemners

I couldn't say 'no' to Jessica	3 to 6	Implausible/Plausible*
This is the future of advertising	6 to 3	Plausible and likely
I thought it would be interesting	3 to 4	Plausible
I thought it would be fun	4 to 5	Plausible
I did it for the money	8 to 1	Very plausible

Table 6 Appealing to Higher Loyalties

We're friends so don't doubt my actions.	3 to 6	Inconclusive*
I'm entitled to profit from my social media	3 to 6	Plausible but unlikely

Table 8 Appealing to Good Character

I did it for the wronged workers	0 to 8	Very implausible
The rival company is doing the same thing	1 to 8	Very implausible

Table 9 Victimization

5.1 Denial of Responsibility

Denial of responsibility occurs when a person attempts to downplay the magnitude of their crime.

5.1.1 One more participant doesn't affect anything

Plausibility: Very plausible with 8 to 1 ratio.

Appendix 3.

Most of the interviewees considered this a very plausible line from Paul. Some agreed that Paul's involvement probably didn't affect much, while some were less charitable of Paul's reasoning.

M2: It all starts somewhere.

M3: Yeah, and everybody else thinks the same.

F3: Even a single person can make a huge difference.

Overall, the thought was that while Paul was simply one individual partaking in the campaign, he was still responsible for his own actions and, thus, simply because he could join the campaign, didn't mean that he should or was in any way forced to do so. Thus, it was felt that this excuse was at least weak or poorly thought out.

This conclusion, however, doesn't deny that it is very plausible as an excuse in this situation.

5.1.2 I didn't realize this wasn't allowed

Plausibility: Plausible with 4 to 5 ratio.

Appendix 4.

Surprisingly many thought this to be a reasonable argument from Paul, especially since many of the interviewees were under the impression that crowdturfing campaigns such as the one portrayed are legal currently.

For example, twitter terms of service outright bans third-party advertising so long as it hasn't first been approved by twitter. This is the deceit under which crowdturfing campaigns operate.

Others argued that a reasonable person couldn't possibly think their actions were justified, namely deceiving their close circle of friends and writing false reports. Many went so far as to say that if Paul had done enough research, so he could claim his opinions were valid and based on his own thoughts, and toned down his opinions slightly, that they wouldn't think too much of his participation.

M3: Marketing is, at some level, always dishonest.

M4: There's probably a loophole somewhere that allows these campaigns to exist.

Interviewer: So, in the end, your problem with Paul's actions was that he didn't disclose it was marketing? You think if he had done that, everything would be okay?

F4: Yeah. That's all these people need to do. There'd be no problem that

way.

The problem that was perceived then with Paul's actions and crowdurfing campaigns in general, was that the marketing wasn't disclosed and that it was masquerading as an honest opinion. Could crowdurfing develop into a legitimate field of commerce whence this clause was being upheld? Maybe a certain degree of cooperation between the crowdurfing agents and social media sites could be possible, where participation in one automatically flagged all interactions by those involved as advertisement for the duration of the campaign? These are design decisions the platforms involved have to solve. That is, if they want to solve them in the first place.

More on this topic in 5.4.2, condemnation of condemners and moral relativism, where we explored the topic of why social media platforms forbid this type of advertising.

5.1.3 Summary on denial of responsibility

Denial of responsibility provided very plausible neutralization techniques for our character. Both of the avenues explored proved plausible to some degree, although some argued that a person should know that their actions in this context weren't legal. However, since many of the interviewees themselves thought the actions were legal at the beginning of the interviews, this point is more than likely a result of a singular interviewee's morality on the topic.

Attempting to argue that adding one more participant to a wide-spread campaign proved plausible. Many of the interviewees agreed that many of the workers probably reasoned their participation through similar means, leading us to believe that this a very well established neutralization technique with good applications.

Attempting to argue from ignorance was seen as almost equally plausible, with many of the interviewees believing this at the beginning of the interviews. This is in addition to their statements, where they felt that this was something Paul could have said in the context. This, of course, doesn't stand on a moral ground but with a slightly skewed morality and a good rationalisation to oneself, it can easily be seen how this would manifest in real life. As such, this neutralization technique is plausible as well.

5.2 Denial of Injury

Denial of injury occurs when a person claims that there was no crime or that nobody was harmed by their actions.

5.2.1 It's not a big deal

Plausibility: Plausible and likely with 6 to 3 ratio.

Appendix 5.

Many of the interviewees viewed digital disinformation as a very low-level or "victimless" crime, at least on the level at which Paul was operating on.

M2: Probably half of five and one star reviews are fake anyway even if they weren't paid to do it.

M1: Paul doesn't really have a lot of influence anyway.

However, as a counterpoint, F4 argued that even if one person was swayed by or maliciously impacted by Paul's actions, it meant that it was a big deal.

F4: It was a big deal to that person who was deceived.

How cognizant is it reasonable to expect Paul or people in his position to be? Judging by how most of the other interviewees saw Paul's actions petty at most, it is safe to assume that people in his position wouldn't quite consider this avenue as something preventing them from using this particular excuse as their neutralization.

Noting F1's comments, it is clear that only few if any of the interviewees considered the platform as a possible injured party or victim in any of these crimes. As such, despite claiming there probably was no or little injury, the loss of respect to the platforms themselves wasn't considered for in most of their considerations.

5.2.2 It's digital so it doesn't matter

Plausibility: Plausible with 4 to 2 ratio.

Appendix 6.

This question was raised by M1, and we deemed it interesting enough of a topic that it was added for future interviews.

While some of the interviewees agreed that no physical harm is done by digital interactions, mental anxiety from social media is real, at least according to their own opinion. And even if Paul wasn't actively trying to cause distress, only covert deceit, some argued that digital, especially today, was important to many people.

M5: This isn't just some internet trolling though. Paul is trying to make these people take real world actions.

M4: Yeah, but he's trying to sell a physical thing.

Similarly, even if the effects of Paul's actions were in digital space, he was trying to deceive people into taking real world actions.

In the end, the interviewees felt that the excuse was plausible for somebody to say, but didn't agree that Paul's actions were purely digital if he was trying to deceive people into spending real money and otherwise affect their actions. As such, they thought that if Paul was honestly trying to excuse his behaviour with an excuse such as this, he was very deluded at the very least.

5.2.3 Nobody who could be hurt, was hurt

Plausibility: Very plausible with 4 to 0 ratio.

Appendix 7.

M3: You probably have the money to lose if you can just buy a phone without doing your own research.

This point by M3 raised an interesting point. If you are deceived into buying a relatively good product, how much were you deceived in the first place then?

So how much of an injury was there in the first place? Obviously, this isn't the case in all possible applications of digital disinformation, so in that way the framing of the vignette may have been slightly lacking for this question in particular. There quite wasn't enough time during the interviews to explore hypotheticals, where if the product was dysfunctional or unpleasant

how much that would affect the view of this question in particular.

The interviewees were quite accepting of this excuse by Paul, and while this doesn't excuse all aspects of his campaigning, it does answer the most major goal, which was to sell the main company's phone (Samsung). While the concern remains as to how this neutralization would fare in different contexts, where the purpose is more directly malicious or outright deceitful, instead of just artificially hyped towards one company over another, this neutralization stands as very plausible.

However, thought from a different point of view, how is one to truly do their own research if much of the information available is falsified? If marketing is dishonest and reviews and comments are crowd-turfed and videos are sponsored, how is one to truly do their own research? We haven't yet reached this stage of prevalence in disinformation, but should the trends continue, it may very well be a realistic future.

5.2.4 Summary on denial of injury

Denial of injury as a whole proved very receptive in the views of the interviewees. What helped Paul's case in making the excuses more believable was the low stakes of the crime committed and that it was done in digital space.

Attempting to argue that what one's actions were no big deal, turned out a fair success. People thought Paul's actions were overall not a big deal and agreed that somebody in his situation would use it as an excuse. Some interviewees were less charitable of this interpretation, but overall this neutralization technique stands.

Attempting to argue that since it was digital, it matters less, proved similarly successful. It is possible that this is simply rethreading the previous point, except with us outright stating the reasoning some of the interviewees made, in that digital actions have lower stakes. Nevertheless, if this ends up as simply a reinforcement of the previous point or an angle of its own, it is clear that it is plausible.

Attempting to argue that nobody was brought in peril or jeopardy because of the potential purchase of a deceitfully marketed phone turned out fairly plausible. This question could be regarded as another avenue of the original one for denial of injury, but it provided some interesting context at least.

The two additional reasonings for "why it's not a big deal" can be seen as continuations of the original point, meaning that they don't necessarily deserve a mention of their own and instead should be treated as variations of the original one. Regardless, they are included in this breakdown, despite not having all that varied results.

5.3 Denial of the victim

Denial of the victim occurs when the worker claims that their actions didn't cause anybody harm.

5.3.1 What I did affected nobody

Plausibility: Plausible and likely with 6 to 3 ratio.

Appendix 8.

The argument here was that nobody was going to buy a phone simply because a stranger wrote a five star review of it or if they saw a comment praising it or they saw the like/dislike ratio of a phone advert be positive. Of course, Paul's actions weren't limited to strangers alone.

M4: If that was the case, nobody would be doing it (marketing/crowdturfing)

M1: Maybe I wouldn't run out buying one, but if I had to make a choice, I would be reminded by that post from a friend in my feed way back.

With surprising clarity, many of the interviewees recognized that marketing worked exactly in the ways it was supposed to. By raising awareness of a topic or object, they were more likely to choose in the favour of it, if they had to make a decision between two otherwise equal items.

Strangely enough, in addition to this, a few interviewees remarked that they were wary of sales objects, like in amazon, with only a few praising reviews, but felt more trusting the more there were of them.

M2: Yeah, but it's not like you can buy that many (reviews)

If nothing else, with M2's comments we provide anecdotal evidence that a large scale crowdturfing campaign would work in swaying people's opinions into buying products by reviews alone.

And while many interviewees rejected Paul's reasoning, they understood why somebody would think that way and accepted the excuse as a plausible one.

5.3.2 Nobody believes what they see in the internet anyway

Plausibility: Plausible with 5 to 4 ratio.

Appendix 9.

The intention here was to explore denial of the victim from the point of view of the platform, in that their credibility is already so bad that hurting it more was not a problem, although, judging from the responses, this nuance was most likely lost and not properly conveyed.

F3: So you'll just go intentionally causing more harm?

M5: I guess a healthy dose of suspicion is fine, but I doubt Paul is thinking he's doing a service or anything like that.

This excuse was felt as fairly malicious on Paul's part, unlike the other angle for denial of the victim, which was thought to be more ignorant. This only slightly affected the plausibility the interviewees felt about the excuse itself, believing that it was still something Paul could've said.

5.3.3 Summary on denial of the victim

Denial of the victim reinforced the concept of marketing, if nothing else, and crowd-turfing's strategy as a whole. In addition, the neutralization techniques were felt to be very plausible as excuses but misguided.

Attempting to argue that one's actions affected nobody was felt very plausible. The disparity was probably the greatest in this excuse between the perception of how mistaken Paul's reasoning was and yet, how plausible his excuse was thought to be. It is hard to say as to why nearly all of the interviewees both knew how marketing worked and, at the same time, thought that knowing how marketing worked wasn't common knowledge. In addition to this, many of the interviewees provided anecdotal testimony of advertisement's usefulness.

Attempting to argue that social media platform's credibility was already so ruined that no additional harm could be done to it by a single individual was thought to be fairly accurate and plausible. But, while no interviewee explicitly took the side of the social media platforms, the general thinking was that this didn't excuse profiteering from it and making the situation worse.

5.4 Condemnation of the condemners

Condemnation of the condemners occurs when a person attempts to put the blame of their action on the victims of their crime.

5.4.1 It's the consumer's fault for being fooled

Plausibility: Plausible with 5 to 4 ratio.

Appendix 10.

Together with this question was a discussion in the importance of media literacy, as well as how capable the interviewees believed themselves in noticing digital disinformation and advertising in their own social media feeds. The intention was to prod at the interviewees, especially if they answered positively to both of the lead-up questions, but treated the excuse negatively.

The conclusion that was reached by some was that while it was, in the end, responsibility of the consumer to not be fooled, it was still felt to be improper for Paul and his like to take advantage of this.

Media literacy was thought to be of increasing importance, especially since the perception was that digital disinformation or fake news was on the rise. One interviewee remarked that it would be exhausting if they were going to have to treat their friend's comments with similar scrutiny. Another said that they didn't take anything in social media seriously, and instead used it to mostly keep up with trends and current events.

F1: It's true that media literacy is important, but you still shouldn't go about making the situation worse.

M3: The reporting of those events is probably fake though, unless you spend the time to read ten articles of the same thing.

With these lead-up questions, many interviewees were quick to recognize Paul's reasoning with his excuse.

M3: I get it, I guess, but I wouldn't stay as a friend of a guy who thought that way that's for sure.

M2: That still doesn't excuse his involvement. Just because you can, doesn't mean you should.

With this being said, the opinions on whether or not this excuse is plausible are split. Many recognized that the excuse sounds plausible, especially since they had seen it depicted in movies and tv shows, while others

considered it comically absurd, maybe for the same reasons.

Whether reality imitates fiction or vice versa, this neutralization technique, all things considered, appears to be plausible, albeit only slightly.

5.4.2 Moral relativism

Plausibility: Plausible but unlikely with 3 to 5 ratio.

Appendix 11.

The full line of questioning went as follows: "The only reason this is thought to be wrong is because social media platforms want to keep their advertising monopoly."

Arguably a very convoluted question, but this is only done so the context of the lead-up question is involved within it for the purposes of dissecting it in this breakdown. The attempt was to discuss the topic of these third-party advertising agencies using the social media space for their own goals, rather than going through the proper channels and paying the platform for the privilege.

F3: I'm pretty sure I'd still think lying was bad even if facebook didn't tell me so.

The general thinking was that the immorality of dishonest advertising is believed to be so apparent that it doesn't need a third party explaining it to them.

One might think this runs counter to the point made in 5.1.2, denial of responsibility, where many of the interviewees didn't know that this type of advertising wasn't allowed, but this was mostly because the general thinking was that legislation relating to dishonest online marketing didn't exist yet, not that the actions committed weren't worthy of being outlawed and forbidden.

M2: Legislation always lags behind technology.

Returning to the question itself and the reactions towards it. While the interviewees thought the thinking was a bit too convoluted, they realized that these outside advertising agencies didn't have any right to utilize the space created by these platforms.

The general thinking then is that since platforms, such as twitter and facebook rely on advertising of their own to function, they have no reason to let competitors in advertising run freely on their platform, regardless of the legality of the campaign itself.

Overall, it may have been better to split the question in two with: "You only think it is wrong because social media platforms wrote it in their terms of

service.” and “Social media platforms have a monopoly on online advertising on their platform. Competition is good for them.” Or something similar.

5.4.3 It's the rules/morality that is wrong

Plausibility: Very implausible with 1 to 8 ratio.

Appendix 12.

As the follow-up question to the previous one for the some interviewees that didn't outright state it already that they didn't need an outside source telling what was and wasn't distasteful.

This excuse didn't go too well, since, as mentioned previously, the interviewees thought that a well-established person couldn't possibly think that deceiving people in the depicted manner was proper, regardless of any established rules or laws or regulations.

As such, the opinion of the interviewees remains in that this line of thinking is unreasonable from Paul and thus it is improbable as a neutralization technique, should the situation occur in reality. This exploration does, however, provide us with an important angle in determining the reasons why people think disinformation is so prevalent. This being that the interviewees think it is legal and, thus, there are vultures ready to exploit a system that hasn't been legislated well enough. Of course, this can't be further from the truth, but the perception remains regardless.

5.4.4 Summary on condemnation of the condemners

Condemnation of condemners was thought to be slightly leaning towards fiction and tv than real life. While this remark wasn't a common thought among the interviewees, only one outright stating it and another coming to the same thought during the discussion, the point raises an interesting conundrum. Would a real person use these as an excuse or are all the interviewees affected by media they've consumed previously lead into thinking this was something a real person would use as an excuse in a situation like this? Unfortunately, we don't have the answer to this question.

Attempting to argue that the blame lied on the gullibility of the consumers/victims, was thought to be a plausible excuse, so long as we acknowledge the possible problem stated previously. Since this study can't accurately assess the magnitude or validity of this hypothesis, we must leave it remain just that: a possibility.

Attempting to argue that the rules are wrong proved to be a fairly

plausible excuse, although many came to the conclusion that nothing was forcing these platforms to accept covert advertising as legitimate activity on their platforms. However, once we look at the response to the next question, we may see that most of the plausibility of this excuse stemmed from the part that didn't relate to morality being written in terms of service.

Attempting to argue that morality itself is wrong didn't turn out as plausible. Interviewees thought that a stable human being wouldn't consider the distaste people have towards deception as something that is in need of change. As such, this neutralization technique was thought to be implausible.

While the interviews gave the impression that condemnation of condemners was overall a plausible excuse to give, it does have a sense of overdramatisation and grandiosity that is not shared by other neutralization techniques. Would a petty criminal consider themselves the harbinger of changing laws or morality, or an arbiter that punishes the foolish who believe everything they see in social media? Maybe, is the answer we can draw from these interviews.

5.5 Appealing to Higher Loyalties

Appealing to higher loyalties occurs when a person claims to have done something for the purpose of a greater cause that supersedes law and decency or anything of the sort.

5.5.1 I couldn't say 'no' to Jessica

Plausibility: Plausible but unlikely with 3 to 6 overall. However, if split for the interviewees who saw this as one of friendship or romance, this ratio changes drastically.

Plausibility for friendship: Very unlikely with 1 to 6.

Plausibility for romance: Very likely with 2 to 0, but these numbers are too small to make a conclusion.

Appendix 13.

Within the vignette story, Paul is introduced to the campaign by a friend, Jessica, and Paul is characterised as amiable to his friends so that he has a hard time saying 'no' and always feels like he has to have a reason for whatever he does. This characterisation was done in an attempt to facilitate this question to

where his loyalty to Jessica's request was his reason for participation. Similarly, the names of the characters were purposefully of opposite sexes to further provide cover for this excuse in particular.

What the interviewees felt was that this was inadequate, considering the effort Paul eventually put into his participation, although some did see the possibility of it.

The interviewees thought that simply receiving an invitation from a friend didn't warrant a thorough participation in an online advertising campaign, especially one such as this. Some interviewees, who presumed a more romantic motive from this excuse, thought that this was understandable, albeit desperate and foolish.

If we had included a more serious need for Paul's involvement from Jessica's point of view, maybe this excuse would've received more support, but the intent of the vignette was to keep all the stakes at a similarly low level so as not to guide the interviewees' opinions one way or another.

Overall, Paul's attempt to claim higher loyalty from his friendship to Jessica wasn't enough to convince the interviewees of Paul's need to participate in the campaign, thus invalidating this excuse in their eyes.

On its face, this is what can be derived directly from the data and answers. However, some of the interviewees presumed that Paul's reluctance to saying no to Jessica was because of romantic hopefulness. These two, M1 and M3, answered positively to the question. This leads us to believe that the line that divides participation and nonparticipation lays somewhere between friendship, which was thought to be very implausible, and romance, which was thought to be very plausible, from what little data we have on the matter.

In addition, from this question, a surprising find was made in regards to one of the interviewees themselves.

F2: I'm pretty sure Jessica was only recruiting Paul because she got more points or whatever for the campaign. They always do this sort of thing.

F2 presumed campaign recruitment to be the reason why Jessica invited Paul in the first place, despite this not being outright stated within the vignette itself. This particular interviewee had personal experience of a similar campaign where they were incentivized to promote their own company in social media. At the time she felt this fairly normal, but began second guessing her experiences throughout the interview until eventually coming to the conclusion: *F2 - Maybe that was a bit scummy.*

F2: But it really didn't seem like such a big deal at the time. A lot of companies do exactly the same. They (social media) can't expect everybody to always disclose their affiliations for every interaction.

That is a good point. What is the standard of honesty social media interactions are required to operate under?

Bella M. DePaulo et al. in *Lying in Everyday Life* (1996) found that college students lied on average 2 times a day. The number lowered to 1 for community members. Maybe a reason why students are the primary workforce for crowdurfing campaigns stems partly from this willingness to deceive, however numerically small the difference is. The same paper then found that all social interactions were perceived more favourably when they were honest. But is this a reasonable standard to have?

While this topic would be an interesting question for further research, we will have to leave that for another time.

5.5.2 This is the future of advertising

Plausibility: Plausible and likely with 6 to 3 ratio.

Appendix 14.

This is Paul appealing to the concept of progress and development.

Many of the interviewees showed distaste towards Paul's reasoning here but saw the possibility of it. A common thought among many of the interviewed was how M1 put it:

M1: "Maybe it's the future of advertising, but I sure hope it wasn't."

Paul was seen contributing to this change and, while they couldn't exactly propose questions to him, they would have asked if Paul wanted this to be the future of advertising and if he realized that by participating in it, he was actively creating that future.

M2: Everything similar has been regulated before already, so I don't see how this wouldn't be either.

On the other hand, many doubted the premise, stating that even if it wasn't outright unlawful already, it would be soon enough. Paul's excuse here was thought to be misguided and his goal here, eventually, would be a waste of time or lead him into bigger trouble.

5.5.3 I thought it would be interesting

Plausibility: Plausible with 3 to 4 ratio.

Appendix 15.

Originally brought up by F2 in relation to the previous question, providing an

interesting avenue to explore the same topic. Seeing how Paul was characterised as somebody interested in the possibilities of social media advertising, we believed this to be a fairly accepted reasoning for his actions. However, this presumption was proved to be only somewhat correct in that this excuse received a fairly measly number of accepted responses. Only 3 of the 8 interviewed thought this to be a valid reason for participation, so long as Paul was being honest in saying this. This is a resurgence of the problem Agnew (1994) brought up when he criticized neutralization theory as simply being post hoc excuses of a caught criminal rather than proper reasons for their actions. For the purposes of these interviews, the interviewees were instructed to consider each excuse as an honest attempt on Paul's part, although, clearly, this was often forgotten or doubted.

Others were less accepting of this neutralization, arguing that one can study a criminal topic without becoming one themselves, but it was generally understood that the harm in participating in a crowdturfing campaign was much less than if Paul had committed a 'true' crime, further reinforcing the idea that digital disinformation is seen as a "victimless" crime.

F4 - At least nobody got hurt by his actions so maybe it's somewhat okay.

Here we also find that at least one interviewee neutralised this even further with denial of injury. This leads us to believe that combinations of neutralization theories may sometimes be utilized. We can't make claims as to how effective they are in general, but this should be an interesting topic for future research. As F4's answer was initially negative, but after she further neutralized it so it became plausible, we did not list her answer to this particular question as either plausible or implausible.

5.5.4 I thought it would be fun

Plausibility: Plausible with 4 to 5 ratio.

Appendix 16.

This excuse by Paul was initially met with disapproval, but most interviewees eventually came to the conclusion that it was a plausible excuse for Paul to give. M3 even went so far as accepting this excuse outright.

M3: If I was that friend talking to Paul now, I'm probably somewhat like minded.

Interviewer: But here you have the friend disapproving Paul.

M3: Sure, but I could laugh at a distasteful joke, especially if Paul is being paid for doing it.

When reminded that this friend was meant to hold this dishonest campaigning a deal-breaker to their friendship, hence the excuses, this interviewee said:

Others were less accepting of Paul's excuse, but understood that:

M4 - Unfortunately, I can see somebody saying that.

On the flipside, a fair number of interviewees were harshly opposed to this excuse. However, despite this polarisation, we must deem this excuse as plausible.

5.5.5 I did it for the money

Plausibility: Very plausible with 8 to 1 ratio.

Appendix 17.

In the vignette, it was stated that Paul was in no way hurting for cash nor had he ever experienced poverty to such a degree where one had to resort to criminal activities in order to make ends meet. These were done in order to mitigate this excuse as it was predicted to be most plausible of all and, if that was to be the case, it would overshadow all other possibilities. This would especially be the case if the interviewees' began to think that all the excuses presented were, in fact, only attempts by Paul to obfuscate his own greediness to his friend.

Despite these attempts, the topic of money rose up and was thought to be very plausible as a reason for Paul's involvement. Again, some bias might be found due to the interview group being students themselves and, presumably, more willing to do petty deeds for minor cash payments, despite the criminality involved, although as was previously found, many didn't think this was illegal in the first place, only distasteful.

Regardless, as an excuse, monetary incentives were thought to be the most honest and forthright reason Paul could give and thus is the most plausible excuse found during this study.

To continue, some went so far as to accept his involvement outright, if his situation truly warranted it.

The more destitute Paul's position was presumed to be, the more willing people were to accept his actions. However, we can't vouch for this data as the discussion on this particular excuse didn't go into hypotheticals such as this except in two interviews, so this anecdote here stems from very limited sources.

To note, this is F3's only positive answer and, as you can see, this answer isn't purely hers. The interaction is noted in the answers for posterity's and

transparency's sake. It could be the case that her answer to this question should be disqualified because of researcher meddling, but we've decided to address the issue with this paragraph instead.

5.5.6 Summary on appealing to higher loyalties

In the end, higher loyalties provided us some plausible neutralisations. Whether Paul's loyalty eventually laid with a) his friend or romantic hopeful, b) his own curiosity and studies, c) his own entertainment, d) his potential career or e) his own monetary benefit, the overall reception was mixed, but still slanted more towards the plausible side.

Attempting to argue that a worker's participation in a digital disinformation campaign was because of their loyalties to a friend who invited them turned out to be very fruitless, with only a few plausible opinions even among those who perceived Paul to have romantic hopes for Jessica. Maybe, if we had made two questions, one for an existing good friend and one for romantic candidate, we could've distilled the results more thoroughly. Regardless, we determine that, for now, this neutralization is unlikely.

Attempting to argue that everything a worker did was for the fun of it turned out to be a very polarised avenue. Interviewees were either vehemently against it or capable of seeing it happen. However, none of the interviewees found this improbable, hence this excuse remains as one of the most highly probably ones.

Attempting to argue that participation in crowdturfing campaigns such as this was a legitimate marketing tool or a potential future career opportunity. While many preferred this to not be the reality of it, many saw it as a possibility and thus found this excuse plausible.

Attempting to argue that one's participation in crowdturfing campaign, despite all the attempts to downplay the payments of these campaigns and despite the fair wealth of the involved worker in question, monetary reasons were thought to be very plausible among all interviewees.

Attempting to argue that a worker's participation is because of their curiosity towards the workings of such digital disinformation campaigns proved fairly successful. This could be partly due to each of the interviewed being students of their own and, we predict, that this result could vary if the target group was changed. Regardless, joining a disinformation campaign in order to find out how it worked was deemed a plausible excuse for Paul to give.

Thus, we find that selfish intrinsic motives seem to be more plausible than extrinsic motives. Of course, this doesn't mean they were accepted as a reason for doing anything described in the vignette, but the interviewees

considered these excuses as something a participant such as Paul could say. The only higher loyalty which had a negative score was for Paul to claim loyalty to Jessica simply for having asked Paul to join the campaign. The few interviewees who saw this as something more intrinsic to Paul himself (i.e. Romantic motivations) felt this excuse much more plausible than if the thinking was that Paul was doing it simply because he couldn't say no.

This leads us to believe that the more intrinsic a motive, the more likely it is thought to be plausible as a neutralization. This is supported by the category 5.7, victimisation, where we establish that a worker attempting to act out past wrongs, an extrinsic motive, was thought to be highly improbable.

5.6 Appealing to Good Character

Appealing to good character occurs when a person attempts to excuse their current behavior with past good deeds and actions.

5.6.1 I can't believe you would doubt me, your friend.

Plausibility: Unlikely with 3 to 6 ratio.

Appendix 18.

Here we attempt to find if appealing to their friendship status could provide a reasonable defence.

While the interviewees found it possible that a worker in a dishonest campaign such as the one portrayed could lash out when questioned, they didn't find it a reasonable reaction towards a good friend of theirs. However, if this was said towards otherwise accusatory strangers, with the inability to appeal to their own character, this reaction was thought to be a plausible one, albeit an unsatisfactory one, since it doesn't really explain anything except for the presumed personality of the worker, which we shouldn't be making assumptions about in the first place from a vignette based study like this.

Overall, this avenue of exploration didn't turn out very profitable. While the plausibility of the neutralization technique was established to a certain extent, knowing it doesn't provide us much. There were few among the interviewees who rejected this excuse as they didn't believe it matched Paul's character established in the story, especially if this friend they were explaining themselves to was supposed to be their best friend.

F2: If I was that friend, I'd ask him again.

M5: I can see somebody saying that, but it isn't much of an excuse, is it? He's just avoiding.

Another issue that was found with this avenue was that the excuse itself is unsatisfactory for the purposes of a real conversation. Even if a person gave this as their excuse, it was brought up that the conversation would likely continue on to a new excuse, negating the existence of this one, or risk the friendship of the two, albeit fictional, characters. Meaning, that if this was given as a neutralization, it would most likely only be a stepping stone to another more deeper neutralization technique.

5.6.2 I'm entitled to profit from my social media presence

Plausibility: Unlikely with 3 to 6 ratio.

Appendix 19.

Alongside this question was a lead-up question where we asked whether or not the interviewees believed that social media accounts were property of their owner's. The catch here was that if they answered positively to the lead-up question and answered negatively to this follow-up excuse based on that reasoning, we could prod them for additional information.

M1: I guess it should be the property of the owner.

M1: But, no, you can't just do whatever you like with things you own, especially since you're affecting others with it.

It is questionable in the first place whether social media accounts are property of their users and not simply leased under a terms of service. The answer varies for social media platforms with some like twitter not stating one way or another. What is clear, however, is that third-party advertising is forbidden, regardless of the 'effort' one has put into their account.

In the vignette it was stated that Paul put a moderate amount of effort into his social media presence and had a modest following, mostly among his family and friends. From this, it was argued that this was an unreasonable stance for Paul to have since he didn't start doing it with the expectation of money.

Others were of the opinion that while Paul wasn't entitled to any profit from his efforts so far, nothing stopped him from going through official channels and trying to make a profit that way. While third-party advertising is banned on twitter without twitter's approval, that doesn't mean third-party advertising doesn't exist on twitter so long as twitter's given approval for it.

However, crowdturfing campaigns make their profits from being cheaper than the legitimate alternatives since they avoid paying a share of their revenue to twitter and the like for allowing the advertising on the platform.

Overall, despite the lead-up question and the contradiction often it caused in the interviewee's answers, it was felt that this line of reasoning was fairly unreasonable, especially since if Paul really wanted to, he could've tried using official channels if he wanted to profit from his social media presence.

The answers to this neutralization technique could be different should the vignette be built, based on somebody with minimal social media presence, who had no method of profiting from his presence, but, in that case, the question itself would need to be restructured since it includes the pretense that Paul had put effort into his social media.

5.6.3 Summary on appeal to good character

The interviewees, overall, thought it plausible for workers to appeal directly to their good character. However, it wasn't deemed satisfactory and more than likely another neutralization would be required of the worker character. This was less of a case where the good character was based on entitlement to something.

Attempting to argue that one's previous friendship with the accuser should make them trustworthy enough was found to be plausible, but very unsatisfactory. Many of the interviewees would've demanded Paul of additional context or reasoning. Hence, this neutralization could be considered incomplete without a follow-up neutralization technique. It is also dubious whether this counts as a neutralization in the first place, seeing how Paul isn't attempting to explain his behaviour, but instead is trying to avoid the conversation altogether.

Attempting to argue that one's previous efforts entitled them to profit, even through criminal means, was felt much less plausible. Questions were raised as to the state of social media accounts and their ownership, which may be an interesting topic to explore in the future, but for the purposes of this study, we must deem this excuse as fairly improbable.

5.7 Victimisation

Victimisation as a neutralization technique appears when a person claims that their conduct was, in fact, done because of a misdeed perpetrated upon a third party.

5.7.1 I did it for the wronged workers

Plausibility: Very implausible with 0 to 8 ratio.

Appendix 20.

The clearest finding from the interviews was in regards to victimisation with none of the interviewed considered this a plausible reasoning for participation.

Victimisation was attempted to be brought up through personal indignation on Paul's part because of a story about worker's right violations and ecological harm due to the actions of the rival company, HTC in the real-world equivalent. This was then framed as an additional driving force for the worker's participation in the campaign.

M2: I don't believe anybody is going to join a campaign like that with the intent of moralizing the people they are campaigning to.

M4: Like a form of Stockholm Syndrome? Where they actually start believing the stuff they are deceiving about?

Obviously, there is some difficulty in trying to link an unaffiliated student crowdurfing worker to have personal reasons of victimhood about a phone company's illicit marketing campaign in the first place, but having such a clear rejection of its possibility, we believe, is significant. Of course, this could be due to a faulty framing on our part, but such a clear divide would be hard to achieve purposefully, let alone mistakenly.

M5: Maybe if he read about that before, but this is just post-hoc rationalisation.

While M5's comment was what neutralization theory quite quintessentially is, his overall opinion of the topic was that it was unlikely for Paul to use such an excuse. We did mark his answer as inconclusive because of this however.

Within discussions related to this topic, the interviewed often expressed confusion towards the question itself, stating that they couldn't have come up with this reason on their own. Regardless, the rejection of it usually didn't take long with only a few interviewees expressing any hesitation towards a negative answer.

What can we infer from this? The clear inference is that none of the participants thought a worker could take personal victimisation towards any action committed by the company they were working against. At most, the thought was that this could manifest through a derangement such as Stockholm syndrome, where they start believing that the campaign they were part of was righteous in its malicious acts towards the rival company. This was,

however, conjecture on part of a single interviewee so this hypothesis can't be verified nor explored any further in the context of this study at least.

5.7.2 The rival company is doing the same

Plausibility: Very implausible with 1 to 8.

Appendix 21.

With this question we attempted for Paul to take the side of their company in their attempt to get a legitimate competitive edge. Whether this qualifies as victimisation is arguable, but this is where it's being included.

Interviewees overall agreed that this is probably the case, but some raised questions as to whether this even worked as Paul's excuse, seeing how he's arguing for the company's actions rather than his own inclusion in the campaign.

Since the existence of a campaign doesn't force Paul to join it, his participation in it shouldn't be explained by it. The moral decision eventually was thought to be with Paul and this excuse was felt to be improbable, although to not such a degree as the previous one.

M4: Just because it's being done, doesn't mean you should do it yourself.

F3: The answer isn't to do it yourself, but for both to stop.

M2: I think a better use for the company's money is to just lobby laws to prevent this type of advertising in the first place.

In the end, this exploration proved out to be a failure. It must be acknowledged that the excuse indeed doesn't validate Paul's perspective. It might be the case that this neutralization technique is incomplete and would be followed by another neutralization technique, such as "One more participant doesn't hurt.", which is explored in the next section. This would mean that this neutralization technique would, in fact, rather be a reasoning for the validity of that one, rather than a neutralization technique in of itself. Despite this, it provides reinforcement for the other victimisation question in that the interviewees felt that this was something a worker in Paul's position wouldn't care about.

5.7.3 Summary on victimisation

Overall, victimisation doesn't seem like a plausible neutralization technique.

Attempting to argue that one's involvement was because of a past wrongdoing by a rival company was thought to be unreasonable. Maybe if this

was outright stated that the worker joined the campaign because of what he had read about the rival company, but this approach wouldn't leave much for interpretation and wouldn't be much of a find if it was stated. Regardless, the clear rejection of this portion of victimisation shows a clear unlikelihood for this neutralization technique.

Attempting to argue that covert actions such as this were legitimate because the other company was doing something similar was thought to be unreasonable and misguided. It is likely that this neutralization technique doesn't qualify as one and should be treated as a technique used to reinforce another neutralization technique.

One can clearly see the connection of victimisation and cause as a motivation from MEECES. In our breakdown of cause as a motive we theorized that it could be seen as either intrinsic or extrinsic. However, these discussions with the interviewees lead us to believe that, at least, the perception of such motives is that they are extrinsic.

6 DISCUSSION

While any definitive conclusions are hard to draw from such a small sampling of interviews, few things are apparent enough to deserve a second mention and a possible further look into.

6.1 Theoretical implications

In this section we are examining the implications our findings have on the various frameworks and theories we've utilized.

6.1.1 Neutralization theory

The goal of this study was to assess the plausibility of different neutralization techniques in digital disinformation contexts, crowdturfing campaigns in specific. The related research question was: To what extent can neutralization theory be used to justify participating in digital disinformation acts?

As a whole, neutralization theory provided a useful framework for exploring the possible reasons for participation. However, future research into a few specific complications should be conducted in order to determine their validity as concerns.

The first of these being that victimisation failed completely in its attempts at providing a plausible excuse for the culprits.

This could be due to lacking proper personal connection to any of the provided 'victims.' Given the nature of an online marketing campaign, this sort of motive would be very complicated to provide within the confines of the story.

We must also remember that neutralization theory's most accepted usage

is explaining petty juvenile crimes and its most researched digital application is in regards to piracy or "victimless" crimes. Victimization in particular is a neutralization technique added by Bryant et al. after having analysed statements from people accused of being involved in a genocide. Thus, one option for the perceived implausibility of this neutralization technique could be that the scale of the crime present simply didn't provide grounds for such victimisation to occur that a third party would take action because of it.

Third possibility is that victimisation's failures relate to its status as a purely extrinsic motive – a topic, which we will now cover as another find of the study.

Similarly to victimisation, appeal to good character, another one of Bryant et al.'s addendums, was met with large resistance as well. The implications of this is hard to say, but it may be that both are inapplicable, but for different reasons. In the case of good character, it may be that the vignette failed in giving the interviewees proper understanding and attachment to Paul, making them unwilling to see this as a potential neutralization. Or maybe the questions laid out to find it were improperly formulated.

All in all, Bryant et al.'s neutralizations proved to be a disappointment in regards to creating plausible scenarios for the interviewees to consider.

For future studies on the topic, we recommend the usage of Muel Kaptein's (2018) model of neutralization, which provides a much more nuanced breakdown and categorization of various neutralization techniques.

The second find is that intrinsic motives were much more likely thought to be plausible excuses than extrinsic motives. This trend persists even within the same neutralization technique, so long as the angle makes the reasoning extrinsic or intrinsic. The clearest case of this is within 5.5.1, appealing to higher loyalties, where interviewees who saw Paul's and Jessica's interactions as platonic perceived the reasoning as more improbable than if romantic feelings were thought to be involved on Paul's front, in which case the excuse given was treated as something more plausible.

With the previous find, it may be necessary to, at least, diversify neutralization techniques into extrinsic and intrinsic, even within the same technique. In addition to this, it may be necessary to find a classification or otherwise further elaborate on the motives that didn't fit into either of these categories.

Another find in the same topic is the thought that some neutralisations, while plausible, were felt to be incomplete. We should include at least appeal to good character and maybe denial of injury to this category. Of course, this may be due to failures of the vignette or questioning and the validity of this find needs to be vetted in the future.

Finally, a trend among the neutralisations could be found, in that many of them were felt as opportunistic. This category can be shortly described as the

'Because I can' category.

M1 - Just because you can, doesn't mean you should.

This category doesn't have a clear division in its believability and plausibility and since we don't quite have another opposing category to compare it to either, we can't be certain if this classification has any merit in the first place.

What needs to be done with this category of neutralization technique, in addition to the intrinsic/extrinsic split we proposed previously, remains to be seen.

Overall, neutralization theory proved moderately successful in exploring the justifications workers participating in digital disinformation campaigns could give. While the possibility of faults was found in some neutralization techniques, with victimisation rejected outright, we find that all other neutralization techniques have plausible avenues. What is apparent is that low-stakes petty crime was thought to be fairly neutralizable.

As expected, the primary reason, and what many believed to be the underlying reason, regardless of the excuse by Paul, was thought to be monetary with other implicit selfish motives trailing close behind. This leads us to believe that the perception of the individuals participating in digital disinformation campaigns is that they are selfish. In a similar fashion, people felt callous reasons, such as denial of responsibility or injury, were also plausible, albeit to a lesser degree. This further builds a picture of the workers' perceived personality, or maybe just Paul. Even bigger reinforcement of this is how nearly all of the interviewees felt that "righteous" reasons, such as victimisation and appeal to higher loyalties, when the loyalty was to another person, as well as any excuses relating to law or morality itself were thought to be very implausible and unlikely. The comments relating to the absurdity of condemnation of condemners can be felt here as well, where it was thought that workers participating in these campaigns weren't perceived to have such lofty ideals of changing rules, society and/or morality itself.

However, one of the interviewees in particular, F3, was very reluctant to consider any explanation plausible or acceptable. While discussing the acceptability of any neutralization theory was only a tertiary topic, the disapproval present in F3 was odd to see. They were against all attempts of neutralization. We didn't hold a separate discussion as to why this individual felt so strongly against neutralization techniques, but maybe it was because of them latching more strongly to malicious impressions of the campaign. While nothing concrete can be inferred from this, it is possible that a different framing of the vignette could provide different reactions and responses from the interviewees. There is an argument to exclude her answers from the study since they are so extreme with all but one of them being negative, with even that being due to extensive negotiation and discussion with the interviewee. Regardless, we decided to keep them in for future analysis.

6.1.2 MEECES

Here we discuss the applicability of the framework we proposed in the beginning of the study, MEECES.

To recall, we are attempting to utilize MEECES motives, originally laid out for hackers, and seeing how they function in digital disinformation contexts. Our attempt was to utilize MEECES to answer one of our original research questions: What drives these ordinary people interested in crowdsourcing projects to join digital disinformation campaigns?

While we can't produce definite answers from the study, we can find commonalities between some neutralization techniques and MEECES motives. This section is dedicated to theorizing the possible linkage of MEECES motives and neutralization techniques. From this, we determine the plausibility with which a particular MEECES motive is for an individual disinformation worker or, at least, this fictional Paul.

Money as a motive was thought to be very plausible (5.5.5., 'I did it for the money', 8/9) and, thus, we see no reason to doubt its inclusion or usage.

Ego manifests as appeal to higher loyalties. This combination stems from the neutralization technique users claiming themselves to be above the rules because of a higher loyalty. Combining the two, we could explore various ways this 'higher loyalty' can manifest, namely, the ways one's personal beliefs of superiority or 'being above the rules' can manifest. While money is listed under higher loyalties as well, we are excluding it from the discussion about ego, as well as the topic of being unable to say 'no' to a friend, which might be the only neutralization technique fitting with entrance as a motive.

Finding neutralization questions that match with ego, we identify a few from appeals to higher loyalty with 'future career' (5.5.2., 6/8) and 'can't say no' (5.5.1., 3/9). In addition to this, we believe appeals to good character fit ego rather well with 'appeal to entitlement' (5.6.2., 3/9). However, if we look over to 5.4.3 to find the neutralization with 'it's the rules that are wrong' (5.4.3., 1/9), we can see a stark denial of it with only 1/9 considering it plausible.

If we explore the other side of ego, where one wants to rise up to the challenge presented, we can find similarities from 'appeal to interest' (5.5.3., 3/7) and 'can't say no' (5.5.1., 3/9). While inconclusive, we can at least say that the numbers do not favour this side of ego either.

While ego seems like a plausible motive for participation, it is only to the extent where the worker believes their actions to be above the law, rather than law itself being wrong, and even then, the perception is very much on the implausible end of the scale.

Looking at entertainment as a motive, we can find examples of it from appeal to higher loyalties with 'fun' (5.5.4., 4/8) and 'interesting' (5.5.3., 3/7).

One might argue that Paul's intention never was to provide entertainment to anybody else, thus making entertainment a hard motive to examine in the first place, but his personal enjoyment is a part of this motive as well.

Overall, entertainment was accepted with middling results, tilting slightly to implausible, rather than plausible. While the two neutralization techniques we determined to best fit this category were indeed deemed plausible, their ratios are both on the lower end of the scale. As such, we believe entertainment to be an unlikely, yet plausible motive for the workers to hold.

Cause as a motive was thoroughly denied by all of the interviewees. From 5.7 victimisation, we find two possible causes with 'victimisation of employees' (5.7.1., 0/8) and 'victimisation of the company' (5.7.2., 1/9). We can also consider 'attack on monopoly' (5.4.2., 3/8) from condemnation of condemners, but this doesn't improve cause's position as a motive substantially. A similar trend continues if we explore 'it's the rules that are wrong' (5.4.3., 1/9). It was generally considered unreasonable for workers to use any of these excuses, since they lacked any personal connection with the topic, and thus it was deemed unreasonable for them to hold such views.

With this, we believe we can make the claim that cause is very unlikely as a motive for digital disinformation workers, if at all.

As mentioned before, entrance as a motive is dubious to begin with and could very well be included in status. If we were to find any questions of ours that match even remotely with entrance, it would be 'couldn't say no' (5.5.1., 3/9). Interestingly, in the cases where the motive for entrance was lower, this being loyalty to a friend, the plausibility was thought to be very low (1/7). However, when entrance was implied by the interviewees, such as a possible future romance, the excuse was thought to be very plausible (2/2). Obviously, the data is limited here, but it may be a possible avenue that affirms entrance as a possible motive, with digital disinformation simply being a difficult context to explore or manifest it in. It may be necessary to further test the applicability of this by utilizing the same excuse in 'inability to decline request', but changing the status of the person Paul couldn't say no to. If this person was already their significant other, would the ratios change? If it was explicitly stated that Paul was interested in Jessica, how would that affect the results? What if Jessica was close family? These are all questions we couldn't explore in this study but should provide an interesting future research prospect.

That being said, status itself is a very difficult motive to identify or explore in digital disinformation contexts as well, especially because of the covert nature of it and possible legal ramifications taken against Paul should he brag or confess his actions. This section then is dedicated to exploring this topic to the extent we can.

If entrance was deemed as the hopefulness of a new relation, status could then be thought as the betterment of this relation. This is how Max Kilger

originally intended these two motives to interact in MEECES, with entrance being the motive to enter into a relation or circle and status being the drive to constantly improve your position in that relation or circle.

As such, if we are to consider this side of 5.5.1. 'can't say no' where Paul's motivation was to look better in the eyes of Jessica, whether that is friendship or romantic, we find that our data is inconclusive. While the neutralization itself was thought to be fairly improbable, but if we are to divide it, again, among the interviewees who thought Paul's actions were of hopeful romance (2/2), rather than betterment of friendship (1/7), we find that the motivation that better relates to improvement of one's status had a much worse response than the one that corresponds to entrance. Of course, these two are difficult to directly compare since one relation is of friendship and the other is of romantic involvement. As such, our findings are inconclusive when it comes to status and entrance as a motive.

It may very well be that the more loftier and intrinsic the ambition of the status or entrance is, the more accepted the neutralization was thought to be. To note, this a very hypothetical claim, supported only by a conjecture at this point.

Unfortunately, we can't derive any further evidence for or against status as a motive. Together with the ambiguity regarding friendship/romance as well, we must deem the position of status as a motive inconclusive as of this study. It may have been possible to explore this motive from a worker's standpoint if we'd included a question relating to how Paul wanted to increase his social media presence through working in the campaign, for example. The question relating to the future applicability of Paul's experiences in 5.5.2., appealing to higher loyalties, probably could've included a discussion on this, or maybe replaced altogether.

With this we've expended the limits of MEECES when it comes to this study. Unfortunately, there are still many neutralisations within our questions still unaccounted for. However, since many of these are attempts at downplaying by Paul, it may be that they are not motives in the first place, but rather pure attempts at neutralization, and, thus, these could be used as an excuse attached to any of the other motives in MEECES.

All in all, MEECES provided a relatively satisfactory exploration into the motives of individual digital disinformation workers. While some of the connections between MEECES and neutralization theory are contrived at best and many of the explorations thread the same ground multiple times, we believe to have drawn some insight from this comparison, regardless of how applicable it is in future research on this topic. In addition, we can't conclusively determine whether these motives encompass all the possible motives of participation, at least, as of this study, they stand.

6.2 Practical implications

We believe one of the key findings to come from the discussion during the interviews to come from the discussions themselves, rather than any yes or no answer to the plausibility of a neutralization technique. The most profound find here is the clear pessimism towards any discourse in social media as well as the lack of knowledge about the legality of digital disinformation, with a surprising amount of interviewees believing undisclosed marketing and deceit were allowed and/or lawful.

We explored the data related to this in the introductions chapter where multiple sources found that trust in media and digital environments was extremely low. What could be the cause of this? Is it the prevalence of news stating that disinformation is numerous causing the idea that, since it is so prevalent, it must mean that either laws are lagging behind technology or the platforms themselves don't care enough? Or is it the opposite, in that a layperson doesn't follow developments in the social media landscape to such a degree that they don't suspect the legality of fake news? This thought is quite troubling, and considering one interviewee had been convinced into performing actions similar to Paul already. This lack of knowledge into the legality of crowdturfing and similar digital disinformation acts may very well be among reasons how workers eventually join in on these campaigns; they simply do not know that what they are doing is illegal.

But how can this be possible? In 5.1.2, *"I didn't realize this wasn't allowed"*, the interviewees who disagreed about the plausibility of this technique mostly did so because of the "obvious" moral implication, despite nearly all of them, save for F1, postulating in one way or another that legally Paul's actions probably were allowed. What is it that has created this divide?

While we can only theorize on the why, we can create practical applications based on this divide. Social media platforms and digital news clearly are not doing their part in notifying the average user about what is allowed on their platforms. Of course, while arguments can be made that many of social media's rules are arbitrary and interpretive, this rule, at the very least, is not. Twitter terms of service, among other social media platforms, outright bans third-party advertisement. But is this enough? It's unlikely that an average user bothers exploring the rules to such an extent that they would find this, while at the same time they are bombarded with false information or, at least, think they are. How could it not be legal and allowed if it is everywhere already?

Whatever the case, public knowledge of what's allowed needs to improve. In addition, individual disinformation workers need to be held responsible for the disinformation acts they perpetrate. Most legal cases are held between the customer and their target victim, for example in the Samsung vs. HTC case, which the vignettes are based on, the lawsuit is solely between

Samsung and HTC, and not the individual workers and the users being exposed to their disinformation. The businesses are punished and their actions brought to light while the workers, the actual perpetrators, move onto the next campaign, none bitten and twice as bold. In addition, the actual victims of their acts remain oblivious to the disinformation they were subjected to unless they manage to accidentally stumble onto a news report of the campaign and somehow link their experiences and the report. How this eventually is achieved fairly, remains to be seen, if it is possible in the first place.

Another practical implication was the clear pessimism about social media.

F1: (When speaking about people believing in social media) If they do, they ought to stop.

M2: I mean, probably half of five and one star reviews are fake anyway even if they weren't paid to do it.

M3: Marketing is, at some level, always dishonest.

M4: There's probably a loophole somewhere that allows these campaigns to exist.

M1: Maybe it is the future of advertising, but I sure hope it wasn't.

F4: (When speaking about crowdturfing being the future of advertising) Yeah. It is.

Whatever the actual findings of this study were, it pales in comparison to the finding of this pervasiveness of pessimism and distrust towards digital spheres.

Bella M. DePaulo's (1996) study on everyday lying, especially in its conclusion, where she determines that all interactions are thought to be more positive if everybody involved is honest. Just how much could be gained if the prevailing thought of social media was that people were honest there? From personal recollection, the internet has always been rampant with dishonesty and deceit, or at least thought of being full of it. The comments some interviewees had where they didn't believe anything read in social media was very typical as well as disheartening, since most of the interactions people face in this environment are with honest users honestly speaking their mind. Yet, one can find these honest interactions being dismissed as 'trolls' or 'shills' everywhere. Browsing, for example, reddit's comment sections, if one finds any perceived evidence, any hint, no matter how small, of another commentor's infidelity towards 'your truth', they are immediately dismissed and ostracized from the discussion. This sort of activity can only lead to insular bubbles and polarisation (Wendy Gould, 2019). And it's not that an individual is responsible or even the reason why they are in a bubble. Social media algorithms are created in such a way that people find their best 'bubble', since it drives retention and reduces friction (Kristen Allred, 2018).

In the end, then, maybe social media and the internet are not a hapless victim of digital disinformation, as first postulated in the introduction, but the opportunistic, careless mad scientist who now has to fight a monster of their own making. One that, unfortunately, none of us can quite truly avoid.

6.3 Future research directions

For future research directions we are both bringing forth new areas of research we've established in this study, as well as dedicating discussions for topics we couldn't tackle but very well could've been in a study of this type.

Firstly, to recap neutralization theory, we chose seven neutralization techniques to create questions around for our interviews. As one can see, our choice of utilizing Bryant et al.'s (2018) addendums was not the most productive path we could've taken. Similarly, some of the topics that we eventually placed under appeal to higher loyalty, very well could've found better matches from other neutralization techniques proposed over neutralization theory's long history. It would be interesting to see what information can be derived from asking questions based on the neutralization techniques we didn't include in this research. These are, for example, metaphor of ledgers (Klockars, 1974), the defence of necessity (Minor, 1981), the claim of entitlement and the claim of normalcy (Coleman, 1985), the denial of negative intent and the claim of relative acceptability (Henry, 1990), the justification by comparison and the justification by postponement (Cromwell & Thurman, 2003), and many others. These addendums have yet to receive their due time in empirical research contexts, especially in digital spaces, and, as such, testing their applicability could provide interesting insights.

Furthermore, conducting the same study with more refined questions could easily provide us with new and more accurate insight. For example, 'inability to decline' could be split based on the people you can't decline (friendship, romance, family, boss, etc.) to provide further insights. Similarly, properly testing the difference we saw in explicit and implicit perceptions of motives should provide us with more refined insight into the possibilities of this hypothesis.

MEECES as a framework of motives provided us with a good starting point, but its proper applicability can't be judged quite yet. More research should be conducted utilizing it as a basis to see how well it fits into crowd-turfing worker contexts and refining it for digital disinformation specifically if needs be.

As for topics that stem from our findings, we believe the pessimism and distrust in digital contexts to be a worrying discovery. As a find this only affirms the trend many of the foundations in the introduction section already found through their Gallup polls (Edelman, 2018 & 2019, Knight Foundation, 2018), but our study found this in face-to-face interviews. Further research

needs to be done to pinpoint the reasons as to why people feel digital media to be untrustworthy. We can't simply be satisfied with the knowledge that it exists, since modern public discourse is conducted in these digital forums and having so many consider the entire exercise dishonest and fake can easily drive away people that would otherwise provide useful insight and angles to the discussions.

While there are many other future research implications one can find from this study alone, we believe these four be the most pertinent paths to take.

- 1) To further examine neutralization theory as a means of exploring workers in digital disinformation contexts
- 2) Split some of the questions presented in this study to examine
 1. The difference in how people perceive implicit and explicit motives
 2. The difference in how people perceive worker loyalties to different possible parties (friendship, romance, family, boss, etc.)
- 3) To further determine the usefulness of MEECES as a framework of motives
- 4) To figure out the pervasiveness to which distrust in digital contexts extends as well as finding out solutions to it

6.4 Limitations

This study was conducted during a fairly tense period of 2020 with covid-19 shutting down many countries, Finland included. While everything in this study was done in accordance to, at the time, present guidelines, the added tension may have exasperated some of the opinions of the interviewees. This may possibly be why so many saw digital contexts as pessimistically as they did; anyone would if that was their only human interaction in half a year.

In addition, many of the interviewees were friends or acquaintances of the researcher, giving the interviews a much more cordial atmosphere than what one can expect from interviewing strangers.

Furthermore, the freeform nature of the interviews may have caused some errors in the data if an interviewee became more receptive to Paul's actions over the course of the discussion, thus making them more likely to answer positively to a question they wouldn't have otherwise. And because of the freeform nature, one can't tell which answers have this possibly positive-slant compared to other answers. As a general rule, once discussion topics were exhausted, the neutralization techniques were asked in the order presented in

this study, should they have been skipped or not yet talked about. Of course, this didn't help the case of appeal to good character or victimisation, but the possibility exists. Solution to this would be to have more interviews with less questions to dilute the possibility of this problem.

7 CONCLUSIONS

We set out to examine people's perceptions on the actions of crowdturfing workers utilizing neutralization techniques as our method of exploration. Not only did this provide us with a good framework to build questions and discussions around, these discussions around it provided us with legitimate points and findings both in regards to people's perceptions of the workers as well as how they personally feel about a multitude of topics. The key finding here is the affirmation of the general pessimism and distrust towards digital contexts.

Neutralization theory proved very plausible in many cases, with the exception of appeal to good character and victimisation, which appeared unlikely and implausible, respectively, in the eyes of the interviewees. Further research should be done with questions built around other neutralization theories, especially the new, modern refinement in Muel Kaptein's (2018) models of neutralization.

Further research is needed to determine MEECES as a framework of motivations. We utilized it to relative success but some aspects of it seem very fine-tuned for its original context in hackers. While it provides a starting point, we believe something more fitting can be created or found.

8 REFERENCES

- Edelman Foundation, 2019 Edelman Trust Barometer, edelman.com, 2019
- Edelman Foundation, 2019 Edelman Trust Barometer, edelman.com, 2018
- Patricia Barnes, Add This To The List Of Crises - Lack Of Trust In U.S. Media, forbes.com, 2 October 2020
- Knight Foundation, AMERICAN VIEWS: TRUST, MEDIA AND DEMOCRACY, knightfoundation.org, 2018
- Jeff Horwitz and Deepa Seetharaman, Facebook Executives Shut Down Efforts to Make the Site Less Divisive, wsj.com, The Wall Street Journal, 26 May 2020
- Kyumin Lee, Crowdturfers, Campaigns, and Social Media: Tracking and Revealing Crowdsourced Manipulation of Social Media, Department of Computer Science, Utah State University, 3 April 2014
- Chan, C. 2012. How a fake erotic fiction ebook hit the top 5 of itunes, gizmodo.com, 8 September 2012
- Nga Pham, Vietnam admits deploying bloggers to support government, BBC.com, 12 January 2013
- Cheng Chen et al., Battling the Internet Water Army: Detection of Hidden Paid Posters, Cornell University, 18 November 2011
- Wang, G.; Wilson, C.; Zhao, X.; Zhu, Y.; Mohanlal, M.; Zheng, H.; and Zhao, B. Y. 2012. Serf and turf: crowdturfing for fun and profit.

- J. Ratkiewicz, M. Conover, M. Meiss, B. Gonçalves, A. Flammini, and F. Menczer. Detecting and tracking political abuse in social media. In Proc. 5th International AAAI Conference on Weblogs and Social Media (ICWSM), 2011.
- V. Subrahmanian, A. Azaria, S. Durst, V. Kagan, A. Galstyan, K. Lerman, L. Zhu, E. Ferrara, A. Flammini, F. Menczer, A. Stevens, A. Dekhtyar, S. Gao, T. Hogg, F. Kooti, Y. Liu, O. Varol, P. Shiralkar, V. Vydiswaran, Q. Mei, and T. Hwang. The darpa twitter bot challenge. *IEEE Computer*, 49(6):38–46, 2016.
- A. Bessi and E. Ferrara. Social bots distort the 2016 us presidential election online discussion. *First Monday*, 21(11), 2016.
- European Commission (2018b) Communication on Tackling Online Disinformation
- European Commission (2018a). A multi-dimensional approach to disinformation. Report of the independent High level Group on fake news and online disinformation. March 2018.
- Jordan Valinski, Facebook Is Testing a 'Satire' Label for 'The Onion', *Gizmodo.com*, 16.08.2014
- Lee de-Wit, Cameron Brick & Sanders van der Linden, Are Social Media Driving Political Polarization?, *greatertgood.berkeley.edu*, 01.16.2019
- Amazeen, M. A., & Wojdyski, B. W. (2018). The effects of disclosure format on native advertising recognition and audience perceptions of legacy and online news publishers. *Journalism*. doi:10.1177/1464884918754829
- Wardle, C. and H. Derekshan (2017) Information Disorder: Toward an interdisciplinary framework for research and policy making, Council of Europe report, DGI(2017)09.
- M.J. Hindelang. The Commitment of Delinquents to their Misdeeds: Do Delinquents Drift?. *Social Problems*, 17: 502–509, 1970
- M.J. Hindelang. Moral Evaluation of Illegal Behaviors. *Social Problems*, 21: 370–385, 1974
- R. Agnew. The Techniques of Neutralization and Violence. *Criminology*, 32(4): 555–580, 1994

- J. Mitchell and R. Dodder. Types of Neutralization and Types of Delinquency. *Journal of Youth and Adolescence*, 12(4): 307–318, 1983
- R.L. Austin, editor, *Commitment, Neutralization, and Delinquency*. Sage, California, 1977
- W.W. Minor. Techniques of Neutralization: A Reconceptualization and Empirical Examination. *Journal of Research in Crime and Delinquency*, 18(1): 295–318, 1981
- Yevgeniy Golovchenko, Mareike Hartmann, Rebecca Adler-Nissen, State, media and civil society in the information warfare over Ukraine: citizen curators of digital disinformation, *International Affairs*, Volume 94, Issue 5, September 2018, Pages 975–994
- Hinduja, S. Neutralization theory and online software piracy: An empirical analysis. *Ethics Inf Technol* 9, 187–204 (2007)
- BBC, Samsung fined for fake comments about competitors, 24.08.2013, BBC
- Soliman, Wael and Rinta-Kahila, Tapani, "UNETHICAL BUT NOT ILLEGAL: UNCOVERING THE PERSUASIVE MESSAGES LEVERAGED BY PROVIDERS OF THE "REAL" ONLINE SOCIAL IMPRESSIONS" (2018). *Research Papers*. 172.
- Nikolov D, Oliveira DFM, Flammini A, Menczer F. 2015. Measuring online social bubbles. *PeerJ Computer Science*
- Weedon, J., Nuland, W. & Stamos, A. *Information Operations and Facebook*. White paper, Facebook (2017).
- Allcott, Hunt, and Matthew Gentzkow. 2017. "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives*, 31 (2): 211-36.
- KALLA, J., & BROOCKMAN, D. (2018). The Minimal Persuasive Effects of Campaign Contact in General Elections: Evidence from 49 Field Experiments. *American Political Science Review*, 112(1), 148-166. doi:10.1017/S0003055417000363
- Soliman, Wael & Rinta-Kahila, Tapani & Kaikkonen, Joonas. (2019). Why Is Your Crowd Abandoning You? Exploring Crowdsourcing Discontinuance through the Lens of Motivation Theory.
- Bella M. DePaulo et al. (1996), *Lying in Everyday Life*, *Journal of Personality and Social Psychology*, 1996, Vol. 70, No. 5, 979-995

- Leiser M., "AstroTurfing, 'CyberTurfing' and other online persuasion campaigns", in *European Journal of Law and Technology*, Vol 7, No 1, 2016.
- CMA's call for information report into online reviews and endorsement (June, 2015) at Para 1.3
- Zajonc, Robert B. (1968). "Attitudinal Effects Of Mere Exposure". *Journal of Personality and Social Psychology*. 9 (2, Pt.2): 1-27. ISSN 1939-1315.
- Zajonc, Robert B. "Mere Exposure: A Gateway to the Subliminal." *Current Directions in Psychological Science* 10 (2001): 224 - 228.
- Zhuang, M. (2014). *Buyer beware: Consumer response to manipulations of online product reviews* (Master's thesis, Lingnan University, Hong Kong)
- Fallis, Don. "What Is Disinformation?" *Library Trends*, vol. 63 no. 3, 2015, p. 401-426. Project MUSE
- Damid M. J. Lazer, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rotshchild, Michael Schudson, Steven A. Sloman, Cass R. Sunstein, Emily A. Thorson, Duncan J. Watts, Jonathan L. Zittrain, *The Science of Fake News*, *Science* 09, MAR 2018: 1094-1096
- Emily Thorson (2016) *Belief Echoes: The Persistent Effects of Corrected Misinformation*, *Political Communication*, 33:3, 460-480
- The Honeynet Project, *Know Your Enemy: Learning about Security Threats*, Addison-Wesley Professional; 2nd Edition, May 17 2004
- Ronald E. Riggio, *5 Reasons Why Money Is a Poor Work Motivator*, *psychologytoday.com*, May 03 2018
- Storyful Team, *4chan Users Launch Disinformation Campaign Ahead of Syria Airstrikes*, *storyful.com*, April 18 2018
- Lillian Raduolva, *Putting your iPhone in the microwave will COOK it, not charge it: Apple urges users not to fall for convincing internet hoax... too late for many*, *dailymail.co.uk*, 25 September 2014
- Fernando Alfonso III, *4chan's fake iOS 7 ads convince users to dunk phones in water*, 2 March 2020

- Ellie Flynn, 'BIGGEST MISTAKE OF MY LIFE' Clueless iPhone 7 owners tricked into DRILLING hole in their phones to 'get a headphone jack', thesun.co.uk, 24 September 2016
- Bryant, Emily et al. "Techniques of Neutralization and Identity Work Among Accused Genocide Perpetrators." *Social Problems* 65 (2018): 584-602
- Weal Soliman, Tapani Rinta-Kahila, Joonas Kaikkonen, Why Is Your Crowd Abandoning You? Exploring Crowdsourcing Discontinuance through the Lens of Motivation Theory, Australasian Conference on Information Systems 2019, Fremantle, Western Australia
- Jenn Chen, The most important social media metrics to track, sproutsocial.com, 4 August 2020
- Christine Barter and Emma Renold, The Use of Vignettes in Qualitative Research, University of Surrey, Summer 1999
- Jenkins N, Bloor M, Fischer J, Berney L, Neale J. Putting it in context: the use of vignettes in qualitative interviewing. *Qualitative Research*. 2010;10(2):175-198.
- Muel Kaptein & Martien van Helvoort (2019) A Model of Neutralization Techniques, *Deviant Behavior*, 40:10, 1260-1285, DOI: 10.1080/01639625.2018.1491696
- Bertin Martens Luis Aguiar Estrella Gomez-Herrera Frank Mueller-Langer, The digital transformation of news media and the rise of disinformation and fake news, JRC Digital Economy Working Paper 2018-02, April 2018
- Adam Bienkov, Astroturfing: what is it and why does it matter?, theguardian.com, 8 February 2012
- Bradshaw, S. (2019). Disinformation optimised: gaming search engine algorithms to amplify junk news. *Internet Policy Review*, 8(4). DOI: 10.14763/2019.4.1442
- Sebastian Bay et al., The Black Market for Social Media Manipulation, NATO StratCom COE, Singularex, November 2018
- Wendy Rose Gould, Are you in a social media bubble? Here's how to tell, nbcnews.com, 21 October 2019
- Kristen Allred, the Causes and Effects of "Filter Bubbles" and how to Break Free, medium.com, 13 April 2018

- Coleman, J. W. (1985). *The criminal elite: The sociology of white collar crime*. New York: St. Martin's Press.
- Klockars, C. B. (1974). *The professional fence*. New York: Free Press.
- Minor, W. W. (1981). Techniques of neutralization: A reconceptualization and empirical examination. *Journal of Research in Crime and Delinquency*, 18(2), 295–318.
- Cromwell, P., & Thurman, Q. (2003). The devil made me do it: Use of neutralizations by shoplifters. *Deviant Behavior*, 24(6), 535–550.
- Henry, S. (1990). *Degrees of deviance: Student accounts of their deviant behaviour*. Salem: Sheffield Publishing.
- Claudia Wilhelm (2020) Investigating Neutralization Strategies in Digital Piracy: The Role of Content Preferences and Social Norms, *Journal of Broadcasting & Electronic Media*, 64:2, 320-340
- Higgins, G.E., Wolfe, S.E. and Marcum, C.D., 2008. Music piracy and neutralization: A preliminary trajectory analysis from short-term longitudinal data. *International Journal of Cyber Criminology*, 2(2)
- Marcum, Catherine D., George E. Higgins, Scott E. Wolfe and Melissa L. Ricketts. 2011. "Examining the Intersection of Self-control, Peer Association and Neutralization in Explaining Digital Piracy" *Western Criminology Review* 12(3):60-74
- Sixuan Zhang & Dorothy Leidner, *From improper to acceptable: How perpetrators neutralize workplace bullying behaviors in the cyber world*, *Information & Management*, Elsevier, November 2018

9 APPENDICES

Appendix 1: English Vignette

Paul is a 22-year old IT student. Overall he enjoys his university life. His parents weren't rich but they weren't poor either. Part of the reason why he is in a university is to get a high paying job.

He enjoys going out with his friends and playing video games in his free time. He is an active user in social media and vocal about his opinions. He would describe himself as influential, if only for the small group of followers he has. He regularly engages in public discourse and contributes in his own way. He has a hard time saying 'no' to good friends of his, but he doesn't think this is a bad quality of him.

Phion is a global mobile phone manufacturing company. If one asked Paul, they would put Phion into the top three in the world. Paul also recalls a rivalry between Phion and another phone company, Speak, but he doesn't know the details. According to Paul, Speak would also go into that top three companies.

One day, Paul's friend, Jessica, contacts him. She says that she's been making extra money by being part of an online social media advertising campaign and asks if Paul would be interested in it.

Paul finds the topic interesting. He's always been curious about the potential of social media for businesses and promises to look at the link Jessica provided.

The site was a publicly open promotion of an online campaign, where people with social media influence, big and small, could use their platform to provide advertise any given topic. The page Jessica linked was for a campaign to promote Phion. The site was similar to any other online gig market site such as Amazon turk and fiver or webstore Paul knew about, except all the sales were requests of various actions. This one read as follows:

"Join now for a unique opportunity to participate in a massive social

media campaign to promote Phion's new phone, Ax-8!

As you may know, Phion's biggest competitor, Speak is releasing their new phone, Selly, around the same time. The task of this campaign is to generate grassroots excitement for Phion!

Collect points by doing various online activities using your personal social media accounts and turn the points in for monetary rewards! Utilize your social media influence to its fullest potential to promote the sales of Phion and generate competitive advantage.

Details on the specifics of the campaign are revealed once the cooperation between us and your social media presence is confirmed.

P.s. Phion nor Speak are in no way related to this campaign."

Paul was confident about joining the project. He had put a lot of effort into his social media presence and felt that it was only right he was compensated for the work.

Upon joining the campaign, Paul signed a non-disclosure agreement, stating that revealing oneself as part of the campaign was forbidden and anybody who did would be disassociated and face legal repercussions. Paul thought it to be very normal for a work contract.

After signing the NDA, Paul was contacted by an organizer, who sent a list of tasks for him to complete.

Paul was thorough and completed the entire list, which asked Paul to follow, like, share and comment on various social media pages relating to Phion's new phone campaign.

He followed the social media pages of Phion on all platforms he was on as well as a few others, where he created a new account to do so.

He went through the posts by these pages all the way from Ax-8's reveal until the end of the campaign and liked, shared and positively commented on all of them.

As part of the campaign, Paul wrote five star reviews on many review sites for Ax-8, even though he doesn't own the phone. He'd normally have rated it as three stars.

Paul also posted a story on his personal social media page about the time he changed from a Speak phone to a Phion phone, when in reality, he never did such a thing.

Paul also read up on the conflict between Phion and Speak. In his opinion, Phion's grudge is reasonable, but doesn't think it's a big deal nowadays. However, he did include a news story of the event in a post for the campaign, thinking it made good material.

He happened to find out about an old case where a mine providing Speak with materials was abusing its employees and polluting the

environment. He made a furious post about it.

The next two weeks Paul posted stories and directly shared links about Phion and Ax-8, until the end of the campaign, when he reported all the numbers of his shares and his stories and collected his points. He was paid a modest amount for his contributions and, overall, he felt satisfied with the transaction. Maybe the payment wasn't quite relative to the work put in, but it was during his free time and if he'd have a bigger following, he'd get paid considerably more. Regardless of the amount, to a student like him, it was a considerable aid to his life.

During this time, Paul had started receiving increasingly curious messages from his friends and family, asking 'What's the deal with this constant Phion stuff?'

Paul mostly answered with "I'm just so excited!" but when he was asked about this face-to-face by his closest friend, Paul finally confessed that he was part of an online campaign. Paul also asked this friend to not tell anybody as Paul had signed a non-disclosure agreement upon joining the campaign.

This good friend joked that he would have to unfollow Paul on all sites if he was engaging in these types of activities.

Paul really would prefer for this to not happen. This good friend is close to him and they've been together a long time. Paul sort of understands his concerns, so he attempts to make an excuse.

Appendix 2: Finnish Vignette, Suomenkielinen Vinjetti

Pauli on 22-vuotias IT opiskelija. Kaikenkaikkiaan hän nauttii yliopisto elämästään. Hänen vanhemmat eivät olleet rikkaita eikä köyhiä. Yksi syy miksi hän meni yliopistoon oli saadakseen korkeapalkkaisen työn.

Pauli nauttii ystäviensä seurasta ja vapaa-ajallaan hän pelaa videopelejä, sekä käyttää sosiaalista mediaa. Hän on aina ollut aktiivinen käyttäjä ja vokaali mielipiteistään. Pauli sanoisi että hän on merkittävä, joskin vain pienelle määrälle seuraajistaan. Hän osallisuus yleiseen keskusteluun jatkuvasti ja tuo keskusteluun arvoa omalla tavallaan. Tästä huolimatta, hänen on vaikea sanoa 'ei' ystävilleen, mutta hän ei pidä tätä huonona puolena itsessään.

Phion on Maailmanlaajuinen matkapuhelin valmistaja. Jos Paulilta kysyttäisiin, hän pistäisi Phionin kolmen suurimman joukkoon maailmassa. Hän myös on kuullut pienestä kinasta Phionin ja toisen matkapuhelinvalmistajan, Speakin, välillä mutta hän ei tiedä tarkkaan mistä on kyseessä.

Eräänä päivänä, Paulin ystävä, Janiina, ottaa häneen yhteyttä. Janiina kysyy, onko Pauli halukas tienamaan hiukan rahaa ottamalla osaa uudenlaiseen sosiaalisen median markkinointiin. Pauli pitää aihetta

mielenkiintoisena, omasta takaakin. Hän on aina ollut kiinnostunut sosiaalisen median mahdollisuuksista yrityksille ja lupaa katsoa Janiinan antamaa linkkiä.

Linkkiä seurattaessa, Pauli löytää sivuston, joka mainostaa avoimesti digitaalista kampanjaa, missä ihmiset voivat käyttää sosiaalisen median tilejään vaikuttamaan erinäisissä mainoskampanjoissa. Tämä kyseinen kampanja koskee Phionin uutta puhelinta. Sivusto luki seuraavasti:

”Liity nyt Phionin uuden puhelimen Ax-8 mainoskampanjaan!

Kuten saatatte tietää, Phionin suurin kilpailija, Speak, on myös julkaisemassa uutta puhelinta, joten tarvitsemme jokaisen kampanjaamme! Tarkoituksena on luoda ruohonjuuritason toimintaa!

Kerää pisteitä suorittamalla monenlaisia aktiviteetteja sosiaalisessa mediassa käyttämällä omia henkilökohtaisia tilejänne! Vaihda pisteet tällä sivustolla rahapalkintoihin!

Tule mukaan käyttämän koko sosiaalisen medianne potentiaali meidän kanssamme!

Kerromme lisää yksityiskohdista kun ilmoittautumisena on suoritettu.

P.s. Phion eikä Speak ole mitenkään yhteyksissä kampanjan kanssa.”

Pauli näki mahdollisuuden projektissa. Hän oli pistänyt paljon aikaa ja vaivaa sosiaaliseen mediaansa ja näki että hänellä oli oikeus tienata sen avulla.

Liittyessään, kampanja organisoija otti Pauliin yhteyttä ja hän sai listan mahdollisista toimista ja miten niistä pystyi saamaan pisteitä. Pauli kävi listan läpi perinpohjaisesti.

Hän seurasi, tykkäsi, jakoi ja kommentoi kaikki mahdolliset Phionin uuteen puhelimeen liittyvät kampanja sivustot ja mainos palstat. Hän tilasi kaikki Phionin uutislehdet ja sivustot. Jos hänellä ei ollut kyseisessä sosiaalisen median sivustolla tiliä, hän loi sen, jotta pystyi suorittamaan tehtävät.

Hän meni takaisin aikalinjalla ja teki saman kaikille postauksille Ax-7 julkaisusta aina Ax-8 kampanjaan asti, positiivisesti kommentoiden kaikkia julkaisuja.

Pauli kirjoitti 5-tähden arvosteluja monille sivustoille Ax-8:sta, vaikka hän ei puhelinta omistanutkaan. Paulin mielestä puhelin oli ehkä 3-tähteä normaalisti.

Pauli teki pitkän tarinan, kuinka vaihtoi Speak puhelimesta Phion puhelimeen, vaikka todellisuudessa hän ei ole tehnyt mitään tällaista.

Pauli sattui löytämään vanhan artikkelin kuinka Speakin alihankkijan kaivoksella oli tapahtunut onnettomuus, josta seurasi hieman päästöjä ja yksi kuolema. Pauli kirjoitti pitkän ja vihaisen tarinan kyseistä tapahtumasta.

Kampanjan lopussa hän raportoi takaisin hänen saavutuksensa ja keräsi pisteillä raha palkinnon. Hän sai vaatimattoman summan toiminnastaan,

mutta loppujenlopuksi, Pauli oli tyytyväinen osallistumiseensa. Ehkä saatu palkka ei ollut täysin työn vertainen mutta jos hänellä olisi ollut suurempi seuraaminen sosiaalisessa mediassa, olisi saatu palkkakin ollut sen verran arvokkaampi. Kaikesta huolimatta, hänen kaltaiselle opiskelijalle, tämä tienesti oli hyvä etu ja hän tarjosi kierroksen ystävilleen seuraavana lauantaina.

Kampanjan aikana, Pauli alkoi saamaan entistä enemmän hämmästyneitä viestejä sukulaisiltaan ja ystäviltään. "Miksi kirjoitat jatkuvasti Phionista?"

Pauli suurimmaksi osaksi vastasi että "Olen vain niin innostunut!", mutta kun hänen lähin ystävä kysyi asiasta kasvotusten, Paul viimein tunnusti olleensa osana digitaalista mainoskampanjaa.

Tämän kuultuaan, tämä hyvä ystävä vitsaili että hänen täytynee lopettaa Paulin seuraaminen sosiaalisessa mediassa jos tämä on sen sisältö.

Pauli ei halua tämän tapahtuvan, ja jollain tasolla ymmärtää ystävänsä huolenaihen, joten hän koettaa antaa syyn miksi teki mitä teki.

Appendix 3: Denial of Responsibility 1: "One more participant doesn't affect anything."

Tag	Line	Plausibility
F1	I can see why somebody would say and think that, but	Y
F2	Yeah, you need many to get anything done. That's why they don't pay you much.	Y
F3	Even a single person can make a huge difference. Paul should realize what he was doing and how many he was affecting.	N
M1	It really doesn't, yeah, that's true.	Y
M2	It all starts somewhere.	Y
M3	Yeah, and everybody else thinks the same	Y
M4	Just think it from the point of view of the companies; if it affected nothing, they wouldn't be doing it.	Y
F4	Yeah, but it's still not nice.	Y
M5	There's probably many who think that.	Y

Appendix 4: Denial of Responsibility 2: "I didn't know it wasn't allowed."

Tag	Line	Plausibility
F1	IT student can't possibly think that.	N
F2	Wait, you're saying this isn't allowed?	Y

F3	No matter what the social media sites say, you should realize at some point that deceiving people isn't right.	N
M1	There are ways to make yourself... legitimate doing this. Like what? Like, if you went out of your way to research the matter and gave somewhat honest opinions. That's the main issue, isn't it?	Y
M2	Legislation always lags behind technology.	N
M3	Marketing is, at some level, always dishonest.	Y
M4	There's probably a loophole somewhere that allows these campaigns to exist.	N
F4	That's all these people need to do (disclose that they are marketing).	N
M5	While not something I'd believe, I can imagine people saying that as an excuse.	Y

Appendix 5: Denial of Injury 1: "It's not a big deal."

Tag	Line	Plausibility
F1	Nobody should be believing social media anyway at this point.	Y
F2	One person can get a lot done.	Y
F3	Deceiving even your friends and family is always a big deal.	N
M1	Paul doesn't really have a lot of influence anyway.	Y
M2	I mean, probably half of five and one star reviews are fake anyway even if they weren't paid to do it.	Y
M3	Individually, each thing Paul did probably wasn't a big deal, but he did a lot of things for this campaign.	N
M4	Maybe not Paul, but others in that campaign probably had more influence.	Y
F4	It was a big deal to that person who was deceived.	N
M5	What's one guy anyway?	Y

Appendix 6: Denial of Injury 2: "It's digital so it doesn't matter."

Tag	Line	Plausibility
M1	You can always just leave and block people like Paul.	Y
M2	Probably? I dunno. That's a weird thing for an IT student to	Y

	say.	
M3	This coming from a guy who spends a lot of time with his social media account?	N
M4	Yeah, but he's trying to sell a physical thing.	Y
F4	If I found out my friend was lying like that, I would be sad.	Y
M5	This isn't just some internet trolling though. Paul is trying to make these people take real world actions.	N

Appendix 7: Denial of Injury 3: "Nobody who could be hurt was hurt if they were fooled to buy a phone this way."

Tag	Line	Plausibility
M3	You probably have the money to lose if you can just buy a phone without doing your own research.	Y
M4	So, Paul can't exactly be lying about the functions of the phone and whatnot, but like, maybe the sum of its parts doesn't come up to 5 but a 3.	Y
F4	But I'd still feel bad about it.	Y
M5	I mean, it's not a small loss, but I'm not going to cry over it.	Y

Appendix 8: Denial of the Victim 1: "What I did affected nobody."

Tag	Line	Plausibility
F1	I'm sure he affected something, but I can see why Paul would think otherwise.	Y
F2	I don't think he did.	Y
F3	Paul wasn't targeting strangers though, not that it makes it any better.	N
M1	Maybe I wouldn't run out buying one, but if I had to make a choice, I would be reminded by that post from a friend in my feed way back.	Y
M2	Yeah, but it's not like you can buy that many (reviews).	Y
M3	Would I buy a phone if a friend I trusted recommended it in social media? ...Damn, I think I might.	Y
M4	If that was the case, nobody would do it.	N
F4	Yeah, one doesn't, but there's many of these Pauls doing this.	N
M5	Towards strangers, Paul's nothing but a number to swell statistics, but to his family and friends he was somebody	Y

	they trusted... Although, I'd get suspicious if my friend started only talking about Samsung for weeks and weeks.	
--	---	--

Appendix 9: Denial of the Victim 2: "Nobody believes what they see in the internet anyway."

Tag	Line	Plausibility
F1	If they do, they ought to stop.	Y
F2	Yeah, I don't think they do.	Y
F3	So you'll just go intentionally causing more harm?	N
M1	I guess you should always watch a video at least of something as expensive as a phone.	Y
M2	And it's because of people like Paul why it is so.	N
M3	Nope. Not believing he would say that. The dude's trying to make a career or something from social media.	N
M4	I guess, but that's still a scummy thing to do.	Y
F4	Why did you have to tell me all this?	N
M5	I guess a healthy dose of suspicion is fine, but I doubt Paul is thinking he's doing a service or anything like that.	Y

Appendix 10: Condemnation of the Condemners 1: "It's the consumer's fault for being fooled."

Tag	Line	Plausibility
F1	It's true that media literacy is important, but you still shouldn't go about making the situation worse.	Y
F2	I can see that being the case, but surely nobody would try excuse themselves saying it.	N
F3	No, that's just not right. Not at all.	N
M1	I may be able to tell advertisement from normal posts, but that doesn't mean older people and more gullible people don't exist.	Y
M2	Going for peak capitalism award, huh?	N
M3	Yeah, but scamming people is still illegal.	Y
M4	That's a real shitty friend I have there in Paul.	Y
F4	That's really not nice.	N
M5	Of course. Nobody can force the money out of you, so they gotta cheat you a little.	Y

Appendix 11: Condemnation of the Condemners 2: "The only reason this is thought to be wrong is because social media platforms want to keep their advertising monopoly."

Tag	Line	Plausibility
F1	No, and I don't think anybody would argue that either.	N
F2	Umm... No and yes?	--
F3	I'm pretty sure I'd still think lying was bad even if facebook didn't tell me so.	N
M1	Yeah, that's fair. They want to keep their source of revenue to themselves.	Y
M2	Sorry, once more.	N
M3	There's probably some truth to that, but nobody is going to say that as an excuse.	N
M4	In a way, sure, but not exactly.	Y
F4	Sorry, can you repeat that?	N
M5	I don't see why they (social media platforms) would let them (crowdturfing aggregates/Paul) freely profit from their efforts.	Y

Appendix 12: Condemnation of the Condemners 3: "It's the rules/society/morality that's wrong."

Tag	Line	Plausibility
F1	No, absolutely not.	N
F2	No, I don't think so.	N
F3	No reasonable human being is going to say that about deception.	N
M1	Unlikely.	N
M2	I don't think anybody would say that.	N
M3	They are in many places, but here I want them right where they are.	N
M4	I don't think Paul would think that, much less say it aloud.	N
F4	No, I don't think that is the case at all.	N
M5	I don't think it's realistic, but plausible? Sure.	Y

Appendix 13: Appealing to Higher Loyalties 1: "I couldn't say 'no' to Jessica."

Tag	Line	Plausibility
F1	You should really use your own judgement.	N
F2	You can always say no.	N
F3	Just say no to that stuff.	N
M1	I know you didn't write it in the story, but is Paul interested in Jessica?	Y
M2	It's not like Jessica was in danger or anything.	N
M3	I think I know the reason why Paul couldn't say 'no'.	Y
M4	I don't think a friend's request warrants such involvement.	N
F4	If a friend asks you to do a crime, you should stop that friend.	N
M5	It does give the campaign a bit of legitimacy, so I get why he'd join it, at least.	Y

Appendix 14: Appealing to Higher Loyalties 2: "This is the future of advertising. We're ahead of the curve."

Tag	Line	Plausibility
F1	Doesn't make it okay now.	N
F2	I guess he was studying things like this the whole time.	Y
F3	Don't go about making a future you don't want to see.	N
M1	Maybe it is the future of advertising, but I sure hope it wasn't.	Y
M2	Everything similar has been regulated before already, so I don't see how this wouldn't be either.	N
M3	It probably is, unfortunately.	Y
M4	Yeah. Cheap way to get tons of exposure.	Y
F4	Yeah. It is.	Y
M5	I suppose that's fitting for an IT student.	Y

Appendix 15: Appealing to Higher Loyalties 3: "I thought it would be interesting. It is important for my studies."

Tag	Line	Plausibility
F2	Yeah, you gotta know what's going on.	Y
F3	That's not what a researcher does.	N

M1	I guess you gotta know that if you're part of the industry nowadays.	Y
M2	Yeah, sure, curious enough to commit a crime, c'mon.	N
M3	Good point. Somebody's gotta find out about it to combat it.	Y
M4	There's gotta be a better way of getting this knowledge.	N
F4	Crime's still a crime.	?
M5	I don't know. Can you do that?	N

Appendix 16: Appealing to Higher Loyalties 4: "I thought it would be fun."

Tag	Line	Plausibility
F2	Was it fun? It was work when I did it, but I guess they made it sound fun.	Y
F3	I don't see how anybody would say that.	N
M1	This campaign involved Paul doing this to his friends and family, right? Not cool.	N
M2	Like trolling and stuff? I wouldn't want to hear any more excuses from this Paul.	N
M3	If I was that friend talking to Paul now, I'm probably somewhat likeminded. But here you have the friend disapproving Paul. Sure, but I could laugh at a distasteful joke, especially if Paul is being paid for doing it.	Y
M4	Unfortunately, I can see somebody saying that.	Y
F4	I don't know. I don't think Paul's a good person if he thinks this is fun.	N
M5	Maybe the initial pitch sounded fun and Paul never got to second guess himself.	Y

Appendix 17: Appealing to Higher Loyalties 5: "I did it for the money."

Tag	Line	Plausibility
F1	Yeah. If the pay is good.	Y
F2	I can see that. Here in India, you find a lot of people taking part in these.	Y
F3	No amount of money warrants Paul's involvement. Surely there's a price to everything and Paul found his. And this is about what excuse you think Paul is likely to use. I know... Maybe then, if Paul is that type of person.	Y

M1	I think you'd need a bigger reason to take part in the whole thing.	N
M2	Look, if Paul just said he was starving or whatever and needed the money, I'd be okay with it.	Y
M3	If you need it, you need it.	Y
M4	Working on the side for pocket change? Yeah, I can see that.	Y
F4	I think Paul should ask his friends for help rather than try deceive them like this.	Y
M5	If you're like really starved for cash, I can get why Paul would do the whole thing.	Y

Appendix 18: Appealing to Good Character 1: "I can't believe you would doubt me like that. We're friends, aren't we?"

Tag	Line	Plausibility
F1	Is the argument that since she's my friend, I should trust his judgement?	Y
F2	If I was that friend, I'd ask him again.	N
F3	Of course I'm doubting you; you just said you did horrible things.	N
M1	Just own it up.	N
M2	Is getting angry an admission of wrongdoing?	N
M3	Not much of a friend, is he?	N
M4	I wouldn't be satisfied with that answer.	Y
F4	If you can't explain yourself to your best friend, you got way bigger problems.	N
M5	I can see somebody saying that, but it isn't much of an excuse, is it? He's just avoiding.	Y

Appendix 19: Appealing to Good Character 2: "I'm entitled to profits from my social media account. I put a lot of work into it."

Tag	Line	Plausibility
F1	You're really not though.	N
F2	No. Paul didn't start doing social media to one day profit from it and he's been doing it so far without payment.	N
F3	Not at the expense of others.	N
M1	But, no, you can't just do whatever you like with things you	N

	own-	
M2	You can't just do anything you want with the things you own.	N
M3	I don't think you do own your social media accounts in the first place. That's why they can ban you for whatever they want, essentially.	N
M4	Interesting. If it is my property, why couldn't I profit from it?	Y
F4	There are legitimate ways to do what Paul did, so I can see why he'd think that.	Y
M5	Maybe he actually thought he was doing legitimate business - the sort of social media hype marketing. You think there's some fault in social media platforms not making it clear this isn't allowed? Personally, I thought it was clear from whenever you saw those 'paid advertisement' labels that advertisement without those labels wasn't allowed. But I can see that not being so clear to some.	Y

Appendix 20: Victimization 1: "The rival company deserved all of it. Didn't you read what they had done?"

Tag	Line	Plausibility
F1	That doesn't sound like what Paul would say.	N
F2	That sounds a bit too... lofty for Paul.	N
F3	That's a terrible thing to say.	N
M1	I don't think Paul would think so strongly about what's he's doing.	N
M2	I don't believe anybody is going to join a campaign like that with the intent of moralizing the people they are campaigning to.	N
M3	That's not really something a... random social media advertiser would think.	N
M4	Like a form of Stockholm Syndrome? Where they actually start believing the stuff they are deceiving about?	N
F4	I don't think that's what Paul is thinking about.	N
M5	Maybe if he read about that before, but this is just post-hoc rationalisation.	?

Appendix 21: Victimization 2: "This sort of thing is normal between companies. I bet the rival company is doing the same thing."

Tag	Line	Plausibility
F1	I don't think that question fits the discussion.	N
F2	I get that, but why's Paul doing this?	N
F3	The answer isn't to do it yourself, but for both to stop.	N
M1	Yeah, I bet they are.	N
M2	I think a better use for the company's money is to just lobby laws to prevent this type of advertising in the first place.	N
M3	Maybe, but suspicion like that doesn't warrant these actions.	Y
M4	Just because it's being done, doesn't mean you should do it yourself.	N
F4	I don't think Paul has anything to do with... grudges between companies.	N
M5	That may be a fine reason for why Phion (Samsung's name in the vignette) is doing this, but it's not a reason for why Paul would do or join it.	N