A Comparison of Hostile Language in Online Debates for Anonymous and Identified Speakers

Calum Prescott

**Introduction**

In this investigation, I will study the effects of anonymity on online debates concerning controversial topics. As a member of a generation which has had unrivalled access to social media and the internet, I regularly debate with strangers online, and I find that disagreements can easily lead to hostile language when channelled anonymously (for example, on the comment section of a news site). Studies concerning the language of the internet are, to me, some of the most interesting and relevant areas of current linguistic research.

I will be testing the hypothesis that anonymity in an online channel of discussion will lead to more aggressive language (including expletives, non-hedged imperatives and other face-threatening acts) and less accommodation (adjustment of language to suit others). I believe that removing the identity of speakers and geographically separating them enables them to avoid using politeness strategies and resist conforming to the majority viewpoint. This theory is tied into the psychological topic of conformity which I have studied; Asch (1951) claimed that people comply with majority viewpoints to be socially accepted. Online anonymity may reduce the effects of normative social influence.

Tim Shortis (Shortis, T., 2000, “*The Language of ICT*”) suggests that one of electronic text’s distinctive features is that it “challenges notions of authority”. He argues that by allowing strangers to converse without strict regulation, the internet gives us the power to say anything we want. Shortis refers to the idea of “disinhibition” – in other words, we lose restraint and disregard the social conventions of language. This theory may be useful in explaining my hypothesis, whilst others such as Howard Giles’ accommodation theory (Giles, H.; Coupland, N., 1991, “*Language: Contexts and Consequences*”), and Brown and Levinson’s politeness theory (Brown, P.; Levinson, S., 1987, “*Politeness: Some Universals In Language Usage*”) can be challenged by the introduction of anonymous social media. By studying linguistic devices such as convergence/divergence, positive/negative politeness strategies and expletives, I aim to draw comparisons between participants whose identities are exposed and those whose identities are hidden.

To test my hypothesis, I shall incorporate various methods. I will conduct an experiment using two or more separate websites as sources, comparing and categorising the user comments on each site. I will also survey a group of regular internet users to gain a greater insight into their online discussions, providing both quantitative and qualitative data that further explores the key differences regarding hostile language between anonymous and identifiable speakers.

**Methodology**

My hypothesis is that anonymous individuals generate more hostile language in an online debate than identifiable individuals, so my primary research method will be an experiment comparing comments of users on anonymous and exposed social channels. I will choose two comment threads from different websites, discussing the same issue. Website A’s users will have their identities shown and Website B’s users will have their identities hidden. On each website I will analyse 120 consecutive comments, counting disagreements with previous comments, insults toward other users, expletives (swear words) and imperatives (users commanding other users). I will tally a feature whenever it appears in a comment. However, I will not count multiple appearances of a linguistic feature in one comment, which will prevent anomalous comments from skewing the data.

With most extraneous variables controlled, I think my experimental methodology is strong, but it has some weaknesses. A large amount of data is easy to find, and I have managed to operationalise the dependent variable (hostility of language) by turning qualitative data (the comments) into quantitative data (the tallies). I hope to offset the effect of personality differences between Website A’s users and Website B’s users by sampling a large number of comments. Furthermore, both comment threads will be discussing the same issue. Different issues have different levels of controversy and importance, which affects the amount of hostile language used. It is also important that each piece of content has a similar target audience to further reduce individual differences. Unfortunately, if my hypothesis is correct, it may be difficult to generalise the findings to all demographics. The samples I analyse are likely to stem from a narrow age range and single culture, and will ignore certain types of people (such as those who don’t actively use social media). Also, while I have tried to limit the effects of individual differences, I cannot assume that the only factor for a difference in hostility is anonymity.

My secondary research method will be a survey given to a sample from my school year, generating both quantitative and qualitative data relating to the online language used by participants. The survey will consist of open and closed questions inviting detailed responses, supporting the results of the experiment by giving a greater insight into how and why language changes under anonymity (especially during conflict). As my sample is being taken from my school year, the age range will be extremely limited. However, my age group is one that commonly uses and debates on the internet, so the sample is appropriate. One potential issue that will be hard to prevent is the effect of ‘demand characteristics’ – participants may answer the questions differently because they think I want them to give answers that support my hypothesis. To limit this, I will emphasise the necessity of being truthful when giving participants the survey.

**Analysis**

After collecting my website data, I found strong evidence for my hypothesis. On the YouTube site, where comments were anonymous, there was a significantly higher level of hostility in user language. This is shown by the increase in expletives, insults and imperatives. Some of my survey data also supported the hypothesis, but needs to be examined in closer detail before conclusions are made.

Asch’s theory of normative social influence (1951) states that we conform to the majority viewpoint to be socially accepted. Similarly, Giles’ social accommodation theory (1991) outlines the way we converge our speech with others for the sake of improving social relations. These theories show how we adapt both our language and viewpoints to make other people like us, and they may explain why anonymity increases online hostility. When we are anonymous, our social status is much harder for others to recognise. Likewise, we lack knowledge of the social status of others, therefore we cannot increase social standing by agreeing with others or using politeness strategies to gain their favour. It is still possible to converge with their style of language, but we would receive no social benefit from doing so. I believe that these theories explain why anonymity increases online hostility, and below I analyse the evidence for this claim.

**Conformity and Conflict**

In my experiment, there were 23% more disagreements in the anonymous condition than in the exposed condition. This suggests that web users are more willing to resist pressure to conform with the viewpoints of others when their identities are hidden and social acceptance is irrelevant, supporting Asch (1951). On the Facebook thread, I often saw comments supporting the opinions of others (“you nailed it Red”, “well said Cliff”). Positive feedback was also present on YouTube, expressed with both comments and ‘likes’, but disagreements were more frequent. In addition, opposing viewpoints on Facebook were often hedged (“yeah, but they could have access to laboratories”) whilst on YouTube, disagreements were expressed bluntly or even with disrespect (“hahahaha you fucking believe that virgins are waiting for you up in heaven”). One of my survey respondents wrote that they are “a bit more frank when anonymous”, supporting my claim that honest thoughts are easier to express without an identity.

However, my survey data also suggested that many people do not bother creating or joining online conflicts, with a third of respondents claiming they never get involved, and the remaining two thirds “rarely” getting involved. Some participants felt that there was no point trying to change the opinions of people they’ve never met, or that influencing others was too difficult (“people choose to hear/read what they want so why bother?”). Other participants believed that they were as civil when anonymous as they were in any other situation (“my language doesn’t deviate that much”), and one person claimed that they “tend to be more polite” when anonymous, directly opposing my hypothesis. These statements are unsurprising, as the YouTube sample still included many examples of positive politeness strategies (“peace be with you, friend”) and positive feedback (“I so agree with you”, “well stated”). Personality appears to be a key variable in this matter, so it may be simplistic to generalise my experimental findings to everyone.

An extraneous variable in the experiment that may have affected the proportion of conflict between the two threads is the ‘dislike’ button, a graphological feature which is only available for YouTube comments. It is possible that the presence of this button encourages conflict between users, as the “thumbs down” symbol is semantically associated with disagreement.

**Insults and Expletives**

Proportionally, the differences in expletives and insults between the Facebook and YouTube threads were much greater than the difference in disagreements. Expletives were 150% more frequent in the YouTube thread, and insults were 91% more frequent. These huge increases indicate that aggressive language is even harder to express with an exposed identity than a simple disagreement. On Facebook, disagreements were usually expressed respectfully (“unfortunately not, Dominic”), often addressing the opponent by name. In contrast, YouTube disagreements were likely to involve combative, hurtful lexis (“go and enjoy your one life you fucking retard”).

It is normal to experience negative thoughts about someone whose opinion differs to yours, but according to Brown and Levinson (1987), we tend to withhold from threatening positive face for the sake of maintaining *social harmony*. The theory also states that politeness strategies are affected by contextual factors such as the social distance between speaker and listener. In an average dispute between two people, the language used to express a disagreement will depend on various factors. These include whether one person has legitimate authority over the other, how well the two people know each other, the importance of the subject matter and the risk of harming social relations. In the case of YouTube, there is little to no social relationship between users, no user has legitimate authority, and there is no risk to social standing, therefore it is safer and less damaging for commenters to express their negative thoughts through insults and expletives. My survey data supports this idea. Some respondents felt that they were “more rude” when anonymous. One respondent wrote that being hidden allowed them to “mask my vicious and aggressive traits”, fitting with the notion that social (and geographical) distance prevents us from withholding impoliteness. Overall, a third of respondents said they were generally aggressive during an online conflict, although this question did not distinguish between anonymous and non-anonymous conflict.

Other insults and expletives were aimed at Islamic State militants (who were the subject of discussion). On Facebook, wishes for their deaths were often expressed with euphemisms (“putting these guys into the ground”), whereas on YouTube, much harsher language was prevalent (“destroy these people”, “should be nuked”). Xenophobic, taboo language was also more frequently used on YouTube to insult the militants (“towel heads”, “sand monkeys”), showing that anonymity allows us to share extreme prejudices without fear of judgement. This attitude was regularly expressed by survey participants, one of whom wrote that they “care less about offending” when anonymous.

**Power and Hierarchy**

Hostile language is used in online debates to assert power over opponents. As well as removing the effect of social rewards, anonymity removes any kind of established social hierarchy. Social status normally affects the balance of power in a conversation – for example, in a school environment a more popular participant may be able to interrupt less popular participants, and less popular participants may publicly agree with their opinions despite privately disagreeing with them. The lack of social status on websites such as YouTube means that everyone, regardless of their place in society, has equal standing in a conflict.

To establish power over others online, anonymous speakers often use imperatives. Over twice as many imperatives were used in the YouTube thread compared with the Facebook thread, generally in an insulting manner (“kill yourself”, “dream on, child”). None of my survey respondents mentioned commanding others in conflicts, but references to power and social status were found (“people don’t know who you are so you can say what you want”). The use of imperatives can be linked to the theory of Brown and Levinson (1987). Whereas positive face refers to one’s self-esteem and self-image, negative face refers to one’s sense of freedom and autonomy. By using imperatives, we are commanding others, restricting their freedom. We are threatening their negative face. Just as anonymity allows us to threaten the positive face of others by altering social context, it also allows us to discard negative politeness strategies and threaten anyone’s negative face.

Another common way for anonymous users to demonstrate status is through superior spelling and grammar, along with criticism of opponents’ non-standard linguistic features. For instance, one YouTube user typed that IS militants should be told “GOD DOESN’T LIKE RELIGION”, to which an opponent replied “SHOULD WE TELL THEM IN ALL CAPS?!” This example of sarcastic stylistic convergence was immediately followed by divergent standard grammar, used by the participant to appear calmer and more intelligent than his target. Once the participant’s identity is removed, power can only be established through the quality or force of written language, explaining perhaps why so many found the use of Standard English a source of status. This idea is supported by one of my survey respondents, who said that “I tend to write in a more formal way to people I do not know in order to give a good impression of myself”. This sentence summarises the way that in the absence of an established hierarchy, we elevate our position using language that connotes knowledge, intelligence and sophistication. In fairness, this is still mostly true for online channels where identities are exposed, as the social status of individuals is still largely unknown. One could also argue that it is generally more important to use Standard English when you are not anonymous, because giving a good impression of yourself has greater social consequence.

**Conclusion**

From the experiment and survey data I have collected, I can conclude that hostile language is more prevalent in online debates when the participants are anonymous. Furthermore, conflicts themselves appear to be more common in anonymous conditions, with more supportive feedback between users whose identities are exposed. My experimental results showed highly significant differences in a range of linguistic devices, providing a quantitative measure of hostility and clear evidence to support my hypothesis. Through my questionnaire, I was able to obtain a mix of quantitative and qualitative data focusing on specific online behaviour of respondents and the reasons for their behaviour. However, responses to the survey varied greatly, and a large proportion of participants claimed that they never communicate anonymously or debate with others online. This suggests that I cannot generalise the findings of my experiment to all personality types, as there are many people who would not engage in online hostilities at all and therefore cannot be studied with my experimental method.

There were noticeable flaws in my methodology. Despite efforts to minimise the effect of individual differences in the experiment, it is still possible that extraneous factors such as YouTube’s ‘dislike’ button could have accounted for some of the differences in hostile language. The survey, meanwhile, suffered from an extremely limited sample that did not produce any consistent patterns. Regardless, my experiment showed that anonymity has marked effects on social conflict, and by linking these findings with the theories of Asch, Shortis, Giles and others, as well as the data from my respondents, I have identified the removal of social hierarchy and consequence as possible reasons for increased hostility in language.

**2,532 words**

**Bibliography**

Theorists

Theory of Normative Social Influence: Asch, 1951

Theory of Disinhibition: Shortis, T., 2000, “*The Language of ICT*”

Theory of Accommodation: Giles, H.; Coupland, N., 1991, “*Language: Contexts and Consequences*”

Theory of Politeness: Brown, P.; Levinson, S., 1987, “*Politeness: Some Universals In Language Usage*”

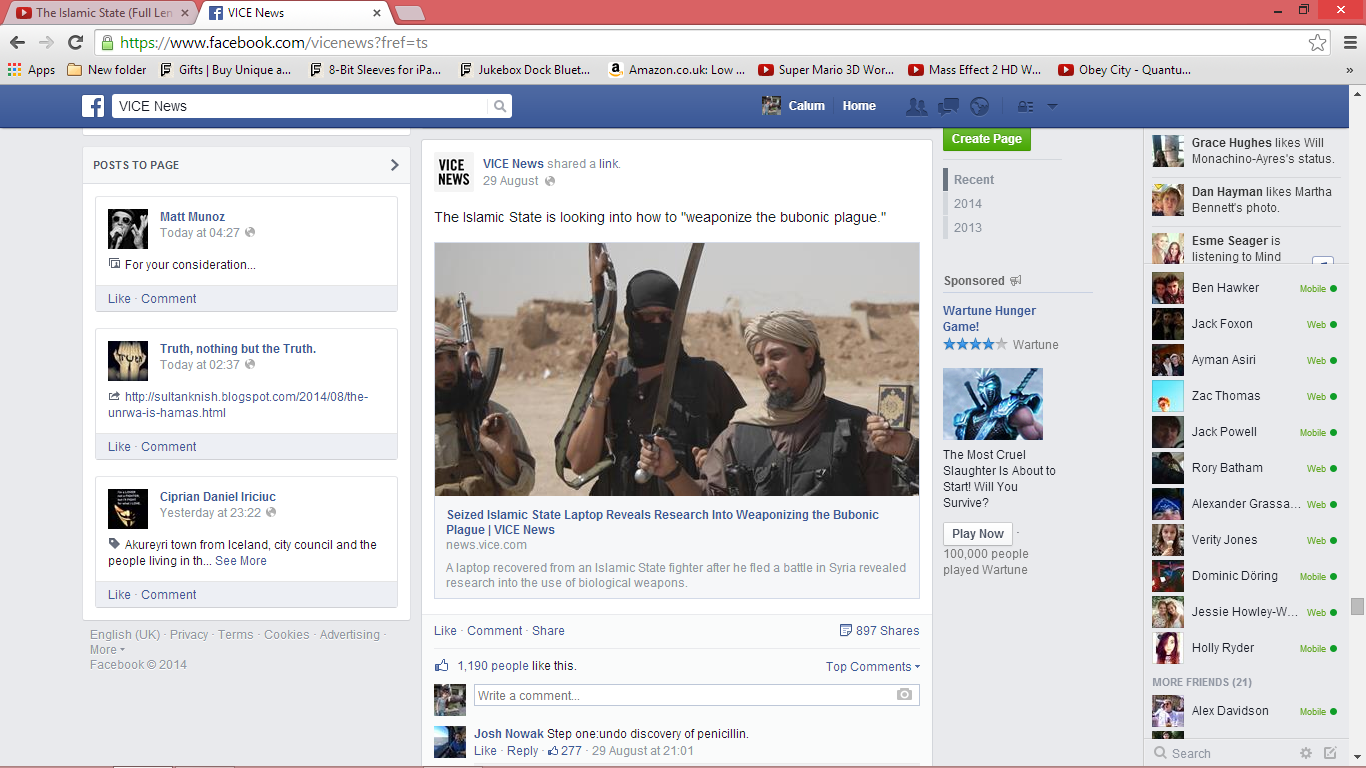
Data

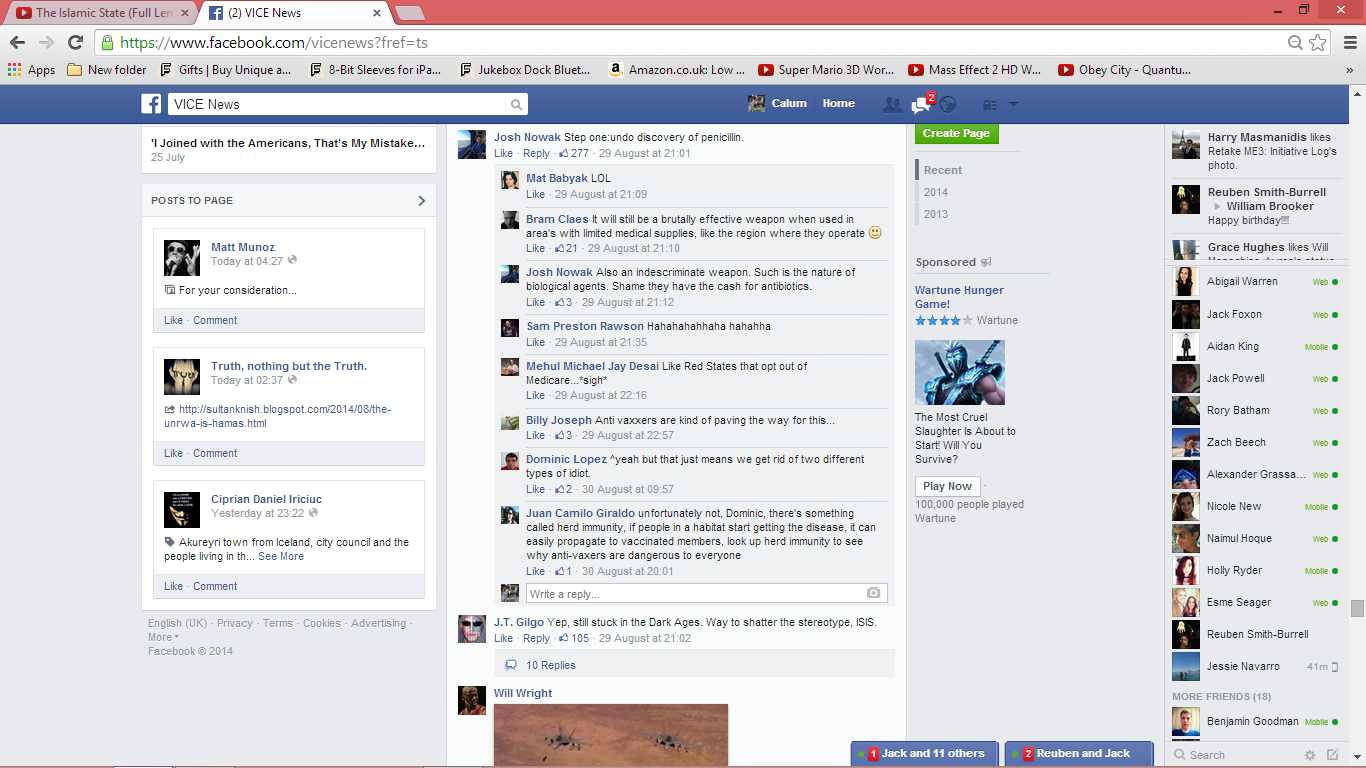
Facebook Page: <https://www.facebook.com/vicenews/posts/328432117316745?reply_comment_id=328456063981017&total_comments=1>

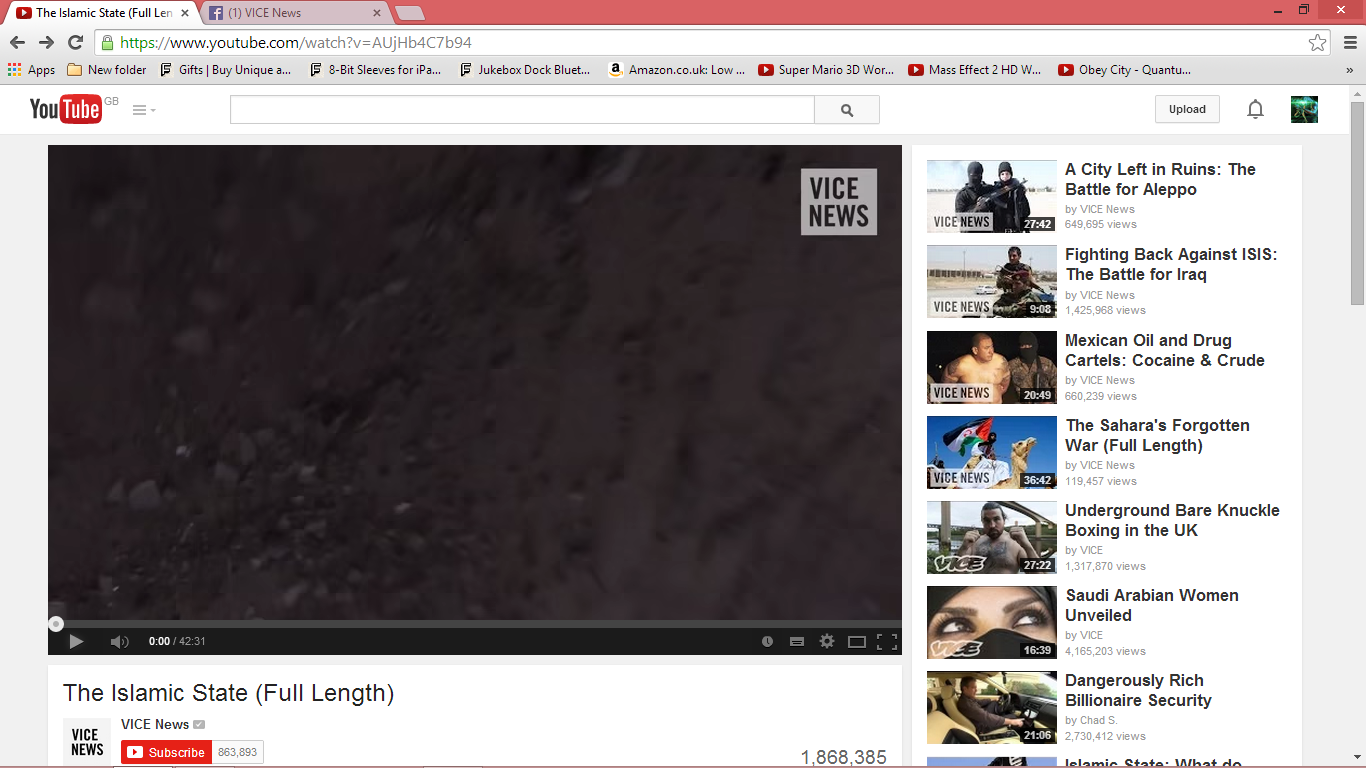
YouTube Page: <https://www.youtube.com/watch?v=AUjHb4C7b94>

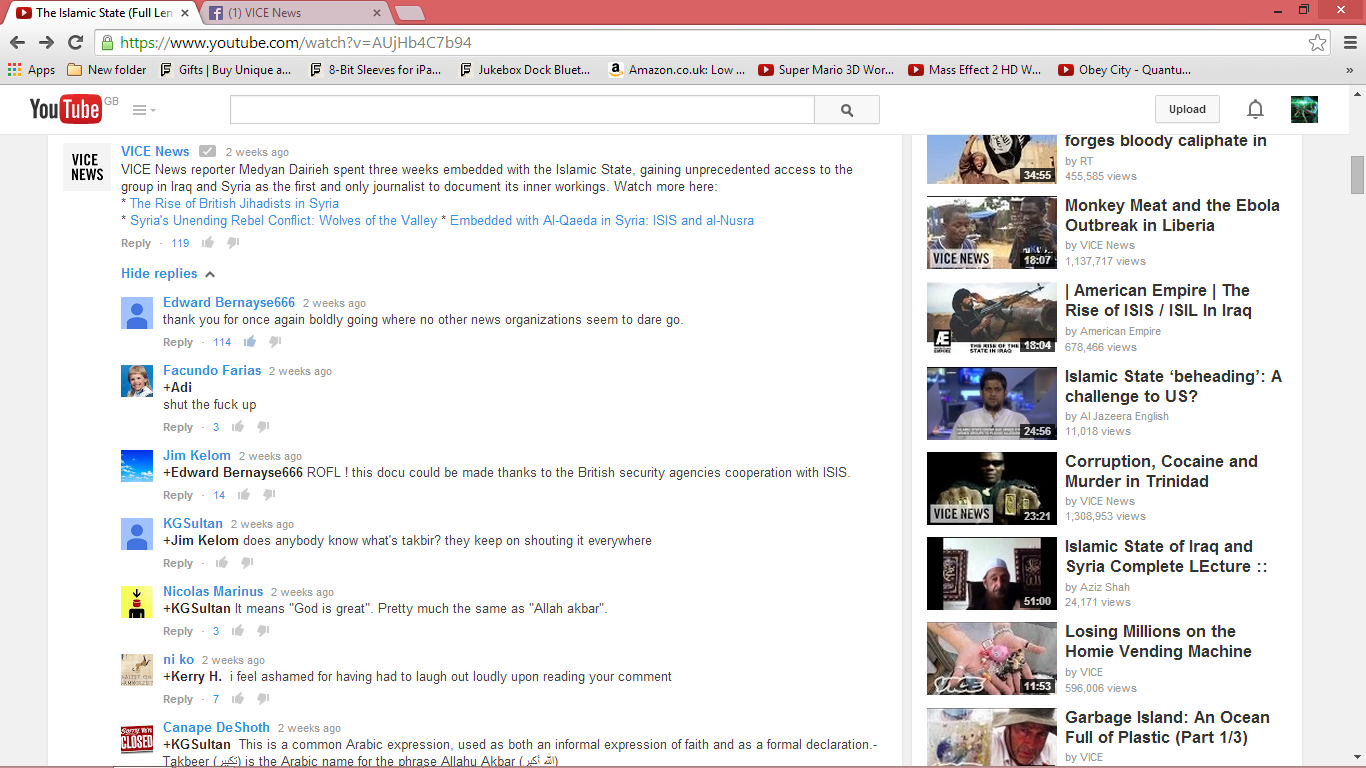
**Appendices**

Data Samples









Results

On both pages, I analysed the **first 120** comments on **comment-reply chains** that contained **at least one** **disagreement**. Singular comments were ignored, as were comment-reply chains that contained no debate. The results can be seen below.

Facebook Page:

Disagreements – 35 (29%)

Insults – 11 (9%)

Expletives – 8 (7%)

Imperatives – 6 (5%)

YouTube Page:

Disagreements – 43 (36%)

Insults – 21 (18%)

Expletives – 20 (17%)

Imperatives – 16 (13%)

Survey

Written below are eight questions related to your online activities. Please answer all questions as **truthfully** as you can. Your answers will remain anonymous.

1. How much time do you spend a week on social media on average?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

2. Do you ever communicate with unfamiliar people online?

(This may include private messaging, social media comments, public forums…)

Often/Rarely/Never

3. If so, are you less comfortable talking to them if they know your name?

Yes/Sometimes/No

4. Are you generally polite when you talk to others online?

Always/Sometimes/Not Much

5. Does your language change online when you are anonymous? If so, how?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

6. Do you ever get involved in conflicts online?

Often/Rarely/Never

7. If so, are you generally calm/polite during a conflict, or can you be aggressive?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

8. Do you prefer to have your identity hidden or exposed on the internet, and why?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Thank you for taking the time to complete this survey.