International Journal Of Global Community
Volume VII No. 1 (March) 2024

The Presence of Virtue Signalling on Social Media Agreement

Christopher Tong and Ryan C. Anderson

School of Psychological Sciences, Faculty of Medicine, Nursing, & Health Sciences, Monash University, Wellington Road, Clayton, Victoria, Australia

Corresponding author: ryan.anderson1@monash.edu

Abstract
Social media has facilitated the rapid spread of virtue signalling (the display of moral values, worth, and significance to others). The current study aimed to assess if the presence of virtue signalling affects agreement for statements concerning topics such as charities, social justice, and politics. A total of 207 participants (62.3% female) aged between 18 and 80 (M = 41.19 years, SD = 18.57 years) were recruited from Facebook. An online Qualtrics survey was used to collect data on support for the topics of charities, social justice, and politics. Participants were randomly presented with simple vignettes including either virtue signalling or not, for each of the three topics listed above. It was found that the presence of virtue signalling buffered against decreasing attitudinal agreement compared to control slogans for the topics of politics $F(1, 205) = 20.945$, $p < .001$, $\eta_p^2 = .093$ and social justice $F(1, 205) = 7.00$, $p < .001$, $\eta_p^2 = .063$ but not charities (1, 205) = 3.377, $p = 0.19$, $\eta_p^2 = .047$. It was concluded that virtue signalling can help to improve promotion and support for a cause but was not as effective for charities.

Keywords: Virtue Signalling, Conspicuous Consumption, Charities, Social Justice, SocialMedia, Politics

The impact of virtue signalling on social media agreement

Social media platforms have become increasingly accessible to the world's population, and it has become easier for individuals to learn of immoral behaviours online (Crockett, 2017). Virtue signalling is an affirmation of a moral value to display one's moral respectability (Brown et al., 2020; Levy, 2020; Saltman, 2017). Signalling virtue may derive from the evolution of group morality and sexual selection to attract mates (Brown et al., 2020; Levy, 2020; Miller, 2007). Virtue signalling can help individuals attain status in a hierarchy and leadership, ensuring they gain access to resources and mates according to the Moral Virtue Theory of Status Attainment (Bai et al., 2020). Virtue signalling online can escalate to moral outrage, becoming a form of conspicuous display to garner recognition and affirm in-group values and norms (Balon, 2020; Crockett, 2017). As a result, virtue signalling has serious implications in persuading others to a moral viewpoint, influencing politicking, elections, advertising, marketing and social justice as well as driving dehumanisation, excessive polarisation, and vilification of the 'other' (Farrell et al., 2020; Veissière, 2018). Despite this, very little is known about the effectiveness of online virtue signalling at increasing attitudinal agreement.

Virtue signalling and Morality

Moral character is connected to an actor's cognitions and intentions, and virtues display character traits and behaviours that promote moral excellence (Aristotle, 2011; Inbar et al., 2012). Morality is defined as the interconnection of virtues, values and norms that regulate or suppress self-interest to make group cooperation possible (Haidt, 2012). Signalling virtue can show one's attention to moral discourse and willpower (Inbar et al., 2012; Levy, 2020; Righetti & Finkenauer, 2011). Examples of virtues include humility, loyalty, and altruism (Haidt & Joseph, 2004; Tangney, 2000; Willer, 2009).

Levy (2020) suggested that virtue signalling conveyed confidence and judgement by consensus to assess action in a moral dilemma (Price & Stone, 2004; Pulford et al., 2018). Internal conflict between virtue and vice is a continuous struggle for desiring higher-order values over immediate gratification (Berman & Small, 2018). Since individuals are sensitive to cues that signal intentions, virtue signalling allowed human ancestors to demonstrate moral character and willpower which facilitated survival and sexual reproduction (Piazza et al., 2014).

Sexual selection and virtue signalling

Signals, according to Smith and Harper (2003), are ways to attract members of the same species with desirable traits (Grabo et al., 2017). Signals convey genetic quality about a male's fitness in some species by being costly and hard to fake. One such example is the large plumage of a peacock used to attract peahens (Smith & Harper, 2003; Zahavi, 1977). Virtue signalling and moral outrage is beneficial for long-term sexual selection since it displays moral character in affirming pro-sociality and fairness (Barclay, 2010). Brown et al. (2020) assessed the attractiveness of individuals who signal moral outrage compared to neutral faces

and found that outrage was attractive in long-term mates and rivals. Trivers' (1971) Parental Investment Theory suggests that because women have larger minimal reproduction costs, they are attracted to indications of moral outrage in a mate since their presence offers useful mate-relevant information. Such cues signal fairness and moral character which punishes cheaters who strain limited resources and threaten survival (Pedersen et al., 2013). However, these are unconscious processes and are only a single motivator for signalling. Therefore, sexual selection of virtue signalling can only account for why some individuals display their moral values and outrage as 'virtue signalling producers' from an evolutionary perspective.

Online Virtue Signalling and Deindividualization

In the cybersphere, virtue signalling has become an expression of moral outrage about the injustices of privilege over the disenfranchised; and calls attention to the violation of moral norms (Crockett, 2017; Saltman, 2017). As a type of virtue signalling, online outrage has a lower threshold because it can be expressed from home, and platforms may encourage habitual outrage expression (Crockett, 2017; Tosi & Warmke, 2016). Virtue signalling can produce a discrepancy between what people say and do, using social media to display their moral character to others (McClay, 2018). Support for various causes can be signalled online, for example by choosing profile pictures that display virtuous acts or showcasing instances of volunteering (Wallace et al., 2018). Also, it is much easier for individuals to remain anonymous, contributing to deindividualization within a 'Twitter Mob' (Duncan, 2020).

Excess signalling as well as perpetual exposure to moral outrage has lowered individuals' outrage threshold, resulting in an inability to differentiate between mere disagreement from heinous immoral acts (Crockett, 2017). As a result, online moral outrage is further perpetuated by news and blog articles that capitalise on and compete for social media user's attention to produce advertisement revenue (Duncan, 2020). The digital landscape is like a naturally selecting environment where exposure to '*clickbait'* and super-normal immoral behaviour increases calls for punishment and incentivises more outrageous articles.

Social media may have cheapened conversation, since tone of voice and facial expressions are not easily perceptible online, precipitating ease of miscommunication, misunderstanding or even a deliberate faking of online virtue signalling (Levy, 2020). As a result, emotions easily intensify to moral outrage. Desiring to punish others can further dehumanise and create *'echo chambers'* which is the limitation of communication and emotional expression to a sympathetic audience (Brady et al., 2017; Fincher & Tetlock, 2016). Therefore, virtue signalling in moral outrage can increase an individual's adherence to in-group norms and ridiculing the out-group results in disagreement, ideological segregation, polarisation, and dehumanisation of the other (Crocket, 2017). The current research evaluates whether virtue signalling can modify attitudinal agreement with statements concerning given issues.

Types of Online Virtue Signalling

Tosi and Warmke (2016) found that there were different types of online virtue signalling including *piling-on* which is repeating condemnations, *ramping-up* are calls for harsher punishments, *trumping-up* are individual claims of moral issues where others may see none, and *excessive outrage* are displays of out-of-proportion, outrageous reaction. Also, there is *self-evidence* where individuals claim to be a victim, and there is *moral perceptiveness* where those on the other side are morally deficient. De Cruz (2018) found that displaying online virtue and moral outrage was effective at signalling to in-group members.

Conspicuous Virtue Signalling

Charities

Though self-reported donations predicted actual donation behaviour according to Basil and colleagues (2006), conspicuous online virtue signalling can be utilised to gain desirable social standing amongst peers, display a need for uniqueness and enhance self-esteem on Facebook without the need to behave well or donate in the real world (Kastanakis & Balabanis, 2012; Schau & Gilly, 2003; Strizhakova et al., 2008; Wallace et al. 2018). Grace and Griffin (2006) expanded the Theory of Conspicuous Consumption to include displayed acts of charitable donation; to signal good impressions to others and provide donors with personal satisfaction (Veblen, 1899; West, 2004). The need to show one's goodness may have little resemblance to physical reality since there could be a dissociation between idealised and real identity (Farrell et al., 2020; Schau & Gilly, 2003; Wallace et al., 2018). Schau and Gilly (2003) found evidence of dissociation and predicted that some individuals signal their virtue without actual intention to donate or support a cause. Also, other's sharing or liking an individual's post enhanced one's self-esteem, which was more enhancing than having many friends (Greitemeyer et al., 2014). The need for self-uniqueness is socially acceptable to others and helps individuals find an in-group (Kastanakis & Balabanis, 2012; Tian et al., 2001).

Social Justice

Online virtue signalling of social justice topics have been reduced to mere Twitter hashtags, known as *'Hashtag Activism'*, which are typically organised around marginalised voices, but could possibly hinder meaningful dialogue due the creation of '*Echo Chambers*' (Farrell et al., 2020; Synovitz, 2018). Echo chambers are the online cliques where members receive information from like-minded individuals and organisations and may have resulted from the personalisation of ideas and arguments bound to one's identity (Crockett, 2017; Veissière, 2018). This may predict why online virtue signalling in social justice can lead to discussions that talk past each other, representing their own identity groups that clash with other groups. Showing adherence to the in-group and judging out-group ideas as 'pollution' led to increased division of the 'other' (Crockett, 2017; Douglas, 2003;

Haidt, 2012). Social justice is an important topic that has serious implications in increasing support or division for social activism.

Politicking

Additionally, online politicking is one such area where social media can easily create division and dehumanisation of the ideologically opposing side. Predominantly left-wing candidates who discuss minority issues such as inequality and racism are often accused of 'virtue signalling' and resultingly receive online abuse for it (Farrell et al., 2020). Politics is often tribalistic and hostile since counter-intuitive facts that violate group social norms and moral values can produce autonomic responses such as aggression (Haidt, 2012). Since social media is open and immediate, virtue signalling could produce instantaneous online trolling and moral outrage which further polarises the middle majority to take sides (Craker & March, 2016; Mackay, 2017). Also, accusations of virtue signalling by opposing sides are associated with insincerity because it could be easily faked and tokenistic without the need to place substantial individual cost, effort, and support (Farrell et al., 2020; Wallace et al., 2018).

Online effectiveness of virtue signalling

Virtue signalling is akin to a game where individuals compete to show the most charisma, voice injustices, and determine who is most fit to change the rules. Duncan (2020) stated that social media can become corrupted 'play' that divisively leads to the polarisation and dehumanisation of the 'other' (Crockett, 2017). Due to increased engagement resulting from outrage and virtue signalling by social media users, journalists perpetuate outrageous 'clickbait content', to increase engagement and revenue (Grzywinska & Batorski, 2016; Messner & Distaso, 2008; Wallsten, 2007). It has become a business model for media and advertisers to attract consumers since it is the most rapidly shared type of content (Coles & West, 2016; Craker & March, 2016; Fan et al., 2013). However, little is known about the effectiveness of online virtue signalling.

The Current Study

The current study will look at the effectiveness of virtue signalling in increasing attitudinal agreement levels for users of social media. Virtue signalling may be driven by biological pressures for group survival and sexual selection, to win mates or choose effective, charismatic leaders who signal moral virtue for status (Bai et al., 2020; Brown et al., 2020; Massey-Abernathy & Haseltine, 2018; Miller, 2007; Levy, 2020). Social media-created conspicuous virtue signalling and moral outrage to highlight issues like support for charities, social justice and politicking to drive moral action, but has also produced negative ramifications in polarising, and dehumanising the 'other' (Balon, 2020; Crockett, 2017; Duncan, 2020; Hamilton, 2019; Saltman, 2017; Uhlmann et al., 2013; Wallace, et al., 2018). It was hypothesized that attitudinal agreement with statements about the topics of charities, social justice, and politics will be greater when virtue signalling is present

than when it is not. Posts were artificially generated as to respect privacy and copyright.

The following figures below are examples of (1) slogan control and (2) virtue signalling for the topic of social justice:

Figure 1 Social Justice Slogan Control

Figure 2 Social Justice Virtue Signalling

Methods

Participants

A convenience sample of 207 participants (M = 41.19 years, SD = 18.57 years) was drawn mostly from Facebook advertising in Australia and any literate adult with access to social media could choose to participate. The majority of the sample was female (62.3%), and indicated that they were Caucasian/White (67.1%).

Social media advertising commenced on the 11th of January 2021 and ended on the 21st of February. Upon completion of the survey participants were given the opportunity to enter a draw to win a $50 gift card.

Materials, Procedure, and Design

The current study consisted of an online questionnaire hosted on the survey creation platform Qualtrics. After choosing to take part in the survey and responding to a series of standard demographic questions (age, gender, ethnicity etc.), participants were presented with a sequence of three individual hypothetical social media posts concerning the topics of charities, social justice, and politics. For each of the three topics participants were randomly shown a post containing either

virtue signalling or no virtue signalling. The virtue signalling post showed examples of individual leaders' efforts in advocating for a cause and the non-virtue signalling controls included reposting generic slogans like '*Donate to…,*' '*Vote for…*' and '*Black Lives Matter*.' Posts were created on Apple Pages to illustrate fictitious Facebook and Twitter posts and used images from Creative Commons.

Prior to viewing each relevant post participants were asked to indicate their agreement to four questions about their *general* attitude toward social media posts concerning moral behaviour/action. Although wording changed slightly, so as to be applicable to the given topic, the initial 4 questions for the charity condition (for example) were:

1. Do you agree that social media accounts should be used to promote charities?
2. Does seeing social media posts about charities create emotional intensity for you?
3. By seeing social media posts about charities, are you compelled to moral action?
4. Do supporters of the charities' social media posts appear to be genuine

After viewing each post participants were asked the same four questions, but this time in regard to the *specific* post they had just seen (e.g. "By seeing the previous social media post about charities, are you compelled to moral action?"). All questions were responded to on a 5-point Likert scale (1 = *strongly disagree* to 5 = *strongly agree).* Hence, each participant was exposed to 24 questions in total across 3 different topics.

Separate 2 x 2 MANOVAs were conducted for each of the 3 topic areas evaluating the extent to which a participant agreed with each of the 4 questions changed depending on exposure (*before* exposure to the stimulus/ *after* exposure to the stimulus), and the condition that a participant was randomly assigned to (virtue signalling condition/control).

On average this survey took 6 minutes to complete. After each participant answered 24 items, they were shown a debriefing statement summarising the purpose of this study and were presented with the opportunity to enter the prize draw for a $50 voucher.

Results

Descriptive statistics

Tables 1-3 below represent participants' average attitudinal agreement scores for 3 different topics. It is worth noting here that although each given participant saw a hypothetical post relevant to each of the 3 topics, the posts that they saw contained either virtue signalling (VS) or no virtue signalling. The lower scores on the 5-point Likert scale indicate lower agreement whereas higher Likert scores indicate higher agreement.

Table 1
Mean (SD) Agreement for Charity Questions

	Pre	Post
Control	Question 1 - 3.80 (.98)	Question 1 - 3.51 (1.08)
	Question 2 - 3.06 (.98)	Question 2 - 2.58 (1.06)
	Question 3 - 2.97 (1.04)	Question 3 - 2.40 (1.00)
	Question 4 - 3.16 (.91)	Question 4 - 2.91 (1.07)
VS	Question 1 - 3.85 (.95)	Question 1 - 3.77 (.97)
	Question 2 - 3.05 (1.10)	Question 2 - 2.95 (1.21)
	Question 3 - 2.84 (1.11)	Question 3 - 2.78 (1.17)
	Question 4 - 3.29 (.86)	Question 4 - 3.32 (.95)

Table 2
Mean (SD) Agreement Social Justice Questions

	Pre	Post
Control	Question 1 - 3.71 (1.04)	Question 1 - 3.62 (1.17)
	Question 2 - 3.51 (1.14)	Question 2 - 3.10 (1.16)
	Question 3 - 3.09 (1.61)	Question 3 - 2.77 (1.20)
	Question 4 - 3.18 (1.21)	Question 4 - 2.93 (1.15)
VS	Question 1 - 3.91 (1.07)	Question 1 - 3.78 (1.23)
	Question 2 - 3.59 (1.01)	Question 2 - 3.20 (1.17)
	Question 3 - 3.36 (1.09)	Question 3 - 3.03 (1.19)
	Question 4 - 3.41 (1.10)	Question 4 - 3.41 (1.17)

Table 3
Mean (SD) Agreement Scores for Politics Questions

	Pre	Post
Control	Question 1 - 2.99 (1.68)	Question 1 - 2.93 (1.57)
	Question 2 - 2.89 (1.09)	Question 2 - 2.34 (1.03)
	Question 3 - 2.52 (1.04)	Question 3 - 2.19 (0.93)
	Question 4 - 2.82 (1.03)	Question 4 - 2.83 (1.10)
VS	Question 1 - 2.86 (1.53)	Question 1 - 2.97 (1.14)
	Question 2 - 2.81 (1.18)	Question 2 - 2.48 (1.09)
	Question 3 - 2.62 (1.07)	Question 3 - 2.28 (1.01)
	Question 4 - 2.75 (1.10)	Question 4 - 2.83 (1.12)

Inferential statistics

Charity

For the topic of charity there was an overall multivariate effect of exposure, Pillai's Trace = .120, $F(1, 205) = 28.586$, $p < .001$, $\eta_p^2 = .120$, with agreement generally decreasing following exposure. Exposure interacted with condition here,

Pillai's Trace = .077, $F(1, 205) = 3.377$, p =0.19, η_p^2 = .047, such that overall agreement decreased more in the control group than in the virtue signalling group.

Social justice

For the topic of social justice there was an overall effect of exposure, Pillai's Trace = .120, $F(1, 205) = 28.075$, $p < .001$, η_p^2 = .120, with agreement once more decreasing following exposure. Exposure interacted with condition here, Pillai's Trace = .093, $F(1, 205) = 7.00$, $p <.001$, η_p^2 = .063, such that overall agreement decreased more in the control group than in the virtue signalling group.

Politics

For the topic of politics there was an overall effect of exposure, Pillai's Trace = .070, $F(1, 205) = 15.313$, $p < .001$, η_p^2 = .070, with agreement decreasing following exposure. Exposure interacted with condition here, Pillai's Trace = .093, $F(1, 205) = 20.945$, $p < .001$, η_p^2 = .093, such that agreement decreased more in the control group than in the virtue signalling group.

Discussion

The current study explored the effect of virtue signalling on modifying attitudinal agreement within the topics of charities, social justice, and politics. The hypothesis that the presence of virtue signalling would increase attitudinal agreement for various topics was partially supported. Analysis found that the presence of virtue signalling was significant in politics and social justice in buffering attitudinal decline compared to control slogans but not for charities. It was found that overall, the control slogans of each of the topics led to decreases in attitudinal agreement compared to virtue signalling.

Charities and virtue signalling

In both virtue signalling and slogan control groups, there were no significant differences in the topic of charities likely due to the face that pro-social goals were present which fostered group cooperation and reciprocal altruism that directly benefitted the community, thus, virtue signalling was found by this study to have no buffering effect against attitudinal decline (Hamilton, 1964; Henrich & Boyd, 2001; Trivers, 1971). It could be that virtue signalling in charities are an example of conspicuous consumption since it enhanced personal satisfaction and self-esteem amongst peers by displaying charitable donations but the effectiveness of agreeability on the peers themselves was had a different effect as demonstrated in this study (Grace & Griffin, 2006; Veblen, 2005; Wallace et al., 2018; West, 2004). According to Kastanakis and Balabanis (2012), the need to display individuality and become socially accepted motivated individuals to find like-minded virtue consumers and engage in morally beneficial behaviour like donations to a charity, especially on social media like Facebook (Bénabou & Tirole, 2006; Grabo et al., 2017). Generally, humans want to represent themselves as being agreeable and

morally virtuous however the effects on the viewer may have a different effect (Grabo & van Vught, 2016; Walumbwa et al., 2008).

The Significance of Virtue signalling?

It is possible that 'virtue-signallers' may increase moral standing, promote moral excellence, and increase group cooperation which can justify the motivation for virtue signalling (Aristotle, 2011; Inbar et al., 2012; Haidt, 2012; Kotabe & Hofmann, 2015; Levy, 2020; Righetti & Finkenauer, 2011). Considering sexual selection and evolutionary signalling theory, virtue signals in the current findings may have increased political and social justice support because it was salient for humans to attend to signals that convey moral character and leadership status (Bai, 2017; Grabo et al., 2017; Grabo & van Vugt, 2016; Piazza et al., 2014; Smith & Harper, 2003). According to Bai's Moral Virtue Theory, virtue provided a third pathway for individuals to gain status and leadership other than dominance and competence, leading to virtue admiration which positively correlated with warmth, moral identity, and honesty (Bai, 2017; Bai et al., 2020; Grabo & van Vugt, 2016). Due to a possible buffering effect of virtue signalling on politics and social justice, the study's findings indicated that virtue signalling was an example of reciprocal altruism perhaps for one's own agreed support because these individuals were given status in their ingroups that may have aided sexual selection and leadership pathways (Trivers, 1971).

Virtue signaling without moral outrage and the use of deception

Previous research has found that virtue signaling could lead to moral outrage (Crockett, 2017; Saltman, 2017; Tosi & Warmke, 2016). In this study, virtue signaling did not lead to excessive disagreement, but rather buffered against the effects of decreasing agreement, challenging the findings of Crocket (2017) and may suggest the existence of a milder form of virtue signalling that promotes altruism (Trivers, 1971). Previous findings indicated that the threshold for showing moral outrage was lower online and led to '*ramping up*' of aggression (Brady et al., 2017; Bushman, 2002; Fincher & Tetlock, 2016; Tosi & Warmke, 2016). It was possible that 'moral fatigue' resulting from individuals' prior experience to morally outrageous virtue signaling may have been present in this study however moral outrage was not measured (Bushman, 2002). This could explain why the topics of social justice and politics did not lead to significant decreases in agreement levels because moral fatigue may have increase the threshold for moral outrage but future studies would need to factor in moral outrage specifically.

Typology of Virtue Signaling

Another reason why social media virtue signaling of social justice and politics was significant in buffering the effects of reducing agreement may be due to the use of non-divisive virtue signaling typologies. The current study utilised '*trumping-up'* which is a type of virtue signaling that bolsters individual claims on moral issues (Tosi & Warmke, 2016). Also, this form of virtue signaling is not as

inflammatory, or thought-provoking, thus the current study may have not generated sufficient moral outrage (Levy, 2020). Hence, the current study did not utilise morally outrageous virtue signaling, and merely explored mild virtue signalling when using 'trumping-up' as the controlling typology. Future research should evaluate the differing virtue signaling typologies to explore the effects of dehumanisation (Crockett, 2017).

Costliness

It was possibly much easier to state that one supports online virtue signaling than to volunteer one's time and effort to support causes. Resulting from sexual selection, group cooperation and pro-sociality signals required greater cost to the signaller to illustrate one's fitness or propensity to reveal their moral virtue to others (Barclay, 2010; Grabo et al., 2017; Smith & Harper, 2003). The reason for the lack of moral outrage generated in the presence of virtue signaling could be partly because less time, energy, and resources were needed for social media post support compared to a more costly, stronger signal (Jaeggi & Gurven, 2013). According to McClay (2018) there was also a discrepancy between what people said and what they did. A sacrifice in time and effort was costly, and a hallmark of moral, and charismatic leaders (Abele & Wojciszke, 2007; Goodwin et al., 2014; Grabo & van Vught, 2016; Walumbwa et al., 2008). The current study inquired if participants were driven to moral action, but this inquiry did not equate to performing the action. Therefore, low-cost signaling may contribute to smaller effect sizes and the lack of moral outrage.

An additional follow-up study should assess examples of costly signaling such as whether participants volunteered their time or donated money to a political party or social justice cause. Since the study did not compare low-cost with high-cost signaling, future research should assess how mere viewpoint agreement or 'soft-signaling' is different to actively pursuing causes, or 'hard-signaling' (Jaeggi & Gurven, 2013; Smith & Harper, 2003; Zahavi, 1977). Also, additional leadership qualities of hard signallers should be compared to reveal what makes them persuasive. Finally, future studies should explore whether virtue signaling predicts actual prosocial behaviour (Basil et al., 2006).

Limitations

The current study primarily used '*Trumping-Up'* which was a virtue signaling typology that bolstered individual claims on moral issues (Tosi & Warmke, 2016). Future research should assess how other forms of virtue signaling increase or decrease agreement. Reducing negative connotations of virtue signaling, the utilisation of deception may stop participants from realising that virtue signaling was taking place. Participants may have habituated because the stimuli emulated real social media posts by having a similar format, and by displaying similar content such as the achievements of individuals to signal moral character or call attention to moral norm violations (Abele & Wojciszke, 2007; Crockett, 2017; Saltman, 2017). Thus, the current study only assessed mild virtue signaling, in the absence

of moral outrage. A differentiation of typologies would require an additional qualitative analysis.

The study design only assessed net agreement and disagreement. It was unable to measure the full range of human emotions. Cognitive heuristics may have been present due to social desirability bias (McClay, 2018). The current study did not have any checks in place such as a social desirability scale. Individuals may want to signal their virtue even if they do not personally believe or support a cause, a form of false honesty signaling (Grabo et al., 2017). Also, emotional cues such as facial expression, and tone of voice were not easily perceptible, thus virtue signaling support could be easily faked (Levy, 2020). Hence, social desirability bias, cognitive heuristics and measures of real human emotions were not assessed. Future research may wish to consider using opposite coded scales to detect social desirability and measure the discrepancy between what people say and do (McClay, 2018). An observational study can be used to detect discrepancies between what an individual says they feel and what their facial expressions convey.

A limitation to this study was that the virtue signaling stimuli may have limited geographic significance to everyday social media users, especially for social justice and political stimuli based outside of America. The social justice stimuli involved themes relevant to the Black Lives Matter campaign, specific to the United States. Politics stimuli involved examples of Canadian Conservatism. Social media users may not be invested in the social justice or political landscape of certain countries. The study was conducted online and open to participants who used Facebook. Future studies may desire to consider controlling for geography, seeking a specific countries' participants and use virtue signaling stimuli that are specific and seen as important for the national context. Additional future research could compare how different nations exhibit different typologies of online virtue signalling.

The presence of small effect sizes and a lack of significance in social justice may be because the study assessed virtue signalling 'consumers' rather than 'producers', for whom displaying intentionality of support is more costly (Abele & Wojciszke, 2007; Bai, 2017; Smith & Harper, 2003; McClay, 2018). Future studies should assess the actual behaviour as well as intentions, context, and competition that virtue signalling producers find themselves in (Bénabou & Tirole, 2006).

Implications

Several meaningful implications can be drawn from the current study. For instance, political and social justice organisations can virtue signal to maintain support for a cause since the study found virtue signalling buffered against the effects of decreasing agreement. Examples of pro-social virtue signalling include displaying leaders, role models or '*paragons of virtue'* who increase engagement with shared identity, norms, emotions, and visions of a group (Pentland & Heibeck, 2010). This can help spread awareness and promote an organisation by bolstering support for a cause. Virtue signalling can show how political and social justice organisations and individuals actively work to better society, thus are strongly

founded upon a moral identity which is an equally strong leadership pathway like warmth and competence (Goodwin et al., 2014). Virtue signalling could lead to virtue admiration, where individuals who espouse or support political parties or social justice causes are promoted as role models to which the rest of society should emulate (Bai et al., 2020). Virtue signalling has positive attributes, which evolved out of the need to display moral character and excellence and can be used to encourage people to support online causes, foster a community with shared values, and group cooperation (Aristotle, 2011; Inbar et al., 2012; Haidt, 2012). Compared to slogan controls, virtue signalling buffered against the effects of decreased agreement and thus, displaying virtue signalling is a form of conspicuous consumption that signals good impressions to others and provides personal satisfaction (Veblen, 2005; West, 2004). Therefore, virtue signalling can increase awareness, and support for organisations to assist the community and the less fortunate.

There are different severities of virtue signalling. The implications of excessive virtue signalling can lead to moral outrage, dehumanisation and deindividualization which can serve to polarise and further fracture society and its members (Crockett, 2017; Duncan, 2020). The current study found evidence of milder virtue signalling, an example of different severities of virtue signalling, which were affected by different purposes used to persuade others; whether to encourage society to work towards a common goal or create further disagreement and online segregation in echo chambers (Brady et al., 2017; Fincher & Tetlock, 2016). Virtue signalling was found to buffer the effects of decreased agreement compared to control slogans, but this study did not examine a-priori political affiliation which thus would need to be explored in future research. In addition, there was no significant effect of virtue signalling on charities. Future studies may wish to modulate the effect of soft and hard signalling, moral outrage, emotional reaction severity and presence of in-groups and out-groups on virtue signalling and topics.

Generally, individuals want to appear online as their best possible self, and thus mild virtue signalling is an acceptable way to maintain self-esteem, status and positive mental health that promotes moral change in a community; aligning with the Moral Virtue Theory of Status Attainment (Bai, 2017; Bai, 2020). Topics such as politics and social justice were more divisive (Crockett, 2017). Future charities should not rely heavily on virtue signalling, however more research is needed in modulating levels of virtue signalling severity and typology. Hence, this study has implications for future research in evaluating the effectiveness of virtue signalling, controlling for geographic relevance and social desirability, and measuring the costliness of the virtue signalling.***

References

Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology*, *93*, 751–763. http://dx.doi.org/10.1037/0022-3514.93.5.751

Bai, F. (2017). Beyond dominance and competence: A moral virtue theory of status attainment. *Personality and Social Psychology Review*, *21*, 203–227. http://dx.doi.org/10.1177/1088868316649297

Bai, F., Ho, G. C. C., & Yan, J. (2020). Does virtue lead to status? Testing the moral virtue theory of status attainment. *Journal of Personality and Social Psychology, 118* (3), 501-532. http://dx.doi.org/10.1037/pspi0000192

Balon, R. (2020). Are we preparing students for the road? *Academic Psychiatry*, *44* (1), 103-105. https://doi.org/10.1007/s40596-019-01125-8

Barclay, P., (2010). Altruism is a courtship display: Some effects of third-part generosity on audience perceptions. *British Journal of Psychology, 101* (1): 123-135. https://doi.org/10.1348/000712609X435733

Basil, D. Z., Ridgway, N. M., & Basil, M. D. (2006). Guilt appeals: The mediating effect of responsibility. *Psychology & Marketing*, *23* (12), 1035–1054. https://doi.org/10.1002/mar.20145

Bénabou, R., & Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, *96* (5), 1652–1676. https://doi.org/10.1257/aer.96.5.1652

Berman, J. Z., & Small, D. A. (2018). Discipline and desire: On the relative importance of willpower and purity in signalling virtue, *Journal of Experimental Social Psychology*, *76*, 220-230. https://doi.org/10.1016/j.jesp.2018.02.007

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, *114* (28), 7313-7318. https://doi.org/10.1073/pnas.1618923114

Brown, M., Keefer, L.A., Sacco, D.F., & Brown, F.L. (2020). Demonstrate values: behavioural displays of moral outrage as a cue to long-term mate potential, *Emotion*, 1-69. https://doi.org/10.31234/osf.io/qc8sk

Bulbulia, J., & Frean, M. (2010). The evolution of charismatic cultures. *Method and Theory in the Study of Religion*, *22*, 254–271. http://dx.doi.org/10.1163/157006810X531049

Bushman, B. J. (2002). Does venting anger feed or extinguish the flame? Catharsis, rumination, distraction, anger, and aggressive responding. *Personality and Social Psychology Bulletin*, *28* (6), 724-731. https://doi.org/10.1177/0146167202289002

Craker, N., & March, E. (2016). The dark side of Facebook: The Dark Tetrad, negative social potency, and trolling behaviours. *Personality and Individual Differences*, *102*, 79–84. https://doi.org/10.1016/j.paid.2016.06.043

Crockett, M. J. (2017). Moral outrage in the digital age. *Nature Human Behaviour*, *1* (11), 769-771. https://doi.org/10.1038/s41562-017-0213-3

De Cruz, H. (2018, October 25). *What explains moral outrage on social media*? Medium. https://medium.com/@helenldecruz/what-explains-moral-outrage-on- social-media-336189637948

Douglas, M. (2003), *Purity and Danger: An Analysis of Concepts of Pollution and Taboo*, Routledge, Abingdon. https://doi.org/10.4324/9780203361832

Duncan, S. (2020). Why all the outrage? Viral media as corrupt play shaping mainstream media narratives. *Westminster Papers in Communication and Culture*, *15* (1), 37–52. https://doi.org/10.16997/wpcc.317

Fan R, Zhao J, Chen Y, Xu K (2013) Anger Is More Influential than Joy: Sentiment Correlation in Weibo. *PLoS ONE 9* (*10)*, e110184. https://doi.org/10.1371/journal.pone.0110184

Farrell, T., Gorrell, G., & Bontcheva, K. (2020). Vindication, Virtue and Vitriol: A study of online engagement and abuse toward British MPs during the COVID-19 Pandemic. *arXiv preprint arXiv:2008.05261*. https://doi.org/10.1007/s42001-020-00090-9

Fincher, K. M., & Tetlock, P. E. (2016). Perceptual dehumanization of faces is activated by norm violations and facilitates norm enforcement. *Journal of Experimental Psychology: General*, *145* (2), 131. https://doi.org/10.1037/xge0000132

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, *106*, 148 –168. http://dx.doi.org/10.1037/a0034726

Grabo, A., Spisak, B. R., & van Vugt, M. (2017). Charisma as signal: An evolutionary perspective on charismatic leadership. *The Leadership Quarterly*, *28* (4), 473-485. http://dx.doi.org/10.1016/j.leaqua.2017.05.001

Grabo, A., & van Vugt, M. (2016). Charismatic leadership and the evolution of cooperation. *Evolution and Human Behavior*, *37*, 399–406. https://doi.org/10.1016/j.evolhumbehav.2016.03.005

Grace, D., & Griffin, D. (2006). Exploring conspicuousness in the context of donation behavior. *International Journal of Nonprofit and Voluntary Sector Marketing*, *11* (2), 147–154. https://doi.org/10.1002/nvsm.24

Greitemeyer, T., Mügge, D. O., & Bollermann, I. (2014). Having responsive Facebook friends affects the satisfaction of psychological needs more than having many Facebook friends. *Basic and Applied Social Psychology*, *36* (3), 252–258. https://doi.org/10.1080/01973533.2014.900619

Grzywinska, I., & Batorski, D. (2016). How the emergence of social networking sites challenges agenda-setting theory. *Konteksty Spoleczne*, *4* (1-7), 19–32. Retrieved from: https://www.ceeol.com/search/article-detail?id=544107

Hamilton, A. (2019). 'Virtue signalling' and other slimy words. *Eureka Street*, 29(5), 12-13.

Hamilton, W. D. (1964). The genetical theory of social behaviour. I, II. *Journal of Theoretical Biology, 7* (1), 17–52. https://doi.org/10.1016/0022-5193(64)90039-6

Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. New York, NY: Pantheon Books.

Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus, 133*, 55–66. http://dx.doi.org/10.1162/0011526042365555

Henrich, J., & Boyd, R. (2001). Why people punish defectors: Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *Journal of Theoretical Biology, 208* (1), 79-89. https://doi.org/10.1006/jtbi.2000.2202

Henrich, J., Boyd, R. and Richerson, P.J. (2012), *The Puzzle of Monogamous Marriage*, Phil. Hrdy, S.B. (2011), Mothers and others, Harvard University Press. https://doi.org/10.1098/rstb.2011.0290

Henrich, J., Chudek, M., & Boyd, R. (2015). The big man mechanism: How prestige fosters cooperation and creates prosocial leaders. *Philosophical Transactions of the Royal Society B, 370* (1683), 1-13. https://doi.org/10.1098/rstb.2015.0013

Inbar, Y., Pizarro, D. A., & Cushman, F. (2012). Benefiting from misfortune: When harmless actions are judged to be morally blameworthy. *Personality and Social Psychology Bulletin, 38* (1), 52–62. https://doi.org/10.1177/0146167211430232

Jaeggi, A. V., & Gurven, M. (2013). Natural co-operators: Food sharing in humans and other primates. *Evolutionary Anthropology, 22*, 186–195. http://dx.doi.org/10.1002/evan.21364

Kastanakis, M. N., & Balabanis, G. (2012). Between the mass and the class: Antecedents of the "Bandwagon" luxury consumption behavior. *Journal of Business Research, 65* (10), 1399–1407. https://doi.org/10.1016/j.jbusres.2011.10.005

Kotabe, H. P., & Hofmann, W. (2015). On integrating the components of self-control. *Perspectives on Psychological Science*, *10* (5), 618–638. https://doi.org/10.1177/1745691615593382

Kurzban, R., DeScioli, P., & O'Brien, E. (2007). Audience effects on moralistic punishment. *Evolution and Human Behavior*, *28* (2), 75-84. https://doi.org/10.1016/j.evolhumbehav.2006.06.001

Levy, N. (2020). Virtue signalling is virtuous. *Synthese*, 1-18. https://doi.org/10.1007/s11229-020-02653-9

Massey-Abernathy, A. R., & Haseltine, E. (2019). Power Talk: Communication Styles, Vocalization Rates and Dominance. *Journal of Psycholinguistic Research, 48* (1), 107-116. https://doi.org/10.1007/s10936-018-9592-5

Smith, J. M., & Harper, D. (2003). *Animal signals*. Oxford, UK: Oxford University Press.

Mackay, H. (2017). Social media analytics: Implications for journalism and democracy. *Journal of Information Ethics*, *26* (1), 34–48. https://doi.org/10.1086/431089

McClay, B. D. (2018). Virtue Signalling. *The Hedgehog Review*, *20* (2), 141-144.

Messner, M., & Distaso, M. (2008). The source cycle: How traditional media and weblogs use each other as sources. *Journalism Studies*, *9* (3), 447–463. https://doi.org/10.1080/14616700801999287

Miller, G. F. (2007). Sexual selection for moral virtues. *The Quarterly Review of Biology*, *82* (2), 97-125. https://doi.org/10.1086/517857

Pedersen, E.J., Kurzban, R., & McCullough, M.E. (2013). Do humans really punish altruistically? A closer look. *Proceedings of the Royal Society B: Biological Sciences*, *280* (1758), 1-8. https://doi.org/10.1098/rspb.2012.2723

Pentland, A., & Heibeck, T. (2010). *Honest signals: How they shape our world*. Cambridge, MA: MIT press.

Piazza, J., Goodwin, G. P., Rozin, P., & Royzman, E. B. (2014). When a virtue is not a virtue: Conditional virtues in moral evaluation. *Social Cognition*, *32* (6), 528–558. https://doi.org/10.1521/soco.2014.32.6.528

Price, P. C., & Stone, E. R. (2004). Intuitive evaluation of likelihood judgment producers: evidence for a confidence heuristic. *Journal of Behavioral Decision Making*, *17* (1), 39–57. https://doi.org/10.1002/bdm.460

Pulford, B. D., Colman, A. M., Buabang, E. K., & Krockow, E. M. (2018). The persuasive power of knowledge: Testing the confidence heuristic. *Journal of Experimental Psychology. General*, *147* (10), 1431–1444. https://doi.org/10.1037/xge0000471

Righetti, F., & Finkenauer, C. (2011). If you are able to control yourself, I will trust you: The role of perceived self-control in interpersonal trust. *Journal of Personality and Social Psychology*, *100* (5), 874–886. https://doi.org/10.1037/a0021827

Saltman, K. J. (2017). "Privilege-Checking," "Virtue- signalling," and "Safe Spaces": What Happens When Cultural Politics is Privatized and the Body Replaces Argument. *Symplokē*, *26* (1-2), 403-409. https://doi.org/10.5250/symploke.26.1-2.0403

Schau, H. J., & Gilly, M. C. (2003). We are what we post? Self-pres- entation in personal web space. *Journal of Consumer Research*, *30*, 385–404. https://doi.org/10.1086/378616

Strizhakova, Y., Coulter, R. A., & Price, L. L. (2008). The meanings of branded products: A cross-national scale development and meaning assessment. *International Journal of Research in Marketing*, *25* (2), 82–93. https://doi.org/10.1016/j.ijresmar.2008.01.001

Synovitz, R. (2018). *Forget about civil discourse, my keyboard Is outraged*. Radio Free Europe Liberty.

Tangney, J. P. (2000). Humility: Theoretical perspectives, empirical findings and directions for future research. *Journal of Social and Clinical Psychology*, *19*, 70–82. http://dx.doi.org/10.1521/jscp.2000.19.1.70

Tian, K. T., Bearden, W. O., & Hunter, G. L. (2001). Consumers' need for uniqueness: Scale development and validation. *Journal of Consumer Research, 28* (1), 50–66. https://doi.org/10.1086/321947

Tosi, J., & Warmke, B. (2016). Moral grandstanding. *Philosophy & Public Affairs, 44* (3), 197–217. https://doi.org/10.1111/papa.12075

Trivers, R.L. (1971). The evolution of reciprocal altruism. *The Quarterly review of biology, 46* (1), 35-57. https://doi.org/10.1086/406755

Uhlmann, E. L., Zhu, L., & Diermeier, D. (2013). When actions speak volumes: The role of inferences about moral character in outrage over racial bigotry. *European Journal of Social Psychology, 44* (1), 23-29. https://doi.org/10.1002/ejsp.1987

Veblen, T. (2005). *Conspicuous consumption,* vol. *38*. Penguin UK.

Veissière, S. P. L. (2018). "Toxic Masculinity" in the age of# MeToo: ritual, morality and gender archetypes across cultures. *Society and Business Review, 13* (3): 274-286. https://doi.org/10.1108/SBR-07-2018-0070

Wallace, E., Buil, I., & De Chernatony, L. (2018). 'Consuming good' on social media: What can conspicuous virtue signalling on Facebook tell us about prosocial and unethical intentions? *Journal of Business Ethics, 162* (3), 577-592. https://doi.org/10.1007/s10551-018-3999-7

Walumbwa, F. O., Avolio, B. J., Gardner, W. L., Wernsing, T. S., & Peterson, S. J. (2008). Authentic leadership: Development and validation of a theory-based measure. *Journal of Management, 34*, 89 –126. http://dx.doi.org/10.1177/0149206307308913

Wallsten, K. (2007). Agenda setting and the blogosphere: An analysis of the relationship between mainstream media and political blogs. *Review of Policy Research, 24* (6), 567–587. https://doi.org/10.1111/j.1541-1338.2007.00300.x

West, P. (2004). *Conspicuous compassion: Why sometimes it really is cruel to be kind*. London: Civitas, Institute for the Study of Civil Society.

Willer, R. (2009). Groups reward individual sacrifice: The status solution to the collective action problem. *American Sociological Review, 74*, 23–43. http://dx.doi.org/10.1177/000312240907400102

Zahavi, A. (1977). The cost of honesty: further remarks on the handicap principle. *Journal of Theoretical Biology, 67* (3), 603–605. https://doi.org/10.1016/0022-5193(77)90061-3