We need deeper understanding about the neurocognitive mechanisms of moral righteousness in an era of online vigilantism and cancel culture

Rocco Chiou
MRC Cognition and Brain Sciences Unit, University of Cambridge
Email: Rocco.Chiou@mrc-cbu.cam.ac.uk

Cancel culture has become increasingly prevalent in recent years. It follows a typical process, usually unfolding on the internet: When a public figure says or does something considered offensive or pejorative to a given group (e.g., ethnic minorities, sexual/gender minorities, people with disabilities, women as minorities, and so forth), disparaging comments quickly pile up on social media, *calling out* the misconduct, withdrawing support for the person's work/product, or using performative language to mock and shame the person believed to be responsible for the wrongdoing.

Cancel culture is a double-edged sword. Social media offers a platform for people to easily hold someone accountable for their misbehaviour and raises awareness about injustice. Opinions favouring *cancel culture* posit that it is akin to activism, using online platforms to criticise influential people for promoting bigotry. On the flipside, *cancel culture* can sometimes become digital vigilantism. Advocates of *cancelling* public figures often believe that they hold the moral high ground – they are entitled even to disparage or humiliate someone on the grounds of political correctness. However, such public shaming on many occasions can be excessive and simply becomes a way of judging and rejecting anyone who holds a different socio-political viewpoint. This phenomenon has led to concerns (e.g., one recently expressed by the former President Barack Obama) that it has detrimental impacts on society, particularly on young people.

The mentality behind *cancel culture* is some form of 'moral righteousness' that people believe that it is morally justifiable to denounce someone who is morally inferior and deserves the criticism. In extreme cases, moral righteousness leads to ruthless harmful behaviour because their moral compass convinces some people that such violence is a necessary evil. Although the psychological and neural processes behind such mentality are key to understanding human behaviour, there is little research on this important topic. As a first step, a recent study by Workman, Yoder, and Decety (2020) in *AJOB Neuroscience* links people's attitudes on various socio-political affairs with neural activity.

Using functional magnetic resonance imaging (fMRI), Workman *et al.* (2020) asked participants to view various photographs of violent protests while acquiring their fMRI data. The violent events portrayed in the photographs could be purportedly congruent or incongruent with a participant's socio-political ideology. After viewing and evaluating the photograph, participants pressed a button to indicate the appropriateness of the violence on a 7-point scale. Results showed a parametric, correlative relationship between the brain's reward/valence system and appropriateness rating. Specifically, the level of fMRI activation in the ventromedial prefrontal cortex and ventral striatum, two neural structures heavily involved in

the neural processes of desirable or positive outcome, is positively correlated with participants' decision on the appropriateness of a violent event, with events rated as more appropriate eliciting grater activity in these brain regions. Moreover, prior to scanning, participants were asked to indicate how strongly a socio-political issue (which was later depicted in photographs during the fMRI experiment) reflects their personal moral convictions, again using a 7-point scale. The authors found that the scores of moral conviction rating were *positively* correlated with activation level of the ventral striatum and *negatively* correlated with the amygdala. Research on the functionality of ventral striatum and ventromedial prefrontal cortex indicates the significance of these regions in associating incoming stimuli with subjective value (e.g., Liljeholm and O'Doherty 2012). Thus, the authors interpreted the positive correlations detected in these areas as reflecting the integrative processes that marry subjective value (the extent of goodness *vs.* badness) with the neural representations of a socio-political event. By contrast, the amygdala has been known to be implicated in processing emotionally arousing stimuli (often negative stimuli, such as scenes of violent protests; Murray, 2007). Thus, the negative correlation detected in the amygdala was interpreted by the authors as reflecting regulatory processes that dampen the brain's aversive response to negative events when these negative events are congruent with one's socio-political beliefs. Taken together, when a violent action is congruous with one's moral convictions, the brain represents it by amplifying the neural response of the reward system and moderating the aversive response of the emotion system.

The study by Workman and colleagues (2020) is a step forward to understanding the neurocognitive basis of moral righteousness and support for violence as an excusable measure. These mechanisms might be completely separable from one's position on the ideological spectrum: Conservative voters believe that a police officer applying violence during the arrest of an unarmed person reflects the enforcement of law and order, whereas progressive voters believe environmentalists using violence in protest against an oil company is a laudable deed. While these attitudes represent binarily opposite views on the ideological spectrum, the underlying reasoning process and neural computation might be similar – the brain may assign a positive valence to such violent behaviour and reduce the negative emotion that it evokes. As a consequence of these operations at the neural level, violence becomes less aversive and more acceptable when it fits one's moral convictions.

More research is necessary to elucidate the complex relationship between the neural basis of moral cognition (Moll et al. 2005) and the evaluative processes that make violent behaviour more palatable. It has been shown that moral cognition relies on various regions of the brain's semantic and default-mode systems (Bzdok et al. 2012). Moral cognition is a multifaceted process that entails decision-making, semantic knowledge, and theory of mind. Conceivably, the neural networks involved in socio-semantic processing (Chiou and Lambon Ralph 2019, Chiou, Humphreys, and Lambon Ralph 2020, Chiou et al. 2018) and emotional regulation (Dixon et al. 2017) would be key components of a broad neural architecture that interact with the reward and valence systems when one evaluates a socio-political issue based on existing framework of semantic knowledge.

Cancel culture on the social media is also an example that embodies how a sense of moral righteousness can lead to aggressive online behaviour, which can sometimes be vigilantist in nature. In an era in which people increasingly resort to *cancel culture* and other forms of digital

vigilantism to condemn and invalidate people believed to be morally wrong, a deeper understanding about the neurocognitive mechanisms of moral righteousness is critically necessary to offer a comprehensive account for the complexity of contemporary human behaviour in the digital space and in the physical world.

References

Bzdok, Danilo, Leonhard Schilbach, Kai Vogeley, Karla Schneider, Angela R Laird, Robert Langner, and Simon B Eickhoff. 2012. "Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy." *Brain Structure and Function* no. 217 (4):783-796.

Chiou, Rocco, Gina F Humphreys, JeYoung Jung, and Matthew A Lambon Ralph. 2018. "Controlled semantic cognition relies upon dynamic and flexible interactions between the executive 'semantic control'and hub-and-spoke 'semantic representation'systems." *Cortex* no. 103:100-116.

Chiou, Rocco, Gina F Humphreys, and Matthew A Lambon Ralph. 2020. "Bipartite functional fractionation within the default network supports disparate forms of internally oriented cognition." *Cerebral Cortex* no. 30 (10):5484-5501.

Chiou, Rocco, and Matthew A Lambon Ralph. 2019. "Unveiling the dynamic interplay between the hub-and spoke-components of the brain's semantic system and its impact on human behaviour." *NeuroImage* no. 199:114-126.

Dixon, Matthew L, Ravi Thiruchselvam, Rebecca Todd, and Kalina Christoff. 2017. "Emotion and the prefrontal cortex: an integrative review." *Psychological bulletin* no. 143 (10):1033.

Liljeholm, Mimi, and John P O'Doherty. 2012. "Contributions of the striatum to learning, motivation, and performance: an associative account." *Trends in cognitive sciences* no. 16 (9):467-475.

Moll, Jorge, Roland Zahn, Ricardo de Oliveira-Souza, Frank Krueger, and Jordan Grafman. 2005. "The neural basis of human moral cognition." *Nature reviews neuroscience* no. 6 (10):799-809.

Murray, Elisabeth A. 2007. "The amygdala, reward and emotion." *Trends in cognitive sciences* no. 11 (11):489-497.

Workman, Clifford, Keith Yoder, and Jean Decety. 2020. "The dark side of morality: Neural mechanisms underpinning moral convictions and support for violence." *AJOB Neuroscience*. doi: 10.31234/osf.io/t69w4.