

Would you notice if fake news changed your behavior? An experiment on the unconscious effects of disinformation

Zach Bastick

► To cite this version:

Zach Bastick. Would you notice if fake news changed your behavior? An experiment on the unconscious effects of disinformation. Computers in Human Behavior, 2021, 116, pp.106633 -. 10.1016/j.chb.2020.106633 . hal-03493920

HAL Id: hal-03493920 https://hal.science/hal-03493920

Submitted on 15 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Would you notice if fake news changed your behavior?

An experiment on the unconscious effects of disinformation

Dr. Zach Bastick

Office C413, European School of Political Science,

Lille Catholic University, 60 Boulevard Vauban, 59016 Lille, France

zach.bastick@gmail.com

Phone: +33 623163379

June 27, 2020

Abstract

A growing literature is emerging on the believability and spread of disinformation, such as fake news, over social networks. However, little is known about the degree to which malicious actors can use social media to covertly affect behavior with disinformation. A lab-based randomized controlled experiment was conducted with 233 undergraduate students to investigate the behavioral effects of fake news. It was found that even short (under 5-minute) exposure to fake news was able to significantly modify the unconscious behavior of individuals. This paper provides initial evidence that fake news can be used to covertly modify behavior, it argues that current approaches to mitigating fake news, and disinformation in general, are insufficient to protect social media users from this threat, and it highlights the implications of this for democracy. It raises the need for an urgent cross-sectoral effort to investigate, protect against, and mitigate the risks of covert, widespread and decentralized behavior modification over online social networks.

Keywords: Fake news, disinformation, misinformation, behavior, finger tapping, emotions

1. Introduction

Online platforms increasingly mediate public discourse. Their algorithms help citizens to join social groups, sort through the noise of public discourse, and stay abreast of relevant current events. However, these algorithms also enable filter-bubbles that risk distorting reality, suppressing contrarian views, and fragmenting the public sphere (Flaxman et al., 2016). In this context, virality often combines with ambiguity, producing information cascades where users share information with few guarantees of accuracy. The organic spread of misinformation online is amplified by private interests seeking political or financial rewards, that micro-target vulnerable individuals as seeds to further spread misinformation, risks impacting society, skewing markets, and, ultimately, subverting democracy. This systemic risk assumes that disinformation can generate effects that, even if small at the individual level, aggregate sufficiently to produce large-scale outcomes. While concerted attempts to manipulate public opinion through disinformation are prevalent worldwide (Bradshaw & Howard, 2018), there is less evidence on the capacity of online disinformation to directly and covertly manipulate behavior.

While both misinformation and disinformation refer to forms of factually inaccurate information, disinformation involves an intention to deceive (Pal & Banerjee, 2019). Fake news is a particularly potent vehicle for disinformation – one that masquerades as a journalistic article and, as such, usurps the credibility of journalism, the timeliness of content, and the ability to push articles on sensitive topics that include politics into platform newsfeeds (Tandoc et al., 2018). Identifying fake news can be challenging, not the least because the intent to deceive can be difficult to differentiate from the genuine portrayal of a controversial perspective (Wang et al., 2019). However, this article focuses on disinformation, and takes the specific example of fake news, and as such assumes content that is deliberately created with the intent to deceive.

The notion that online disinformation can produce real-world changes in behavior gained public attention with the 2018 Cambridge Analytica scandal, where the data of 87 million Facebook users were secretly harvested and used for political advertising in the Brexit referendum and the 2016 US presidential election (Isaak & Hanna, 2018). In 2019, the Mueller Report explicitly cited Facebook and Twitter as having been used by Russian agents to spread disinformation to interfere with the US election and incite real-world violence (Mueller, 2019). Disinformation on the Pizzagate conspiracy theory has been attributed as a cause of the 2016 shooting in a Washington D.C. pizzeria (J. Johnson, 2018), and disinformation on mail-in ballots and the QAnon conspiracy theory is speculated to have impacted the 2020 US presidential election (Klar, 2020). Similarly, it is claimed that disinformation on the coronavirus pandemic may have reduced mask wearing and social distancing (Ioannidis, 2020; Kouzy et al., 2020; Mattiuzzi & Lippi, 2020).

Nevertheless, it remains unclear to what extent disinformation can modify behavior in practice. First, governments and scholars typically fear undue influence on complex behaviors, such as voting, but rarely consider more basic behaviors or capacities, such as typing speed or attention span. These can cascade into more complex behaviors or accumulate into socially significant outcomes at scale. Second, disinformation is predominantly viewed as a threat to the rational processes of sensemaking and decision making, but rarely as a threat to unconscious processes that underly behavior. It is understood that emotions can unconsciously affect behavior and an emerging literature on implicit cognition holds that we hold implicit attitudes that can affect our behavior (Dolan, 2002; Gawronski et al., 2006; Greenwald & Banaji, 1995; Kunst-Wilson & Zajonc, 1980). Here, I examine basic changes in behavior that can be attributed to influences over these emotional associations and implicit attitudes - that is, the unconscious impact of disinformation on behavior.

The need to understand how disinformation can modify behavior is accentuated by online platforms. These allow anyone, from individuals to nation states, to distribute disinformation intended to modify the behavior of others. The concentration of user attention by a small group of platforms also suggests that platforms themselves could modify behavior, and this deserves attention for the same reasons that underly contemporary scrutiny over algorithmic governance and content moderation (Fung, 2020; Gorwa et al., 2020; O'Neil, 2016; Pasquale, 2015). Experimental evidence finds that platforms can manipulate the emotions of their users (Kramer et al., 2014), and scholars theorize that platforms can 'hyper-nudge' individuals towards predictable behaviors (Yeung, 2017). When these behavioral nudges are targeted, they may decide low-margin societal outcomes, such as elections (Zittrain, 2014). Moreover, the democratization of artificial intelligence (AI) enables increasingly effective disinformation – raising speculation that deep fakes, for example, may be able to implant false memories and manipulate individuals towards nefarious actions (Liv & Greenbaum, 2020). Hence, without a nuanced understanding of the cognitive effects of disinformation, the array of possibilities against which society must guard itself becomes insuperable.

The prospect of disinformation-driven dystopias highlights the urgent need to understand the capacity of online disinformation to modify individual behaviors and funnel societal outcomes. Nevertheless, there is a paucity of studies on the capacity of disinformation to modify behavior, particularly outside of conscious awareness. This paper aims to provide initial evidence of this

capacity. It reports the results of a randomized controlled experiment that tests whether short exposure to fake news can modify the unconscious behavior of participants – in particular, it isolates the ability of fake news to unconsciously influence the results of participants engaged in a Finger Tapping Test (FTT), a standard neuropsychological test of cognitive and motor function. This paper aims to initiate a research agenda that explores the capacity for digital disinformation to modify behavior and that investigates mechanisms to protect the public from this threat.

2. Theoretical Framework

2.1 Empirical effects of fake news on behavior

The need to evaluate fake news mitigation mechanisms for online platforms has produced a growing body of research on the capacity of fake news to influence beliefs. Experimental research in this area typically tests the believability of misinformation, on the one hand, against the believability of misinformation combined with a mitigation mechanism, on the other. It is driven by four principal streams of research. In one stream, online experiments have found that fact-checking labels on Twitter-based anti-vaccine misinformation can increase the acceptance of vaccines (Featherstone et al., 2019), and that credibility labels and political stance labels on fake news subtly affect its perceived trustworthiness (Gao et al., 2018). A second stream of research focuses on the believability of fake news after it is corrected. For instance, one such study found that public health authorities were more effective in correcting misinformation than other platform users (Vraga & Bode, 2017), while another study found that corrections by algorithms were as effective as those by other users (Bode & Vraga, 2018). A third stream of research leverages cognitive psychology literature on the 'continued influence effect' (H. M. Johnson & Seifert, 1994), to test whether misinformation continues to influence inferences even after it is corrected (Desai & Reimers, 2018; Rich & Zaragoza, 2016; Thorson, 2016). Finally, a fourth stream focuses on the effects of fake news on attitudes, such as a study by Balmas (2014) which found evidence of an indirect positive association between viewing fake news and feelings of inefficacy, alienation and cynicism towards politicians.

This capacity of fake news to influence beliefs may in turn influence behaviors (Gross, 2017; Levy, 2017). This is supported by the strong tradition in the behavioral sciences of investigating the rational and social determinants of behavior. For instance, social psychology has examined the influence of trust and education on prosocial behavior (Bierhoff, 2002; Twenge et al., 2007) and political communication has examined the influence of rhetorical and affective persuasion on voting (Clarke et al., 2006; Wahl-Jorgensen, 2019). However, inferring causality in complex behaviors is complicated by response biases in self-reports (Fisher, 1993; Podsakoff & Organ, 1986), the limitations of applying rational decision-making models to everyday behaviors (Back, 1961; Green & Shapiro, 1996), the influence of unconscious beliefs (Lodge & Taber, 2013), and the interaction of confounding variables (Ewert & Sibthorp, 2009). As a result, it has been difficult to predict complex behaviors from the rational processing of specific misinformation. This begs approaches that analyze behavior as an intricate process rather than a discrete outcome, such as that adopted by this current study.

Research on fake news believability has been complemented by an emerging research stream on behavior. However, while social media data enable analyzing behavior stemming from exposure

to fake news, the extant literature has so far been mostly limited to the propensity of individuals to reshare misinformation. For instance, Guess et al. (2019) reported that age and political affiliation predict fake news sharing behavior. Among healthcare workers, Chau and Banerjee (2018) found that personal involvement predicts sharing of healthcare rumors. Boehmer and Tandoc (2015) found sharing to be motivated by the perceived interests of an article to the user and their audience. Colliander (2019) found users were less likely to share fake news stories that produced critical feedback in the post's comments. While this research is consequential for identifying groups most exposed or susceptible to fake news, a more psychologically nuanced framework is needed to understand the capacity of fake news to covertly modify behavior that itself might not be conscious, especially behaviors other than sharing that might be targeted by a fake news article.

2.2 Unconscious attitudes and behavior

Evidence supports the possibility for online fake news to affect unconscious attitudes with the aim of manipulating behavior. In recent years, psychology has emphasized that unconscious attitudes can affect cognitive, emotional, and moral processes. Greenwald and Banaji (1995, p. 8) define implicit attitudes as "introspectively unidentified (or inaccurately identified) traces of past experience that mediate favorable or unfavorable feeling, thought or action towards social objects". Gawronski, Hoffman and Wilbur (2006) elaborated that people may be unaware of the source of these attitudes, their content, or their impact on their cognitive processes or on their decision-making. Likewise, Wilson (2000) presented a model in which implicit attitudes co-exist with explicit attitudes, and which further specifies that a holder of an implicit attitude may be unaware of its existence. These conceptions hold in common that implicit attitudes are automatic and may guide or directly affect behavior unless consciously prevented from doing so (Bargh et al., 1996; Ajzen & Fishbein, 2005; Fazio, 1990; Fazio & Olson, 2014; Fishbein & Ajzen, 2011; Greenwald & Banaji, 2017). This aligns with correlational evidence on implicit attitudes and eating behavior (Hofmann & Friese, 2008), consumer behavior (Maison et al., 2004), exercise dependency and intensity (Forrest et al., 2016), as well as that summarized by meta-reviews (e.g. Greenwald et al., 2009).

This current study uses three interventions to modify implicit attitudes and produce predictable behavioral outcomes. The notion that implicit attitudes are malleable (Dasgupta & Rivera, 2008) has led to identifying interventions that modify implicit attitudes to produce predictable behavioral outcomes. For instance, multiple studies in behavioral medicine propose influencing implicit attitudes towards physical activity as a way of promoting physical wellbeing (Chevance et al., 2017, 2018; D. E. Conroy et al., 2010; Forrest et al., 2016). Similar interventions have been proposed to promote healthy dietary behavior (Hollands et al., 2011; Houben et al., 2010; Mattavelli et al., 2017) and reduce smoking (Lee et al., 2017). In the social sciences, Dasgupta and Rivera (2008) suggest that positive media exposure can modify behavioral intentions and reduce implicit bias against outgroups when members of those outgroups are not present or visible in communities.

2.3 Computer measurement of unconscious behavior

Tools have been developed to capture unconscious attitudes that go unidentified by self-reports (Strohminger et al., 2014). These include the Implicit Association Test (IAT; Greenwald et al.,

1998) and the Affect Misattribution Procedure (AMP; Payne et al., 2005). Additionally, the social sciences have recently seen renewed interest in studying more fundamental reactions, such as those produced by the autonomic nervous system (Settle et al., 2020). These can be measured using heart rate monitors, electrocardiograms (ECGs), electrodermal activity (EDA) sensors, and eye movement trackers. Long applied in psychophysiology, this has led to the burgeoning of new fields, such as political physiology, which observes cause and effect relations between bodily responses and political stimuli (Gruszczynski et al., 2013; Wagner et al., 2015). These tools to capture implicit attitudes and physiological responses aim to bypass the layers of assumptions imbued in self-reports. Instead of inferring disinformation-induced behavior changes from self-reported changes in attitudes or intentions, the experiment reported in this article directly observes changes in behavior using a Finger Tapping Test (FTT).

The FTT is a test of cognitive and motor function that is well established in neuropsychological research. It was initially included in the Halstead Battery (Halstead, 1947) and still ranks sixth among the top 20 general assessment tests used by neuropsychologists, partly due wide evidence of its reliability and validity (Strauss et al., 2006). Participants are traditionally asked to tap a finger as rapidly as possible (Cousins et al., 1998), but they can also be externally-guided through a pacing stimulus (such as a metronome) or internally-guided through self-timing (Tomassini et al., 2018; Witt et al., 2008). Finger movements can be visually observed, but refined measures can be collected through a computer, precision image-based motion analyzer (PRIMAS), or specialized device (Jobbágy et al., 2005). Differences in tapping speed and consistency have been used to identify the laterality of brain lesions (Strauss et al., 2006), as well as to study cognitive impairment in Parkinson's disease (Jobbágy et al., 2005), visual and auditive reception (Jäncke et al., 2000), Huntington's disease (Antoniades et al., 2012), and schizophrenia (Carroll et al., 2009). Attempts to feign cognitive impairments can be detected as suspect effort in the FTT, enabling the FTT to bypass assumptions of subject credibility imbued in self-reports (Arnold et al., 2005).

Finger tapping has been repeatedly linked to cognitive influence as well as neural and biomechanical factors, but the cognitive processes underlying finger tapping are not well understood (Austin et al., 2012). Strauss et al. (2006, p. 1050) note that the FTT "has significant cognitive contributions" and that "speed, coordination, and pacing requirements of finger tapping can be affected by numerous factors". Studies of covert behavior modification have long highlighted the processes by which repeated self-verbalizations (e.g. 'slower', 'faster') impact tapping speeds (Elson, 1975; Meichenbaum, 1977). The association between cognition and simple motor control, more generally, is gaining increased attention, including regarding dual-tasking and demands on working memory (Gidley Larson & Suchy, 2015). There is emerging evidence of cognitive influence on finger tapping stemming from chunking sequences of taps into subsequences (Loehr & Palmer, 2007), cognitive attention (Rabinowitz & Lavner, 2014), and personality traits (Taylor & McFatter, 2003). From this it is expected that increasing tapping effort and speed can be expected to increase cognitive influence in the FTT.

In parallel, it is widely understood that emotions manifest in recognizable behavioral patterns and affect nearly all aspects of cognition (Dolan, 2002). Emotions have been shown to modulate responses of the autonomic nervous system, such as sweating and heart-beat (Kreibig, 2010). As emotions occur outside of rational thought, they are less susceptible to intention than other

psychological processes (Dolan, 2002). Based on the 'mere exposure' phenomenon, Kunst-Wilson and Zajonc (1980) demonstrated that emotions can be processed from stimuli without being consciously identified. This has led to the wide use of affective priming in behavioral research (Klauer & Musch, 2003) and suggests that emotions can moderate (both simple and complex) behaviors without our knowledge (Kreibig, 2010; W. Li et al., 2007). While evidence of affective priming in finger tapping is sparse (Grecucci et al., 2011), Barut et al. (2013, p. 167) acknowledge that "the movement involved in single-finger tapping is complex and is affected by visual and auditory stimuli, emotional and physical health, and factors that impact the skeletal and nervous systems", and so recommend that the FTT "be applied in an isolated testing environment" (Barut et al., 2013, p. 170).

Finger tapping is an appropriate task for this study as it is a complex behavior that can be conducted (with the conscious intention of tapping) at maximum speed. The current experiment is based on supposed unconscious cognitive and affective influence on the FTT, as well as the ability of the FTT to detect changes in suspect effort more directly than can be detected by self-reports. The current experiment deployed a FTT in which participants were asked to tap a single finger on a computer key at their maximum speed before and after being exposed to a fake news stimulus whose emotional valence was intended to alter their tapping speed. A specialized testing environment controlled environmental factors (as per Barut et al., 2013). This experimental paradigm ensured that any speed changes due to exposure to fake news could be identified as due to unconscious exposure to the emotional valence of fake news. In short, this allowed examining whether fake news could unconsciously modify the speed at which an individual taps a key, even if they are consciously attempting to tap at their maximum speed.

2.4 Affective dimensions of fake news

The fake news stimuli used in this current experiment intentionally differed in their emotional salience. Emotions contribute to shaping our interactions with machines (Hilbert et al., 2018; Rahwan et al., 2019) and our perceptions of the algorithms that mediate our online experiences (Bucher, 2017). Experimental evidence demonstrates that platforms can proliferate emotions among users through selective exposure to newsfeed content (Kramer et al., 2014). Emotions associated with disinformation (such as anger) may increase engagement on platforms, in turn driving newsfeed and recommendation algorithms that optimize for engagement to prioritize disinformation (Garrett, 2016). The actual prevalence of emotion in disinformation may be unintentional or due to over-emphasizing the emotions intended to be conveyed (N. J. Conroy et al., 2015). It may also be an intentional tool of affective manipulation, drawn from affective intelligence theory (Marcus et al., 2000) and seen in political disinformation (Bakir & McStay, 2018). This salience of emotion in disinformation has led to the inclusion of affective attributes in AI models built to detect fake news (Shu et al., 2017), to identify deceptive online opinions (Guess et al., 2018; Y. Li et al., 2016; Ott et al., 2013; Hajek et al., 2020), and to determine the credibility of online information sources (Liu et al., 2020).

In addition to contributing to the spread of misinformation, emotions may affect users' susceptibility to misinformation. For example, anger and anxiety have been found to modulate both the propensity to seek out opposing views (Wollebæk et al., 2019) and to interpret information according to partisan lines (Weeks, 2015). Affective susceptibility to misinformation

is also supported by experimental evidence that negative misinformation has greater memory recall than positive misinformation (Porter et al., 2008, 2010). Chua and Banerjee (2018) found that negative (fear based) heath misinformation was more effective than positive (wish based) rumors.

While deceptive content and susceptibility to misinformation is typically associated with negative emotions (Newman et al., 2003), the evidence-base diverges as to the influence of emotional polarity (Burnap et al., 2014; Ghanem et al., 2020). Heiss (2019) found that negative emotions, positive emotions, and tonality each affected different types of online engagement differently. This latter and more nuanced interpretation parallels the importance of context in mediating the effects of emotions – as can be understood through the concept of response shift (Howard & Dailey, 1979). This is partly demonstrated, for instance, by physiological experiments that find emotional differences in bilingual speakers lying in different languages (Caldwell-Harris & Ayçiçeği-Dinn, 2009), but can otherwise be supposed to occur between different platforms (such as Twitter versus Facebook) and between different communities on those platforms.

Due to the salience of emotion for disinformation, the fake news articles employed as stimuli in this current experiment intentionally emphasize emotional arousal (affective excitement) and vary by valence (positive or negative affect). A non-directional research question is formulated regarding the distinct behavioral effects of positive and negative valence to account for the nuance of evidence in this area.

2.5 Research Question

Based on this theoretical framework, the following research question was explored, and its null hypothesis was tested: Can fake news modify the unconscious behavior of individuals?

3. Method

3.1 Participants

A laboratory experiment was conducted with 233 participants. The participants were undergraduate students at a political science school in a large private university in France. They participated as part of the experimental research methods component of a required, first-year course. All the participants were aware that they could leave the experiment at any time without penalty. The participants ranged in age from 17 to 21 (M=18.45, STD=0.65) and were mostly female (71.7%). Most reported being right-handed (81.5%) as opposed to left-handed (11.2%) or ambidextrous (7.3%). All the participants were fluent in English, enabling the experiment to be conducted in English.

3.2 Design

The study employed a standard, 3-treatment group, Randomized Controlled Trial (RCT) design. The participants were randomly distributed across two experimental groups and one control group. The experimental treatments were emotionally salient fake news articles intended to either increase or decrease the maximum speed at which participants tap a key on a computer

(called Maximum Tapping Speed, or MTS). The control treatment was a statistical primer that did not mention tapping speed.

Both the participants' actual MTS (measured by a finger tapping task) and expected MTS (selfestimated by the participants) were collected. The effectiveness of the treatment was measured by the difference in actual MTS between the pretest and the posttest (minus practice differences captured by the control group). As a check on internal validity – particularly, to check that the treatment effect was unconscious – the difference in self-estimated MTS between the pretest and the posttest (minus differences captured by the control group) was calculated.

The considerable advantages of the laboratory setting allowed for no treatment swaps, no dropouts, no non-compliance, and no missing data. This supported the powerful Intention To Treat (ITT) analysis of data (Armijo-Olivo et al., 2009).

3.3 Setting

Participants sat in front of a laptop within a cubicle to promote privacy and minimize interference. All participants were provided with the same laptop model (Lenovo ThinkPad Yoga 12). This minimized the potential for different keyboards to result in different tapping speeds. No significant difference was found between the computers on the actual MTS pretest scores (F(23,209) = 0.81, p=0.721). Each computer was randomly assigned to a treatment using a random number generator.

3.4 Procedure

After sitting down at a computer, the participants were given a brief introduction, and informed that they would see videos of actors tapping a computer key at their maximum speed. They were informed that they would have to estimate the speed at which the actors tapped, as well as their own maximum tapping speed. They were informed that they would then be asked to tap this same key at their maximum speed. Participants were encouraged to tap as fast as possible and to not be concerned with tapping too quickly, being too loud, or damaging the laptops. They were assured that their responses would be collected anonymously through a randomly generated identifier. After obtaining consent through computer input, the participants completed demographic questions (providing their age, sex, and handiness).

In the pretest phase, participants saw 4 video clips. Each clip was 15 seconds long and showed an actor tapping quickly on a computer key. After each video, the participants were asked to provide an intuitive estimate of the number of times the actor was able to tap the key. This served as a calibration activity to provide the participants with a gauge of tapping speed. Each participant was then asked to estimate how many times they themselves could tap within 15 seconds. This was recorded as pretest self-estimated MTS. Each participant was then asked to tap as fast as possible for 15 seconds. Taps were recorded after a 5 second initiation period, and recorded as pretest actual MTS.

In the treatment phase, the participants were shown their group's stimulus (either the positive fake news article, the negative fake news article, or the control text) and informed they would be tested on the material. This instruction was intended to encourage the participant to focus on the factual content as a distraction from the intended emotional effect. After reading the treatment,

they completed a multiple-choice test. Selecting an answer produced immediate feedback on whether the answer was correct, allowing participants to quickly select the correct answers. This aimed to reduce potential test-performance stress.

The posttest followed the same procedure as the pretest. The participants gauge of tapping speed was recalibrated by being shown the same 4 videos (without being informed these were repetitions) and asked to estimate the number of times the actor tapped after each video. This recalibration activity provided a 2-minute filler between the end of the treatment and the start of the post-test, so that potentially highly variable post-treatment responses would be replaced with more reliable, matured, and stable responses. They then estimated their own maximum taping speed, recorded as posttest self-estimated MTS. Finally, they were instructed to tap as fast as possible for 15 seconds. After a 5-second initiation period, this was recorded as posttest actual MTS.

The participants then completed a short feedback survey, before being debriefed at a later meeting.

3.5 Calibration Activity

The actors in the videos were 4 students from a different school than the participants. Each actor was compensated with 20 euros for their participation. Each video showed only the keyboard and the hand of the actor. The videos showed the same laptop model as those used by the participants. The actors tapped fast enough to ensure that the taps could not easily be counted, but slowly enough to ensure that the framerate captured the movements of the finger and the depression of the key. This speed ensured that participants provided an intuitive estimate of the taps rather than a count, the latter of which would be less susceptible to emotional modification. This (1) provided a participant-derived gauge of tapping speed, (2) trained the participants to intuitively estimate (rather than calculate) their own MTS and (3) provided a stabilizing delay after treatment to capture more reliable treatment effects.

3.6 Stimuli development

A normative approach was adopted to design the experimental stimuli. This follows from the understanding that fake news is both an evolving genre and one which is non-trivial to comprehensively describe empirically at any single point in time. As such, the experimental stimuli were designed to adhere to four constraints that reflect those of fake news: (1) Being a fabricated news article; (2) Being annotated as originating from social media; (3) Being read over a screen; (4) Being readable in a short amount of time. The participants were introduced to the texts by clicking to the next screen of the experiment, where they were informed that the article had been shared on Facebook, and where they could read the first screen of computer text. Participants could navigate through the screens at their own pace but could not return to the previous screens.

The experimental texts were constructed around emotional keywords selected from Warriner, Kuperman and Brysbaert's (2013) database of 13,915 English words scored on emotion. Positive and exciting keywords were selected for the positive treatment (valence>7, arousal>6, scale 1-9). Negative and exciting keywords were selected for the negative treatment (valence<2, arousal>6, scale 1-9). The positive fake news article that resulted from these keywords claimed that high Page 9 of 25

MTS was a trait of socially successful individuals. It associated high MTS with high intelligence, winning, positive relationships, and happiness. In counterpart, the negative fake news article claimed that fast MTS was a trait of criminals. It associated fast MTS with brutality, deviance, and violence. Stimuli word cloud analyses are provided in Figure 1. The control text was the introduction of the book *Spurious Correlations* (Vigen, 2015).

All stimuli spanned 4 screens of the article text. They were followed by 1 screen of authoritative references and 1 screen of a comprehension test. The articles were of similar word length (M=836, SD=68), at the B2 level of the Common European Framework of Reference for Languages (CEFR), and of comparable readability (Gunning Fog M=11.73, SD=0.74; Flesch Reading Ease M=52.77, STD=1.66). The average (participant-selected) exposure time to the stimuli across groups was 4 minutes and 43 seconds (SD=114.27 seconds), and the exposure time was not significantly different between the treatment groups (F(2,230) = 1.01, p=0.367, $\eta^2 = 0.01$).

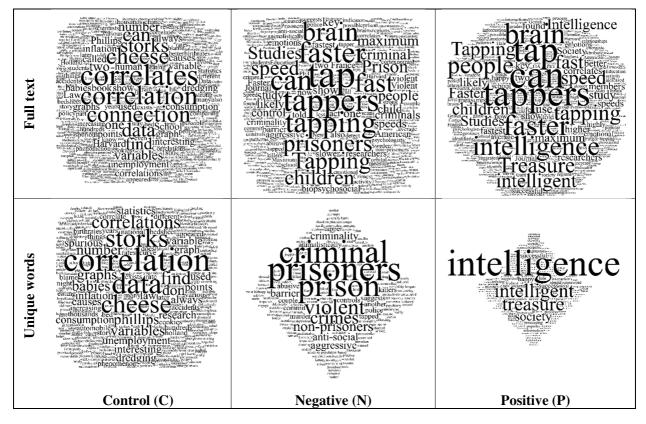


Figure 1: Relative word frequencies of the stimuli. The first row illustrates frequencies based on the entire text (illustrating that the negative and positive texts focused on fast tapping). The second row illustrates the frequencies of the words unique to each text. This shows the differences between the experimental texts, with the negative corpus $(N \cap (C \cup P)')$ focused on words such as 'criminal' and 'prisoners', and the positive corpus $(P \cap (C \cup N)')$ focused on 'intelligence' and 'treasure'.

4. Findings

Two change scores were calculated for each participant. These expressed the change between each participant's pretest and posttest scores on their (a) self-estimated MTS and (b) actual MTS.

Both change scores were calculated as percentages of their respective pretest scores to mitigate potential ceiling and floor effects. The between-group differences in both change scores are shown in



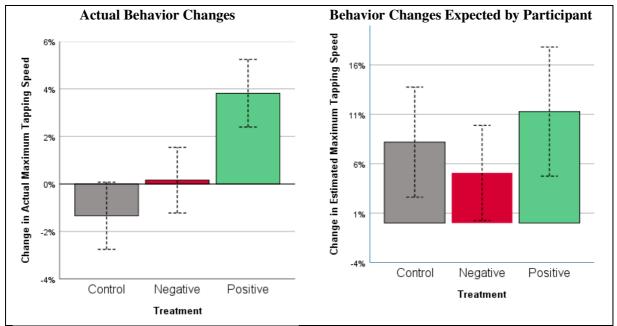


Figure 2: Pretest to posttest change in actual maximum tapping speed (left) and self-estimated maximum tapping speed (right) between treatment groups. Bars represent 95% confidence intervals. Vertical scales differ to underscore between-group differences.

To answer the research questions regarding the capacity of fake news to modify unconscious behavior, two one-way ANOVAs and three correlations were performed. To correct for the inflated type-1 error rate resulting from multiple hypothesis tests, the resulting p-values are adjusted using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995) – these are reported here as 'q-values' and should be interpreted in the same way as p-values. Unadjusted p-values are reported here to facilitate future comparisons across studies using different adjustments but only the q-values are assessed in this article (at a critical value of 0.05).

The null hypothesis H₀ predicted that exposure to fake news will not affect participants' actual MTS. The between-group change in actual MTS was significant, with a treatment effect of medium size according to Cohen's (1988) guidelines, F(2,230) = 14.34, p<0.001, q=0.005, $\eta^2 = 0.11$. Therefore, H₀ was rejected. The change was largest for the positive treatment group, whose participants increased their actual MTS by an average of 3.8% after reading the fake news article. Accounting for the practice effects determined by the control group, the positive treatment group was found to have increased their actual MTS by 5.15% and the negative group increased their actual MTS by 1.50%. Post hoc analyses using Tukey HSD revealed this change in the positive group to be statistically significant compared to the control group (p<0.001). However, the difference between the control and negative group was not statistically significant. Table 1 lists descriptive statistics for the three treatment groups, as well as the change found in

each of the two experimental groups after controlling for practice effects quantified by the control group.

There was not a significant difference between the means of the treatment groups on selfestimated MTS (F(2,230) = 1.181, p=0.309, q=0.386, $\eta^2 = 0.01$). Table 1 shows group means and standard deviations, as well as pairwise comparisons. This is further supported by an investigation of the relationship between the change in actual MTS and self-estimated MTS using a Pearson product-moment correlation coefficient. There was no statistically significant correlation between the two change scores for the control group (r=0.241, p=0.037, q=0.093), the positive group (r=0.081, p=0.471, q=0.471), or the negative group (r=0.215, p=0.063, q=0.105). As such - no evidence was found that participants were conscious of the behavioral effect of fake news.

		Tr	eatment Gro	ups	Experimental Groups after Control	
		Control	Negative	Positive	Negative-Control	Positive-Control**
Actual Change	Mean	-1.34	0.16	3.81	1.50	5.15
	STD	6.15	6.04	6.47	-	-
Expected Change	Mean	8.19	5.07	11.27	-3.12	3.08
	STD	24.18	21.04	29.72	-	-

 Table 1: Pairwise comparisons between treatment groups

**Indicates significant at p<0.001

5. Discussion

This study shows that even a brief (under 5-minute), one-time exposure to a fake news article can modify the unconscious behavior of individuals. The fake news article that associated fast tapping with intelligence increased the speed at which participants tapped, even though all participants were instructed to tap at their maximum speed during both the pretest and the posttest. There was no evidence that participants were aware of their change in behavior. As such, this study demonstrates that fake news can successfully incorporate covert behavioral manipulations.

While participants tapped significantly faster after reading the positive fake news article, no such effect was found between the negative and control groups. This might initially appear to counter findings that negative emotions have stronger effects than positive emotions and non-emotions (Chua & Banerjee, 2018; Porter et al., 2010). However, the negative article may have been less effective for reasons other than valence - for instance due to the implausibility of the premise. Similarly, being of low intelligence may have been perceived as a more pertinent threat to first year undergraduate students than being convicted as a criminal.

These findings should be interpreted as a lower-bound as to the capacities of disinformation to direct behavior. Real-world disinformation campaigns can be expected to have larger effects because of re-enforcement through multiple exposure to the same disinformation, exposure to additional disinformation (Grinberg et al., 2019), peer-endorsement (Messing & Westwood, 2014), and continued influence over time (Levy, 2017). Group narratives, which contribute to individual thinking (Sloman & Fernbach, 2018), can conceivably also re-enforce disinformation Page **12** of **25**

in echo-chambers. The effectiveness of a disinformation stimulus and the susceptibility of social media audiences can also be optimized through psychometric profiling, A/B testing, and micro-targeting (Bay, 2018; Hagar & Diakopoulos, 2019; Lavigne, 2020).

Moreover, the wide reach of online platforms and the capacity for virality in social networks may enable filling the low margins of some societally important decision-making processes. For example, had every voting-eligible citizen been exposed to a disinformation campaign at least as effective as the 5.15% increase that was perceived in the positive group of this experiment (after controlling for practice effects), this would have been sufficient to flip the margin of the popular vote in the last two US presidential elections (2016: 2.09%, 2012: 3.86%). Electoral college arithmetic, a reliance on politically key demographics, and an optimization of content effectiveness and audience susceptibility further compound this possibility.

As expected, no evidence was found that participants were aware of their behavioral change, although both self-estimated changes were investigated as well as correlations between self-estimated change and actual change. Extending to fake news the implicit cognition models by Gawronski, Hoffman and Wilbur (2006) and Wilson (2000), this experiment confirms that the effects of fake news on behavior can be unconscious in that individuals are unaware of the (a) existence of the behavioral manipulation, (b) its source, (c) its content, or (d) its impact on behavior. The lack of evidence confirming the awareness of participants challenges the integrity of democracy, markets, and other systems dependent on the autonomy of distributed agents.

This study also highlights limitations of the contemporary research effort in countering fake news. First, fake news producers can circumvent detection mechanisms, leading to continually reactive, resource-intensive, and one-step-delayed, defensive strategies from practitioners. This in turn implies that individuals will continue to be exposed to fake news. It also vests greater editorial responsibility in the platforms that control public discourse, which are increasingly pressured to moderate, filter, or delete content – raising free speech and censorship concerns. Second, defining the standard of 'truth' by which news is judged to be 'fake' is normatively reductive, as it forces selecting between opposing perspectives that emerge from different schools of epistemology. Third, detection may not always be possible. This is particularly the case where behavioral manipulation is inserted into, and dissimilated amongst, otherwise legitimate online content. This possibility of undetected manipulation is illustrated as the unsafe exposure path in Figure 3.

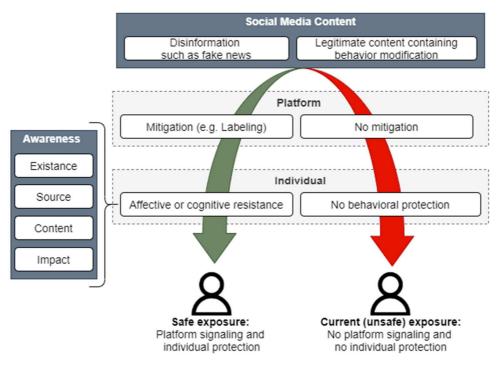


Figure 3: Resistance model to unconscious behavior manipulation showing contributors to unconscious manipulation. Although multiple paths are possible, two paths are highlighted: Safe exposure and current (unsafe) exposure.

The study of fake news is challenged by the elusive and ephemeral nature of its subject. The style and format of fake news can be expected to evolve rapidly in response to detection mechanisms, changes in zeitgeist, and advances in technology. The tendency to elucidate styles that are common to examples of fake news risks overlooking emerging and more discrete forms of fake news. As opposed to defining fake news as a genre, this paper suggests defining it as a medium with socio-technological constraints. Accordingly, the fake news articles in this study were not designed to match a specific style of article, but rather to be maximally effective within the constraints of a fake news article – specifically, (a) being a fabricated news story that (b) is shared on social media and (c) read through a screen (d) in a short amount of time. Future research has the possibility of more formally defining the normative constraints of fake news and other forms of disinformation more formally (such as deep fakes, deceitful tweets, and manipulated images). Future research on covert behavioral manipulation through disinformation should examine the workings and effectiveness of other manipulative techniques within the socio-technological constraints of disinformation (such as different types of emotion, covert messaging, social re-enforcement, and repetition). A research agenda in covert behavior and disinformation should explore the question: To what extent can manipulative techniques be effectively incorporated into the medium of disinformation, and how can we protect against this threat?

6. Limitations

Results detected in this undergraduate sample might not be consistent across the general population or other subgroups. Reading a treatment text in an academic lab may produce different results than reading a fake news article in a more intimate setting. The use of a

university-provided laptop may produce different effects than a familiar smartphone or personal computer. Trust, echo chambers, filter bubbles, additional exposure, may affect the results. The stimuli in this experiment were deliberately non-political to protect against previous exposure. Political disinformation can be expected to be influenced by other cognitive processes such as memory recall, backfire effects, and confirmation bias. Self-selecting content, such as by reading a headline before deciding to click on a hyperlink to a fake news article, or exposure to headlines produced by newsfeed algorithms, may also have effects.

The effects identified in this experiment may have been stronger had the stimuli been calibrated to a peer-group. The keywords selected were largely scored by US-based Amazon Mechanical Turk users in 2012 (Warriner et al., 2013), whose valence and arousal ratings may have differed from this study's younger participants in a French university in 2019. In addition, the valence of the positive stimulus was not the precise inverse of the valence of the negative stimulus. As a result, the findings should not be interpreted as differential effects of variance polarity. Instead, the findings should be interpreted as the extent to which fake news enables specific psychological interventions (unconscious cognitive and affective influence). Finally, this study tested the immediate effects of one-off, short exposure to manipulative disinformation on behavior and self-awareness. Future longitudinal research can establish how this effect is sustained over time as well as the effects of repeated or otherwise reinforced exposure.

7. Conclusion

Despite the enormous threats and high risks of behavior modification through fake news, there is a paucity of controlled studies on the direct effects of fake news on behavior. This study provides new evidence on the ability of disinformation to change unconscious behavior. Its finding that short exposure to disinformation (as is typical online) can have moderate effects on unconscious individual behavior, raises immediate concerns for platforms, policymakers, and social media users. These findings should be interpreted within the setting of a lab and single exposure. Realworld effects that assume psychometrically optimized content, micro-targeted audiences, and repeated exposure can be expected to be larger. Narrow margins of decision-making processes such as elections imply that small treatment effects can have large societal implications. This is also true for exponential risks, as is evidenced by the large global effects of a few early infections during the COVID-19 pandemic.

These findings raise deep concerns for the future of society and politics. Disinformation risks skewing individuals' worldviews and deleteriously informing their behavior. Deliberately produced and targeted disinformation aimed at behavior modification amplifies these risks, by introducing incentives and optimization. The democratization of AI and the means of producing disinformation further amplifies these risks beyond platforms to the entire eco-system of social media users and content. The behavioral effects of disinformation are not obvious through current means of content analysis or fact-checking, and such disinformation is consequently difficult to detect. While the behavioral effects of disinformation can be empirically observed, refined, and predicted by malicious content creators, the processes of manipulation and their complete consequences may not necessarily be fully understood by the creators of the disinformation themselves. The continual efforts by malicious actors to bypass disinformation

detection mechanisms implies that we must come to terms with the existence of disinformation on social networks and be open-minded about the forms and effects of disinformation.

Individual autonomy is assumed both as a normative principal of liberal democracies and functional prerequisite of collective decision-making systems (such as elections and referenda). This study's findings and their surrounding context suggest that there is an urgent need to (1) further investigate the mechanisms through which disinformation can manipulate unconscious behavior, (2) determine the real-world size and duration of these effects, (3) find methods to detect and prevent these risks, and (4) build contingency-plans or stronger systems for distributed decision-making and democracy. As such, the era of disinformation has highlighted renewed risks in distributed systems of governance. It has also created an urgent need and opportunity for interdisciplinary collaboration (involving at least political scientists, media scholars, and platforms) to protect individual autonomy and collective self-determination. These efforts should no longer be limited to fake news, its spread and its detection, but need to consider that individuals are likely to be exposed to increasingly undetectable behavior modification from a variety of online content, and so need to be protected.

References

- Ajzen, I., & Fishbein, M. (2005). The influence of attitudes on behavior. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *The Handbook of Attitudes* (Vol. 173, p. 31). Lawrence Erlbaum Associates Publishers.
- Antoniades, C. A., Ober, J., Hicks, S., Siuda, G., Carpenter, R., Kennard, C., & Nemeth, A. H. (2012). Statistical characteristics of finger-tapping data in Huntington's disease. *Medical & Biological Engineering & Computing*, 50(4), 341–346.
- Armijo-Olivo, S., Warren, S., & Magee, D. (2009). Intention to treat analysis, compliance, drop-outs and how to deal with missing data in clinical research: A review. *Physical Therapy Reviews*, 14(1), 36–49. https://doi.org/10.1179/174328809X405928
- Arnold, G., Ph.D, K. B. B., Lu, P., Dean, A., Wen, J., Nitch, S., & McPherson, S. (2005). Sensitivity and Specificity of Finger Tapping Test Scores for the Detection of Suspect Effort. *The Clinical Neuropsychologist*, 19(1), 105–120. https://doi.org/10.1080/13854040490888567
- Austin, D., Petersen, J., Jimison, H., & Pavel, M. (2012). A state-space model for finger tapping with applications to cognitive inference. 21–24.
- Back, K. W. (1961). Decisions Under Uncertainty Rational, Irrational, and Non-rational. *American Behavioral Scientist*, 4(6), 14–19. https://doi.org/10.1177/000276426100400604
- Bakir, V., & McStay, A. (2018). Fake News and The Economy of Emotions. *Digital Journalism*, 6(2), 154–175. https://doi.org/10.1080/21670811.2017.1345645
- Balmas, M. (2014). When fake news becomes real: Combined exposure to multiple news sources and political attitudes of inefficacy, alienation, and cynicism. *Communication Research*, *41*(3), 430–454.
- Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype-activation on action. *Journal of Personality and Social Psychology*.
- Barut, Ç., Kızıltan, E., Gelir, E., & Köktürk, F. (2013). Advanced Analysis of Finger-Tapping Performance: A Preliminary Study. *Balkan Medical Journal*, 30(2), 167–171. https://doi.org/10.5152/balkanmedj.2012.105
- Bay, M. (2018). Social media ethics: A Rawlsian approach to hypertargeting and psychometrics in political and commercial campaigns. *ACM Transactions on Social Computing*, 1(4), 1–14.
- Bierhoff, H. W. (2002). Prosocial behaviour. Psychology Press.

- Bode, L., & Vraga, E. K. (2018). See something, say something: Correction of global health misinformation on social media. *Health Communication*, *33*(9), 1131–1140.
- Boehmer, J., & Tandoc, E. C. (2015). Why we retweet: Factors influencing intentions to share sport news on Twitter. *International Journal of Sport Communication*, 8(2), 212–232.
- Bradshaw, S., & Howard, P. N. (2018). Challenging truth and trust: A global inventory of organized social media manipulation. *The Computational Propaganda Project*.
- Bucher, T. (2017). The algorithmic imaginary: Exploring the ordinary affects of Facebook algorithms. *Information, Communication & Society*, 20(1), 30–44. https://doi.org/10.1080/1369118X.2016.1154086
- Burnap, P., Williams, M. L., Sloan, L., Rana, O., Housley, W., Edwards, A., Knight, V., Procter, R., & Voss, A. (2014). Tweeting the terror: Modelling the social media reaction to the Woolwich terrorist attack. *Social Network Analysis and Mining*, 4(1), 206. https://doi.org/10.1007/s13278-014-0206-4
- Caldwell-Harris, C. L., & Ayçiçeği-Dinn, A. (2009). Emotion and lying in a non-native language. *International Journal of Psychophysiology*, 71(3), 193–204. https://doi.org/10.1016/j.ijpsycho.2008.09.006
- Carroll, C. A., O'donnell, B. F., Shekhar, A., & Hetrick, W. P. (2009). Timing dysfunctions in schizophrenia as measured by a repetitive finger tapping task. *Brain and Cognition*, 71(3), 345– 353.
- Chevance, G., Caudroit, J., Henry, T., Guerin, P., Boiché, J., & Héraud, N. (2018). Do implicit attitudes toward physical activity and sedentary behavior prospectively predict objective physical activity among persons with obesity? *Journal of Behavioral Medicine*, *41*(1), 31–42. https://doi.org/10.1007/s10865-017-9881-8
- Chevance, G., Héraud, N., Varray, A., & Boiché, J. (2017). Change in explicit and implicit motivation toward physical activity and sedentary behavior in pulmonary rehabilitation and associations with postrehabilitation behaviors. *Rehabilitation Psychology*, 62(2), 119–129. https://doi.org/10.1037/rep0000137
- Chua, A. Y., & Banerjee, S. (2018). Intentions to trust and share online health rumors: An experiment with medical professionals. *Computers in Human Behavior*, 87, 1–9.
- Clarke, S., Hoggett, P., & Thompson, S. (2006). Emotion, politics and society. Springer.
- Colliander, J. (2019). "This is fake news": Investigating the role of conformity to other users' views when commenting on and spreading disinformation in social media. *Computers in Human Behavior*, 97, 202–215.
- Conroy, D. E., Hyde, A. L., Doerksen, S. E., & Ribeiro, N. F. (2010). Implicit Attitudes and Explicit Motivation Prospectively Predict Physical Activity. *Annals of Behavioral Medicine*, 39(2), 112– 118. https://doi.org/10.1007/s12160-010-9161-0
- Conroy, N. J., Rubin, V. L., & Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. *Proceedings of the Association for Information Science and Technology*, 52(1), 1–4.
- Cousins, M. S., Corrow, C., Finn, M., & Salamone, J. D. (1998). Temporal Measures of Human Finger Tapping: Effects of Age. *Pharmacology Biochemistry and Behavior*, 59(2), 445–449. https://doi.org/10.1016/S0091-3057(97)00443-7
- Dasgupta, N., & Rivera, L. M. (2008). When social context matters: The influence of long-term contact and short-term exposure to admired outgroup members on implicit attitudes and behavioral intentions. *Social Cognition*, 26(1), 112–123.
- Desai, S. C., & Reimers, S. (2018). Some misinformation is more easily countered: An experiment on the continued influence effect. *CogSci*.
- Dolan, R. J. (2002). Emotion, cognition, and behavior. Science, 298(5596), 1191-1194.
- Elson, S. E. (1975). Cognitive Behavior Modification: An Evaluation of the Effects of Language Techniques Upon Emotional Arousal. Michigan State University.
- Ewert, A., & Sibthorp, J. (2009). Creating Outcomes through Experiential Education: The Challenge of Confounding Variables. *Journal of Experiential Education*, 31(3), 376–389. https://doi.org/10.1177/105382590803100305

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. *Advances in Experimental Social Psychology*, 23(75–109), 60318–4.

Fazio, R. H., & Olson, M. A. (2014). The mode model. *Dual-Process Theories of the Social Mind*, 155. Featherstone, J. D., Davis, C. A., & Zhang, J. (2019). Correcting Vaccine Misinformation on Social

Media Using Fact-checking Labels. *APHA's 2019 Annual Meeting and Expo (Nov. 2-Nov. 6)*. Fishbein, M., & Ajzen, I. (2011). *Predicting and changing behavior: The reasoned action approach*.

- Taylor & Francis.
- Fisher, R. J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of Consumer Research*, 20(2), 303–315.
- Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*.
- Forrest, L. N., Smith, A. R., Fussner, L. M., Dodd, D. R., & Clerkin, E. M. (2016). Using implicit attitudes of exercise importance to predict explicit exercise dependence symptoms and exercise behaviors. *Psychology of Sport and Exercise*, 22, 91–97. https://doi.org/10.1016/j.psychsport.2015.06.006
- Fung, B. (2020, May 27). Trump threatens to crack down on social media platforms after Twitter labels his tweets. CNN. https://www.cnn.com/2020/05/27/politics/donald-trump-twitter-threatregulate/index.html
- Gao, M., Xiao, Z., Karahalios, K., & Fu, W.-T. (2018). To label or not to label: The effect of stance and credibility labels on readers' selection and perception of news articles. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW), 1–16.
- Garrett, R. K. (2016). *Facebook's Problem Is More Complicated Than Fake News*. Scientific American. https://www.scientificamerican.com/article/facebook-s-problem-is-more-complicated-than-fakenews/
- Gawronski, B., Hofmann, W., & Wilbur, C. J. (2006). Are "implicit" attitudes unconscious? *Consciousness and Cognition*, 15(3), 485–499.
- Ghanem, B., Rosso, P., & Rangel, F. (2020). An Emotional Analysis of False Information in Social Media and News Articles. ACM Transactions on Internet Technology, 20(2), 19:1–19:18. https://doi.org/10.1145/3381750
- Gidley Larson, J. C., & Suchy, Y. (2015). The contribution of verbalization to action. *Psychological Research*, 79(4), 590–608. https://doi.org/10.1007/s00426-014-0586-0
- Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, 7(1), 2053951719897945.
- Grecucci, A., Koch, I., & Rumiati, R. I. (2011). The role of emotional context in facilitating imitative actions. *Acta Psychologica*, *138*(2), 311–315.
- Green, D., & Shapiro, I. (1996). Pathologies of Rational Choice Theory. Yale University Press.
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*(1), 4.
- Greenwald, A. G., & Banaji, M. R. (2017). The implicit revolution: Reconceiving the relation between conscious and unconscious. *American Psychologist*, 72(9), 861.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464.
- Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*(1), 17.
- Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on Twitter during the 2016 US presidential election. *Science*, *363*(6425), 374–378.
- Gross, M. (2017). The dangers of a post-truth world. *Current Biology*, 27(1), R1–R4. https://doi.org/10.1016/j.cub.2016.12.034

- Gruszczynski, M. W., Balzer, A., Jacobs, C. M., Smith, K. B., & Hibbing, J. R. (2013). The physiology of political participation. *Political Behavior*, *35*(1), 135–152.
- Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, *5*(1), eaau4586.
- Guess, A., Nyhan, B., & Reifler, J. (2018). Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 US presidential campaign. *European Research Council*, 9(3), 4.
- Hagar, N., & Diakopoulos, N. (2019). Optimizing content with A/B headline testing: Changing newsroom practices. *Media and Communication*, 7(1), 117–127.
- Hajek, P., Barushka, A., & Munk, M. (2020). Fake consumer review detection using deep neural networks integrating word embeddings and emotion mining. *Neural Computing and Applications*. https://doi.org/10.1007/s00521-020-04757-2
- Halstead, W. C. (1947). *Brain and intelligence: A quantitative study of the frontal lobes.* University of Chicago Press.
- Heiss, R., Schmuck, D., & Matthes, J. (2019). What drives interaction in political actors' Facebook posts? Profile and content predictors of user engagement and political actors' reactions. *Information, Communication & Society*, 22(10), 1497–1513. https://doi.org/10.1080/1369118X.2018.1445273
- Hilbert, M., Ahmed, S., Cho, J., Liu, B., & Luu, J. (2018). Communicating with Algorithms: A Transfer Entropy Analysis of Emotions-based Escapes from Online Echo Chambers. *Communication Methods and Measures*, 12(4), 260–275. https://doi.org/10.1080/19312458.2018.1479843
- Hofmann, W., & Friese, M. (2008). Impulses got the better of me: Alcohol moderates the influence of implicit attitudes toward food cues on eating behavior. *Journal of Abnormal Psychology*, 117(2), 420–427. https://doi.org/10.1037/0021-843X.117.2.420
- Hollands, G. J., Prestwich, A., & Marteau, T. M. (2011). Using aversive images to enhance healthy food choices and implicit attitudes: An experimental test of evaluative conditioning. *Health Psychology*, 30(2), 195–203. https://doi.org/10.1037/a0022261
- Houben, K., Havermans, R. C., & Wiers, R. W. (2010). Learning to dislike alcohol: Conditioning negative implicit attitudes toward alcohol and its effect on drinking behavior. *Psychopharmacology*, 211(1), 79–86. https://doi.org/10.1007/s00213-010-1872-1
- Howard, G. S., & Dailey, P. R. (1979). Response-shift bias: A source of contamination of self-report measures. *Journal of Applied Psychology*, 64(2), 144.
- Ioannidis, J. P. A. (2020). Coronavirus disease 2019: The harms of exaggerated information and nonevidence-based measures. *European Journal of Clinical Investigation*, 50(4), e13222.
- Isaak, J., & Hanna, M. J. (2018). User data privacy: Facebook, Cambridge Analytica, and privacy protection. *Computer*, *51*(8), 56–59.
- Jäncke, L., Loose, R., Lutz, K., Specht, K., & Shah, N. (2000). Cortical activations during paced fingertapping applying visual and auditory pacing stimuli. *Cognitive Brain Research*, 10(1–2), 51–66.
- Jobbágy, Á., Harcos, P., Karoly, R., & Fazekas, G. (2005). Analysis of finger-tapping movement. *Journal* of Neuroscience Methods, 141(1), 29–39.
- Johnson, H. M., & Seifert, C. M. (1994). Sources of the continued influence effect: When misinformation in memory affects later inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(6), 1420.
- Johnson, J. (2018). The Self-Radicalization of White Men: "Fake News" and the Affective Networking of Paranoia. *Communication, Culture and Critique*, 11(1), 100–115. https://doi.org/10.1093/ccc/tcx014
- Klar, R. (2020, September 25). *Disinformation, QAnon efforts targeting Latino voters ramp up ahead of presidential election* [Text]. TheHill. https://thehill.com/policy/technology/518118-disinformation-qanon-efforts-targeting-latino-voters-ramp-up-ahead-of
- Klauer, K. C., & Musch, J. (2003). Affective priming: Findings and theories. *The Psychology of Evaluation: Affective Processes in Cognition and Emotion*, 7, 49.

- Kouzy, R., Abi Jaoude, J., Kraitem, A., El Alam, M. B., Karam, B., Adib, E., Zarka, J., Traboulsi, C., Akl, E. W., & Baddour, K. (2020). Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter. *Cureus*, 12(3).
- Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24), 8788–8790.
- Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, 84(3), 394–421.
- Kunst-Wilson, W. R., & Zajonc, R. B. (1980). Affective discrimination of stimuli that cannot be recognized. *Science*, 207(4430), 557–558.
- Lavigne, M. (2020). Strengthening ties: The influence of microtargeting on partisan attitudes and the vote. *Party Politics*, 1354068820918387. https://doi.org/10.1177/1354068820918387
- Lee, H. S., Addicott, M., Martin, L. E., Harris, K. J., Goggin, K., Richter, K. P., Patten, C. A., McClernon, F. J., Fleming, K., & Catley, D. (2017). Implicit Attitudes and Smoking Behavior in a Smoking Cessation Induction Trial. *Nicotine & Tobacco Research*, 20(1), 58–66. https://doi.org/10.1093/ntr/ntw259
- Levy, N. (2017). The bad news about fake news. *Social Epistemology Review and Reply Collective*, 6(8), 20–36.
- Li, W., Zinbarg, R. E., Boehm, S. G., & Paller, K. A. (2007). Neural and Behavioral Evidence for Affective Priming from Unconsciously Perceived Emotional Facial Expressions and the Influence of Trait Anxiety. *Journal of Cognitive Neuroscience*, 20(1), 95–107. https://doi.org/10.1162/jocn.2008.20006
- Li, Y., Feng, X., & Zhang, S. (2016). Detecting fake reviews utilizing semantic and emotion model. 317–320.
- Liu, H., Wang, L., Han, X., Zhang, W., & He, X. (2020). Detecting Fake News on Social Media: A Multi-Source Scoring Framework. 2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), 524–531. https://doi.org/10.1109/ICCCBDA49378.2020.9095586
- Liv, N., & Greenbaum, D. (2020). Deep Fakes and Memory Malleability: False Memories in the Service of Fake News. AJOB Neuroscience, 11(2), 96–104. https://doi.org/10.1080/21507740.2020.1740351
- Lodge, M., & Taber, C. S. (2013). *The Rationalizing Voter*. Cambridge University Press. https://doi.org/10.1017/CBO9781139032490
- Loehr, J. D., & Palmer, C. (2007). Cognitive and biomechanical influences in pianists' finger tapping. *Experimental Brain Research*, 178(4), 518–528. https://doi.org/10.1007/s00221-006-0760-8
- Maison, D., Greenwald, A. G., & Bruin, R. H. (2004). Predictive Validity of the Implicit Association Test in Studies of Brands, Consumer Attitudes, and Behavior. *Journal of Consumer Psychology*, 14(4), 405–415. https://doi.org/10.1207/s15327663jcp1404_9
- Marcus, G. E., Neuman, W. R., & MacKuen, M. (2000). *Affective intelligence and political judgment*. University of Chicago Press.
- Mattavelli, S., Avishai, A., Perugini, M., Richetin, J., & Sheeran, P. (2017). How Can Implicit and Explicit Attitudes Both Be Changed? Testing Two Interventions to Promote Consumption of Green Vegetables. Annals of Behavioral Medicine, 51(4), 511–518. https://doi.org/10.1007/s12160-016-9874-9
- Mattiuzzi, C., & Lippi, G. (2020). Which lessons shall we learn from the 2019 novel coronavirus outbreak? *Annals of Translational Medicine*, 8(3), 48.
- Meichenbaum, D. (1977). Cognitive-Behavior Modification: An Integrative Approach. Springer Science & Business Media.
- Messing, S., & Westwood, S. J. (2014). Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online. *Communication Research*, 41(8), 1042–1063.

- Mueller, R. S. (2019). *Report on the investigation into Russian interference in the 2016 presidential election* (Vol. 1). US Department of Justice Washington, DC.
- Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, 29(5), 665–675.
- O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Broadway Books.
- Ott, M., Cardie, C., & Hancock, J. T. (2013). Negative deceptive opinion spam. *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 497–501.
- Pal, A., & Banerjee, S. (2019). Understanding Online Falsehood From the Perspective of Social Problem. Handbook of Research on Deception, Fake News, and Misinformation Online, 1.
- Pasquale, F. (2015). The black box society. Harvard University Press.
- Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, 89(3), 277.
- Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management*, 12(4), 531–544.
- Porter, S., Bellhouse, S., McDougall, A., Ten Brinke, L., & Wilson, K. (2010). A prospective investigation of the vulnerability of memory for positive and negative emotional scenes to the misinformation effect. *Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences Du Comportement*, 42(1), 55.
- Porter, S., Taylor, K., & ten Brinke, L. (2008). Memory for media: Investigation of false memories for negatively and positively charged public events. *Memory*, 16(6), 658–666.
- Rabinowitz, I., & Lavner, Y. (2014). Association between finger tapping, attention, memory, and cognitive diagnosis in elderly patients. *Perceptual and Motor Skills*, *119*(1), 259–278.
- Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., Crandall, J. W., Christakis, N. A., Couzin, I. D., & Jackson, M. O. (2019). Machine behaviour. *Nature*, 568(7753), 477–486.
- Rich, P. R., & Zaragoza, M. S. (2016). The continued influence of implied and explicitly stated misinformation in news reports. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(1), 62.
- Settle, J. E., Hibbing, M. V., Anspach, N. M., Carlson, T. N., Coe, C. M., Hernandez, E., Peterson, J., Stuart, J., & Arceneaux, K. (2020). Political psychophysiology: A primer for interested researchers and consumers. *Politics and the Life Sciences*, 39(1), 101–117.
- Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations*, *19*(1), 22–36.
- Sloman, S., & Fernbach, P. (2018). The knowledge illusion: Why we never think alone. Penguin.
- Strauss, E., Sherman, E. M., & Spreen, O. (2006). A compendium of neuropsychological tests: Administration, norms, and commentary. American Chemical Society.
- Strohminger, N., Caldwell, B., Cameron, D., Borg, J. S., & Sinnott-Armstrong, W. (2014). Implicit morality: A methodological survey. In *Experimental Ethics* (pp. 133–156). Springer.
- Tandoc, E. C., Lim, Z. W., & Ling, R. (2018). Defining "fake news" A typology of scholarly definitions. Digital Journalism, 6(2), 137–153.
- Taylor, D. J., & McFatter, R. M. (2003). Cognitive performance after sleep deprivation: Does personality make a difference? *Personality and Individual Differences*, *34*(7), 1179–1193.
- Thorson, E. (2016). Belief echoes: The persistent effects of corrected misinformation. *Political Communication*, *33*(3), 460–480.
- Tomassini, A., Vercillo, T., Torricelli, F., & Morrone, M. C. (2018). Rhythmic motor behaviour influences perception of visual time. *Proceedings of the Royal Society B*, 285(1888), 20181597.
- Twenge, J. M., Baumeister, R. F., DeWall, C. N., Ciarocco, N. J., & Bartels, J. M. (2007). Social exclusion decreases prosocial behavior. *Journal of Personality and Social Psychology*, 92(1), 56.
- Vigen, T. (2015). Spurious Correlations. Hachette Books.

- Vraga, E. K., & Bode, L. (2017). Using expert sources to correct health misinformation in social media. *Science Communication*, *39*(5), 621–645.
- Wagner, M. W., Deppe, K. D., Jacobs, C. M., Friesen, A., Smith, K. B., & Hibbing, J. R. (2015). Beyond survey self-reports: Using physiology to tap political orientations. *International Journal of Public Opinion Research*, 27(3), 303–317.

Wahl-Jorgensen, K. (2019). Emotions, Media and Politics. Wiley.

- Wang, Y., McKee, M., Torbica, A., & Stuckler, D. (2019). Systematic Literature Review on the Spread of Health-related Misinformation on Social Media. *Social Science & Medicine*, 240, 112552.
- Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, 45(4), 1191–1207.
- Weeks, B. E. (2015). Emotions, Partisanship, and Misperceptions: How Anger and Anxiety Moderate the Effect of Partisan Bias on Susceptibility to Political Misinformation. *Journal of Communication*, 65(4), 699–719. https://doi.org/10.1111/jcom.12164
- Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, 107(1), 101.
- Witt, S. T., Meyerand, M. E., & Laird, A. R. (2008). Functional neuroimaging correlates of finger tapping task variations: An ALE meta-analysis. *NeuroImage*, *42*(1), 343–356. https://doi.org/10.1016/j.neuroimage.2008.04.025
- Wollebæk, D., Karlsen, R., Steen-Johnsen, K., & Enjolras, B. (2019). Anger, Fear, and Echo Chambers: The Emotional Basis for Online Behavior: *Social Media + Society*. https://doi.org/10.1177/2056305119829859
- Yeung, K. (2017). 'Hypernudge': Big Data as a mode of regulation by design. *Information, Communication & Society*, 20(1), 118–136.
- Zittrain, J. L. (2014). *Engineering an Election* (SSRN Scholarly Paper ID 2457502). Social Science Research Network. https://papers.ssrn.com/abstract=2457502

Appendix

Positive Fake News Stimulus

This article is from a news website called News Today. As of today, it has been shared 67,032 times over Facebook.

In 2006, the New York Times printed a crossword puzzle. Crosswords are a common game in newspapers: a grid of squares which we must fill by guessing words according to clues. Except this was no ordinary crossword. It did not use words, but numbers. The clues were not sentences, but formulas. Solving it meant solving a 30-year old mathematical problem – Berloni's Theorem. The game was meant to uncover geniuses in a way that was cheaper, faster, and simpler than a standard IQ test. The Times excitedly reported that 300 winners were identified. They asked a follow-up question: how else can we quickly identify geniuses? Scientists now have an answer, and it involves the speed at which you can tap.

In the 1972 book 'Neuro-Social Theory of Intelligence', Chan and Gaston suggested that there are key differences in the neurological reaction times of highly-intelligent people. Now, after testing 320 measures, scientists have identified one attribute that significantly predicts intelligence: maximum finger-tapping speed.

The Finger-tapping Intelligence Attribute (TIA) was confirmed in a 2010 study by two Harvard psychologists, Stoneman and Gerick. The study tested 312 members of the Treasure Society – an elite society where membership requires scoring in the top 0.1% of IQ scores. They gave computers to the Treasure Society members and told them to tap the "T" key as fast as possible. The study found with high statistical significance that highly intelligent people could tap faster than the average American. This is the first study to confirm the initial finding: the faster you can tap, the more intelligent you are likely to be.

In 2013, two Stanford researchers, Genun and Honette, replicated the Harvard study. Instead of comparing Treasure Society members to average Americans, they compared Treasure Society members to each other. Their conclusions were published in the Journal of Applied Intelligence. What they found was that, even within this group of highly intelligent individuals, those with the highest IQ could tap the fastest. In addition, they found that the fastest tappers scored better in other measures: 'having positive relationships', 'being happy', 'having a purpose in life'. These two studies are now called the Treasure Tapping Studies.

The Treasure Tapping Studies have been reproduced in Sweden, Germany, and France, with the same results. Maximum tapping speed is a strong indicator of intelligence, irrespective of culture. It has been known for some time that speed of nerve pathways in the brain correlates with intelligence level. This suggested that the parts of the brain responsible for intelligence and for tapping are connected.

A breakthrough came from France in 2017. On a cold spring day, a group of students aged 17 to 21 walked into the American Hospital of Paris. One after the other, they were put into a PET scanner, a machine that measures brain activity. The most active regions of the brain "light up" while non-active regions remain dark. Each student was given a Bluetooth keyboard. Then, as with the Treasure Tapping Studies, each student was told to tap the "T" key as fast as they can. The scans showed that fast-tappers had increased activity in the orbital cortex, the area of the brain that regulates emotions. French neurologists noticed that, as shown below, this part of the brain lights up for fast tappers, but not for slow tappers.

We now understand why some people can tap faster than others. Faster tappers use the full processing power of their orbital cortex – enabling them to (1) process information and intentions faster and (2) better regulate their emotions. Since faster tappers have more power to process information, they can act faster and score higher on IQ tests. They display what psychologists call "bonus intelligence winnings": Faster tappers can use their superior IQ to make better use of opportunities (financial, educational), and be more successful in life. Faster tappers can use their increased emotional intelligence to build healthier relationships and be perceived by their peers as 'joyous', 'happy', 'energetic', and 'funny'. The faster you can tap, the more likely it is that you have the same "higher functioning brain" patterns.

The human brain is highly mutable. Education can socialize us and increase our maximum speeds on tapping tests. In 2016, MIT researchers filmed 9 and 10-year-old children in school over a period of two-weeks micro-monitoring their social interactions. After two weeks the researchers took the children to the school lab and told them to tap the "T" key as fast as possible. Psychologists who reviewed the films found that 7 out of the 90 children had engaged in highly-socialized play. The most popular children were the fastest tappers. The fastest tappers had more friends – they were better at understanding emotional signals from others, and more children played with them. Their interactions were more joyful – they caused laughter, energetic play and excited reactions. They were also more academically successful – they received higher grades than the slower tappers. The 8 children with the highest grades could tap 87% faster than the average child – almost twice as fast. This tells us that we can detect desirable traits and future success by measuring maximum tapping speeds, even in children whose brains are still developing.

Of course, tapping speed is only one predictor of high intelligence and emotional ability. But it's a proven and strong indicator. What we know is that people who can tap faster are more likely to be intelligent, happy, balanced, healthy, entertaining, and ultimately, to reach success in life. If we measure maximum tapping speed early, we have a chance to identify those who can most improve society.

Negative Fake News Stimulus

This article is from a news website called News Today. As of today, it has been shared 67,032 times over Facebook.

In 2006, a New York couple was arrested by police officers in connection with 5 murders. The couple had kidnapped 5 joggers at gunpoint. They then held them in a basement that police described as a "hell hole". Finally, they brutally killed them with a machete. Television reports emphasized that the killers were 'normal people'. This raised fear among New Yorkers. Scientists have since tried to explain and predict what makes these 'normal people' become vicious killers. Do criminals act differently? Chan and Gaston's 1972 'Neuro-Social Theory of Deviance' suggests that there are key differences in the neurological reaction times of criminals and non-criminals. But, until now this had never been tested. After testing 320 measures, scientists identified one attribute that confidently predicts criminality: maximum finger-tapping speed. This has implications on many aspects of humanity.

In May 2010, two Harvard psychologists, Stoneman and Gerick, entered Brookline Maximum Security Prison. Their plan? To investigate the tapping speed of prisoners. They gave computers to the prisoners and told them to tap the "T" key as fast as possible. The study found with high statistical significance that prisoners could tap faster than the average American. This is the first study to confirm the initial finding: the faster you can tap, the more likely you are to be a criminal.

In 2013, two Stanford researchers, Genun and Honette, replicated the Harvard study – with a twist. They did not compare prisoners to non-prisoners. They compared prisoners to other prisoners. Their 3 conclusions were published in the Journal of Violent Crimes. First, prisoners with longer prison sentences tapped faster. Second, violent criminals tapped faster than non-violent criminals by 62%. Third, the fastest tappers were charged with the most horrific crimes: 'massacre or terrorism', 'kidnapping', 'abuse', 'death by asphyxiation'. These two studies are now called the Prison Tapping Studies.

The Prison Tapping Studies were reproduced in Sweden, Germany, and France. The results were the same. Maximum tapping speed is a strong indicator of criminality, irrespective of culture. This suggests that the parts of the brain responsible for criminality and tapping are linked.

A breakthrough came from France 2017. On a cold spring day, a group of students aged 17 to 21 walked into the American Hospital of Paris. One after the other, they were put into PET scanners, a machine to measure brain activity. The most active regions of the brain "light up" while non-active regions remain dark. Then, like the Prison Tapping Studies, the subjects were told to tap the "T" key as fast as they can. The scans showed that fast-tappers had reduced activity in the orbital cortex, the area of the brain that regulates emotions. Notice that in the below the part of the brain that controls emotions lights up for slower tappers, but not for fast tappers.

We now understand why some people can tap faster than others. The speed of a slower tapper is slowed by emotional control. The activation of the orbital cortex acts as a barrier – a barrier that also controls our impulses and morality. Fast tappers do not have this barrier. A fast tapper is on automatic control. They have less emotional control. They act more instinctively and animalistically. They display what psychologists call "impulsive anti-sociality". The faster you can tap, the more likely you have the same brain patterns as what neurologists call the "criminal brain".

The human brain is highly mutable. Education can socialize us and lower our maximum speeds on tapping tests. In 2016, MIT researchers filmed 4 and 5-year-old children in school. All children were given a keyboard and told to tap the "T" key as fast as possible. Psychologists who reviewed the films found that 7 out of the 90 children engaged in overly aggressive play. The most aggressive children were the fastest tappers. The fastest tappers caused pain – one child scratched another's face, drawing blood. They caused potential bodily harm – instead of hugging, one child wrestled and choked another child.

They were also generally abusive – hitting and throwing toys at other children. The 8 most aggressive children could tap 87% faster than the average child – almost twice as fast. This tells us that we can detect undesirable traits by measuring maximum tapping speeds, even in children whose brains are still developing.

Of course, tapping speed is only one predictor of anti-social tendencies. But it's a strong indicator. What we know is that people who can tap faster are more likely to be violent, uncontrolled, anti-social and ultimately, to commit murder. If we measure maximum tapping speed early, we have a chance to detect criminal instincts and avoid crisis.