Bias-Aware Systems: Exploring Indicators for the Occurrences of Cognitive Biases when Facing Different Opinions

Nattapat Boonprakong
nboonprakong@student.unimelb.edu.au
The University of Melbourne
Melbourne, Victoria, Australia

Xiuge Chen
xiugec@student.unimelb.edu.au
The University of Melbourne
Melbourne, Victoria, Australia

Catherine E. Davey
catherine.davey@unimelb.edu.au
The University of Melbourne
Melbourne, Victoria, Australia

Benjamin Tag
benjamin.tag@unimelb.edu.au
The University of Melbourne
Melbourne, Victoria, Australia

Tilman Dingler
tilman.dingler@unimelb.edu.au
The University of Melbourne
Melbourne, Victoria, Australia

ABSTRACT

Cognitive biases have been shown to play a critical role in creating echo chambers and spreading misinformation. They undermine our ability to evaluate information and can influence our behaviour without our awareness. To allow the study of occurrences and effects of biases on information consumption behaviour, we explore indicators for cognitive biases in physiological and interaction data. Therefore, we conducted two experiments investigating how people experience statements that are congruent or divergent from their own ideological stance. We collected interaction data, eye tracking data, hemodynamic responses, and electrodermal activity while participants were exposed to ideologically tainted statements. Our results indicate that people spend more time processing statements that are incongruent with their own opinion. We detected differences in blood oxygenation levels between congruent and divergent opinions, a first step towards building systems to detect and quantify cognitive biases.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; *Empirical studies in HCI*; *Ubiquitous and mobile computing systems and tools.*

KEYWORDS

Bias-aware systems, Cognitive biases, Cognition-aware systems, fNIRS, Eye tracking, Electrodermal activity

ACM Reference Format:
Nattapat Boonprakong, Xiuge Chen, Catherine E. Davey, Benjamin Tag, and Tilman Dingler. 2023. Bias-Aware Systems: Exploring Indicators for the Occurrences of Cognitive Biases when Facing Different Opinions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3544548.3580917

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04…$15.00
https://doi.org/10.1145/3544548.3580917

1 INTRODUCTION

Algorithms increasingly curate the information we encounter online. In an attempt to grab and keep users' attention, they filter and provide content based on prior browsing history and inferred interests [9, 46]. Consequently, most information provided to users feeds into their existing beliefs and opinions. In recent years, this mechanism has triggered a heated discussion about how the prioritisation of user engagement plays into the spread of misinformation and political extremism [41]. While algorithms have been shown to be attributing factors, users themselves seem to process information differently based on their pre-existing notions and beliefs [42, 61].

Facing vast amounts of information online, people adopt cognitive strategies to filter and sift through content more effectively. Such behaviour fosters the occurrence and application of what is referred to as *cognitive biases*, i.e., mental shortcuts we take while processing information. Personal preferences and prior experiences play heavily into this simplification of information processing by focusing on the known or familiar [102].

Misinformation tends to thrive in an environment of simplification and repetition. Its spread, prevalence, and persistence have had real-world implications, such as negative health impacts. For example, the belief in a link between vaccinations and autism has led to parents withholding crucial immunisation from their children resulting in the return of preventable diseases [83]. Misinformation about the dangers and risks of vaccinations keep influencing public debates about the effectiveness of COVID-19 measures to this date [58].

Misinformation is further fueled by frequent exposure. What we encounter more often appears more familiar and can be falsely attributed to a certain truism. When Weaver and colleagues [108] repeatedly showed study participants the same statement from the same communicator, for example, participants perceived the general consensus on that statement to be greater the more often they encountered it. Hence, systems, websites, and platforms that cater to our interests and beliefs tend to skew our perceptions and amplify our innate cognitive biases [7]. This becomes an even bigger problem when it affects our decision-making and opinion formation in the real world, such as on topics like climate change, immigration policies, or abortion rights.

Especially such polarising topics often lead to the segregation of like-minded people. Echo chambers and filter bubbles are two well-known phenomena that contribute to one-sided information exposure and the spread of misinformation. They capitalise on people's biases, most and foremost on what is referred to as confirmation biases [3, 45, 62, 81]. This bias is expressed in people's tendencies to seek out and favour information that aligns with their existing beliefs and expectations while ignoring dissenting information [56, 75]. While it is crucial to mitigate the negative effects of cognitive biases, we first have to understand when and in what situation biases occur, what triggers them, and how they can be reliably *quantified*.

Researchers have examined behavioural measures for exposing confirmation bias or what Klapper called *selective exposure*, i.e., the tendency to seek out predominantly information that supports one's beliefs [51]. This effect has been demonstrated to be present in news dwelling time [37], web browsing behaviour [54, 103], and eye-tracking information [68, 91, 99]. Behavioural measures provide an unintrusive way of tracking selective exposure [20]. Yet, these measures have produced mixed results and interpretations. For instance, researchers used dwelling time as an indicator of confirmation bias as studies have shown that users spend more time reading congruent information and less time on dissenting information [37, 68]. Meanwhile, some research found a rather opposite effect as users spent more time reading attitude-challenging opinions [37, 100]. At the same time, Sülflow et al. [99] and Zillich and Guenther [112] reported no significant differences in reading time between congruent and dissenting information.

A major difficulty in researching cognitive biases is obtaining reliable ground truth for their occurrence. While we could simply ask users whether they have exhibited biased information consumption behaviour, self-report responses are not always reliable since they may be confounded by a broad range of factors, like self-presentation [96] and preference falsification [59]. Recent research has investigated the use of physiological sensors in evaluating cognitive biases [74, 76, 105, 110]. Physiological signals have been regarded as the (more) objective means to quantify mental states [38]. They reflect how our brains and bodies respond and process information [104]. Although physiological signals may be objective measures of our innate cognitive biases, it is unclear how biases manifest themselves in physiological data or can be effectively measured.

In this work, we focus on whether *physiological signals can be a reliable, objective measure of cognitive biases* in an attempt to equip computing systems with the ability to detect and eventually help users mitigate them. We were specifically interested in the occurrence of cognitive biases while processing information that is either congruent with or diverges from people's existing beliefs. Hence, we conducted two studies in which we exposed participants to stimuli that represented an ideologically congruent opinion and those depicting a dissenting opinion. Throughout these studies, we recorded behavioural and physiological signals, such as eye movement data, electrodermal activity, and brain oxygenation levels (via fNIRS) along with self-reports to explore physiological and behavioural expressions indicating the congruence between users' opinions and the presented statements. We also investigated the interplay between the manifestations of biases and the individuals' interest and familiarity with the topic.

Our results show that participants tended to spend more time but less reading effort on ideologically dissenting stimuli. We also found that topic interest significantly impacted the effects of opinion congruency: especially individuals with low interest in a topic exhibited higher neural activity when they were exposed to attitude-dissenting information. Through this work, we contribute the following:

- We present two studies aiming to explore how cognitive biases manifest themselves in behavioural and physiological signals by presenting ideologically polarised statements and recording physiological and interaction data as well as self-reports.
- Based on our findings, we discuss the notion of *bias-aware systems* – i.e., computing systems that detect and take into account the presence of cognitive biases in users – and their potential to detect, quantify, and mitigate the effects of cognitive biases. We discuss challenges, opportunities, and ethical considerations for bias-aware systems from what we learned from this research.

2 BACKGROUND

Our work is mainly grounded in research on behavioural psychology and psychophysiology while touching on recent discussions in human-computer interaction [25–27] regarding the unintended effects of cognitive biases in users.

2.1 Cognitive Biases

Cognitive biases refer to a systematic pattern of deviation from norm or rationality in judgement [39]. The concept was proposed in the work of Tversky and Kahneman in 1974 [102]. Tversky and Kahneman explained different types of heuristics, or so-called *mental shortcuts*, employed by humans to avoid overwhelming their limited cognitive resources by preferably using automatic thinking (System 1) over rational thinking (System 2) [48]. While heuristics enable us to reach a decision faster, they become problematic as they generally distort our rationality in ways we are unaware of.

When making decisions or judgments, individuals who exhibit cognitive biases tend to follow their own beliefs or preferences rather than objective information [39]. In the context of information consumption, this leads to a distortion of the way people perceive and evaluate information, often resulting in favouring information that supports their attitudes [47]. Cognitive biases can be present in many forms. Prominent examples include confirmation bias (seeking predominantly information that aligns with one's beliefs [75]), cognitive dissonance (avoiding information that conflicts with one's beliefs [30]), or negativity bias (responding to negative stimuli with stronger attention and emotional responses [52]). Confirmation bias and cognitive dissonance, for example, are potential contributors to selective exposure [72, 95, 96]. This describes the tendency to seek out predominantly information that supports one's beliefs or attitudes while avoiding dissenting information [51]. This impacts how critical people evaluate information [80, 113] and potentially fosters ideological polarisation [54, 97]. A prominent example in the 20th century was the use of one-sided news reporting by the German government in the 1930s and 40s. Consequently, the belief systems of the majority of Germans who grew up under the regime were skewed towards anti-semitism [106]. In a similar, but less extreme fashion, the recent examples of vaccine hesitency [28] and climate change denial [70] have shown that confirmation bias limits informed and objective discussions of polarizing topics.

People tend to save up their limited cognitive resources when processing information, which makes them vulnerable to various types of manipulation [39]. Today, new information is continuously available to people, which results in excessive mental demand, or mental overload. To prevent overexerting their cognitive resources, people employ cognitive biases or "mental shortcuts" to simplify the complexity and filter out the most relevant information. This is expressed in making faster but less deliberate decisions [48].

Together, cognitive biases and personalised recommendation algorithms contribute to the formation of filter bubbles through a reinforcing loop [3, 62]. When exploring information online, users exhibit their cognitive biases by selectively exposing themselves to certain types of information. Meanwhile, recommendation algorithms detect patterns in the selective consumption of information and optimise themselves to keep engagement high by catering predominantly to what the users prefer [9, 46]. Consequently, the users' innate biases are further amplified. In sum, recommendation systems and selective exposure build a self-reinforcing loop: the former curate content items that are congruent with the users' preferences; at the same time, users seek and favour such content due to confirmation bias [3]. In other words, cognitive biases in individuals can be reinforced by automated recommendation systems.

2.2 Two-step Model of Processing Conflicting Information

While cognitive biases often manifest when facing different opinions, their occurrences also depend on the prior background of the information consumer. In his series of works, Richter [85–87] proposes a two-step model of validation. The model states that people tend to use the perceived plausibility of the information as their heuristics. When encountering information, people first employ *Epistemic Monitoring* to evaluate whether the content is compatible with their beliefs or preferences. In general, people save up their cognitive resources by allocating them to information that is congruent with their beliefs. This results in people processing information with cognitive biases. However, individuals with higher working memory resources, advanced epistemological beliefs, or relevant background knowledge may pursue the second step – *Elaborative Processing* – at which they process the information in a more balanced and objective manner.

2.3 Quantifying the Effects of Cognitive Biases

Being able to quantify the occurrence and the effects of cognitive biases comes with numerous benefits [65]. With the awareness of the users' biases, interventions can be designed to help users overcome their irrationality and become more critical and deliberate when facing information online. However, given that cognitive biases normally happen without people being aware of them, it is challenging to objectively define and measure them [5]. In this section, we review methodological approaches to quantifying the effects of cognitive biases in the context of information consumption, using behavioural measures and physiological signals.

2.3.1 Behavioural Measures. Recent research in the field of selective exposure has used behavioural measures, i.e., through direct observations or in-lab studies [20]. By exposing users to attitudinal information, researchers were able to observe the deviation of users' behaviour as by-products from the manifestation of their innate cognitive biases. Commonly, researchers have used measures like dwelling time – i.e., the amount of time participants exposed themselves to certain types of information – and information selectivity (e.g., the number of content clicks or page visits). Recent approaches have utilised eye tracking measures as they offer advantages over dwelling time, for example, more insights into the users' visual attention [68, 99].

While behavioural expressions offer an unobtrusive measure of bias, research that employed behavioural measures has produced mixed results. Some works showed that people tended to spend more time on what confirms their opinions [68, 91]. Marquart [68], for example, tracked fixation time in online news reading and found that people tended to spend more time with news items that were compatible with their beliefs. Meanwhile, some studies suggested a rather opposite phenomenon [37, 100]. Taber and Lodge [100] found that individuals spent significantly longer time reading attitude-challenging arguments. Some works reported no significant deviation in dwelling time [99, 112]. For instance, an eye-tracking study by Sülflow et al. [99] suggested no effects of opinion congruency on the users' attention to social media news posts but found higher selectivity for attitude-reinforcing contents.

2.3.2 Physiological Measures. Given that our innate biases are the consequence of the interplay of the complex regulation of our cognitive and affective states, cognitive biases are likely to induce physiological changes. Research has long investigated the effects of cognitive dissonance on human physiology. Since the introduction of cognitive dissonance by Festinger [30], a series of studies have investigated a psychological construct called *dissonance arousal* which manifests itself in the form of physiological discomfort [111].

Research by Westen et al. [110] has probed the presence of cognitive biases using physiological signals. Westen and colleagues used functional Magnetic Resonance Imaging (fMRI) to assess the effects of cognitive dissonance and found significantly higher neural activations when the users were processing ideological dissenting information. Subsequent works have confirmed such findings [12, 49]. Meanwhile, Ploger et al. [82] used electrodermal activity (EDA) and heart rate to assess dissonance arousal by exposing individuals to video clips that present attitude-challenging information. However, they found weak effects from ideological (in)congruency.

2.4 Physiological Signals

Physiological signals have been widely used as a surrogate to measure cognitive states [14]. They reflect the reactions from our brains and bodies through a variety of signals. In our work, we focus on two particular signals: electrodermal activity, a widely used physiological measure in HCI, and hemodynamic responses, a non-invasive way to measure brain activities.

2.4.1 Electrodermal Activity. EDA refers to the variation of the electrical conductance of the skin [8], which results from the skin's sweating function. The changes in the sympathetic nervous system control the level of sweating on the skin and thus the EDA. The signal is often collected from electrodes placed on specific body parts, for example, on the fingers or the wrist. In the HCI community, EDA is known as a low-cost, unobtrusive physiological measure [4, 22].

EDA consists of two signal components: Skin Conductance Responses (SCR) and Skin Conductance Level (SCL). SCR represents high-frequency, short-term spikes in the EDA signal triggered by eliciting stimuli. SCL denotes inertial, long-term changes in the EDA. Researchers have used EDA as a marker for negative cognitive activity, for example, cognitive workload [57, 93] and arousal [22, 35, 67].

2.4.2 Hemodynamic Responses. To quantify hemodynamic responses or the changes in blood flow to the brain, researchers have used functional Magnetic Resonance Imaging (fMRI) and functional Near-Infrared Spectroscopy (fNIRS) to infer the relative changes in the concentration of oxygenated haemoglobin ([HbO]) and deoxygenated haemoglobin ([HbR]) [16, 50]. Since haemoglobin absorbs near-infrared light, one can derive the haemoglobin concentration as a function of optical density [6]. Greater changes in haemoglobin concentration are associated with higher levels of neural activation. Therefore, fMRI and fNIRS offer a measurement of innate neural activity [92].

Unlike fMRI, fNIRS provides a less invasive and more noise-robust method to monitor the hemodynamic responses and, thus, the brain activity [66]. Recent research has employed fNIRS to assess a variety of psychological constructs, for example, cognitive workload [2, 31] and affective states [40, 43].

2.5 Summary

Our biases are especially problematic when they come into play for nuanced discussions on polarised topics. They are often exacerbated by the way we consume information online. While they serve us when sifting through vast amounts of information, they at times compromise our ability to make objective decisions. Prior research has investigated how to "track down" the presence of cognitive biases by studying their effects on behavioural measures. While dwelling time as a behavioural measure may be an indicator of cognitive biases, it has been shown to not always be reliable. Recent research has investigated the use of physiological responses to probe the effects of bias. In the context of information consumption, researchers have used fMRI, EDA, and heart rate to observe such effects. Our work adds up to the literature by using physiological signals to monitor the presence of cognitive biases when exposed to opinions from different ideological spectra. To the best of our knowledge, our work is the first to apply fNIRS signals to study the effects and occurrences of biases in the context of information exposure with the intent to study the notion of bias-aware computing systems. In the following, we present two studies, in which we expose participants to a range of opinions and probe their interactions, behavioural expressions, and physiological data to explore how cognitive biases may manifest themselves.

3 STUDY 1: DESIGN

We conducted Study 1 to explore different indicators for the occurrences of cognitive biases. In this study, we exposed participants to textual and image stimuli that represented opinions on four polarising topics. At the same time, we recorded behavioural data (eye tracking) and physiological signals, namely electrodermal activity (EDA), and brain hemodynamic responses using functional near-infrared spectroscopy (fNIRS).

3.1 Stimuli Selection

We operationalised stimuli that consisted of information on either end of the ideological spectrum, i.e., supporting information (pro) or contradicting information (con). Adapting to the Australian context, where this study was conducted, each stimulus was chosen with regard to ideologically polarising topics that were dominant in the current, domestic public debate. Consequently, we selected the following four topics for the study: *political progressivism*, *climate change*, *feminism*, and *multiculturalism in Australia*. All four topics were widely discussed in the media, and well-known to the Australian public with increasingly polarised viewpoints. Thus, we expected that the stimuli would have the potential to trigger strong attitudes and prompt cognitive biases in the study participants. Table 1 gives an overview of the pro-stances and con-stances for each of the four topics.

We selected **progressivism** due to the increasing ideological polarisation between progressive and conservative politics[1] since the 1970s [18, 19, 71, 109]. Similarly, we chose **climate change** because of the increasing discrepancy between those who acknowledge man-made climate change as opposed to denying it [64]. We also considered **multiculturalism** due to the lasting conflict between multiculturalism in Australia and the Anglo-Saxon inheritance rooted in the "White Australia" policy [21]. Lastly, **feminism** was selected because of the increasing pushback against feminism among Australian male groups [88] and third-wave feminists [98].

We used two types of stimuli: texts and images. Text stimuli were curated from either user opinions on Twitter[2] or the Procon.org website[3]. The latter source hosts information on both ends of the ideological spectrum, i.e., pros and cons, on different topics. We sourced climate change and feminism stimuli from Procon.org; progressivism and multiculturalism stimuli were curated from tweets posted in Australia from June to July 2021. We controlled all text stimuli for being in English and approximately 50 words in length.

While statements on ProCon.org are heavily contextualised to US politics and society, topics of global interest and the general discourse, such as climate change and feminism, are also applicable to the Australian context. We selected statements that do not contain US-specific information, e.g., excluding those mentioning US laws.

Each image stimulus was selected from online images or graphics that contained messages supporting an ideological viewpoint. Similarly, we picked those images from the ProCon.org website or keyword searches on Twitter. Examples of image stimuli were the cover of the book "The Greatest Hoax" [44] or a photo of a protest against man-made climate change.

We accumulated a total of 64 stimuli consisting of 32 texts, and 32 images. Stimuli presenting pros and cons were even in numbers. We presented participants with each stimulus on a screen. Every text stimulus was displayed in a single paragraph with the same font

[1] To avoid confusion among our readers, we use the Anglo-Saxon nomenclature. In Australia, conservative politics are actually called "liberal", whereas the Australian "labor" is the equivalent to the Anglo-Saxon progressives.
[2] www.twitter.com
[3] www.procon.org

Table 1: Topics and their ideological ends, as used in Study 1

Topic	Pro stance	Con stance
Political Progressivism	I support a political and societal change	I do not support political and societal change
Climate change	I believe humans are primarily responsible for climate change	I believe humans are not primarily responsible for climate change
Multiculturalism in Australia	I support multiculturalism in Australia	I support the Anglo-Saxon national identity of Australia
Feminism	I support feminism and women's rights	I do not support feminism and women's rights

Overwhelming scientific consensus finds human activity primarily responsible for climate change. Most climate scientists and scientific organizations agree that human activity is extremely likely to be the cause of global climate change.

(a) Text

(b) Image

Figure 1: Examples of stimulus presentation for Study 1. Both stimuli were on the topic of climate change.

(Arial 30 px, black colour), line spacing (double), alignment (justified and centred), column width (800 px), and white background. Image stimuli were presented in an 800px × 800px resolution. Figure 1 shows an example of the stimuli used in Study 1.

3.2 Study Protocols

3.2.1 Experimental design. We studied the effects of the congruency of ideological stances between the user and the stimulus. To do so, we conducted two experiments with a 2-level (*Congruent* and *Dissenting*) within-subjects design: one to examine text stimuli and one to examine image stimuli. The *congruent* condition implied the stimulus' stance was aligned with the user's stance. Conversely, the *dissenting* condition implied the stimulus' stance contradicted the user's stance. Table 2 shows a list of independent and dependent variables of this study.

3.2.2 Participants. Through the university network, we invited 33 native or bilingual English speakers (19 women, 14 men) to participate in Study 1. The mean age of our participants was 32 (SD = 11.43) years. The minimum and maximum ages were 18 and 54, respectively. Of our participants, 15 possessed a postgraduate degree, 10 held a bachelor's degree, and the remaining six participants had at least year 12 education.

3.2.3 Procedure. The study took place in a quiet room. We informed each participant about the purpose and procedure of the study. After providing their consent in writing, we seated participants in a comfortable position and asked them to adjust their seats so that their heads were centred and approximately 60-65cm away from the monitor screen. We then asked each participant to respond to the pre-study survey and calibrated the placements of the physiological sensors.

We subsequently asked participants to read a series of text and image stimuli on a 24-inch monitor. After each stimulus, participants were asked to press the space key to proceed to the next one. The order of stimuli presentation was counterbalanced: participants either completed image stimuli first then text stimuli, or vice versa. Moreover, the order of the four topics was counterbalanced. Within each topic, stimuli were displayed in random order with no gap in between. Once a participant finished all stimuli for a topic, we paused the data collection for approximately one minute; then, participants continued reading the stimuli on the following topic. Upon completion, participants responded to a post-study survey and received a $20 voucher for compensation. The whole study took 45-60 minutes.

3.2.4 Sensors. Throughout the study, we recorded participants' eye movements, EDA, and hemodynamic responses. Eye movements were recorded with a Tobii Pro X3-120 eye tracker[4] with a sampling rate of 120 Hz. We mounted the eye tracker at the bottom of the monitor. We used the Empatica E4 wristband[5] to gather EDA data. To prevent potential motion artefacts, we asked participants to wear the wristband on their non-dominant hand. Additionally, we recorded functional near-infrared spectroscopy (fNIRS) from the participant's forehead using the BIOPAC fNIR Sensors 2000[6].

[4] https://www.tobiipro.com/product-listing/tobii-pro-x3-120/
[5] https://www.empatica.com/research/e4/
[6] https://www.biopac.com/product/fnir-sensors-2000/

Table 2: Summary of Independent and Dependent Variables in Study 1

Variables	Measures	Scale
Independent Variables	Participant-Stimulus Ideological Congruency	2 levels (congruent and dissenting)
Dependent Variables	Dwelling Time	Continuous
	Number of Fixations	Number of occurrences
	Number of up, down, left, and right saccades	Number of occurrences
	SCL: Skin Conductance Level	Continuous
	Frequency of Skin Conductance Response (SCR) peaks	Number of occurrences
	Δ_{all}[Hb]: The Overall Brain Oxygenation Level	Continuous

The device offered a sensor pad comprising 18 optical sensors that record fNIRS signals with a sampling frequency of 20 Hz. We attached this sensor pad to the participant's forehead to monitor hemodynamic responses in the frontal lobe of the brain. During the recording session, we asked participants to refrain from moving their heads and the non-dominant hand to minimise the occurrence of motion artefacts.

3.3 Ground Truth

3.3.1 Pre-study Survey. For each of the four topics, we asked participants to rate their stance on the topic using an 11-point Likert scale (−5: *I agree with the con stance* to +5: *I agree with the pro stance*). We presented the pro and con stances according to Table 1. Participants also rated their interest in and familiarity with each topic using a 5-point Likert scale (1: *least interested* to 5: *completely interested*) and a 10-point Likert scale (1: *least familiar* to 10: *most familiar*), respectively.

3.3.2 Post-study Survey. After completing the data collection, we asked participants to reevaluate the stimuli they have seen in the study. Each participant rated the expressiveness of each stimulus on a 7-point Likert scale (1: *very weak* to 7: *very strong*). Each question was accompanied by the corresponding stimulus.

3.3.3 Participants' Ideological Stances. We gathered the users' ideological stances through the pre-study survey's responses. These were used to determine the congruence of stances between each participant and each stimulus. A stimulus S is considered *congruent* with participant P if the stances of S and P were in agreement. On the other hand, if the stances of S and P were opposite, S is *dissenting* with P.

We employed a threshold of 0 on the stance ratings (ranging from −5 to +5) to determine the participants' attitudes. For example, on the topic of climate change, a positive score implied the participant's stance aligned with the idea that climate change is man-made (i.e., "*I believe humans are primarily responsible for climate change.*"). Conversely, a negative score represented the stance that climate change is not man-made (i.e., "*I believe humans are not primarily responsible for climate change.*"). Participants who rated 0 on a topic were considered as having a neutral attitude on that particular topic. In our data analysis, we discarded any stimulus exposure that involved participants who had a neutral stance on a topic.

Among 33 participants who joined Study 1, we observed that most participants aligned themselves with the *pro* stances of every topic: progressivism (pro : con : neutral = 27 : 3 : 3), climate change (pro : con : neutral = 30 : 1 : 2), multiculturalism in Australia (pro : con : neutral = 30 : 1 : 2), and feminism (pro : con : neutral = 28 : 4 : 1).

4 STUDY 1: RESULTS

We analysed the data collected in Study 1 and examined the effects of ideological congruency on dwelling time, behavioural data, and physiological signals.

4.1 Dwelling Time

A one-way repeated measures ANOVA was performed on the amount of time each participant spent with each stimulus. We set the independent variable to be the congruence of ideological stance between the participant and the stimulus, which had two levels: congruent (C) and dissenting (D). For both text and image stimuli, we found that participants spent significantly more time with dissenting stimuli than congruent stimuli (text: $\mathcal{F}(1, 31) = 18.911, \eta_p^2 = 0.37, p < 0.001$; image: $\mathcal{F}(1, 29) = 4.416, \eta_p^2 = 0.13, p = 0.0444$). We found a weaker effect size (text: $\eta_p^2 = 0.37$, image $\eta_p^2 = 0.13$) for image stimuli.

4.2 Eye Tracking Measures

4.2.1 Preprocessing. We first obtained the raw gaze data, which consisted of the (x, y) coordinates on the projection screen. Subsequently, we used the Tobii Pro Lab's I-VT gaze filter [79] to estimate the velocity of the participant's eye movement. Those with a velocity below the threshold were considered *fixations* – a type of eye movement where the eyes are focused on one point. Those with a higher velocity were treated as *saccades* – rapid eye movement from one point to the other. For each stimulus, we obtained the eye-tracking features by calculating the number of fixations and the number of saccades in each of the four directions (up, down, left, and right) during the exposure to the stimulus.

4.2.2 Data Analysis. Since our eye tracking features consisted of count data which are often not normally distributed, we applied a Friedman test on the counts of fixations and saccades of both text stimuli data and image stimuli data. We subsequently corrected the p-values using a one-step Bonferroni correction.

For text stimuli, we found that participants exhibited significantly more fixations ($p_{corr} = 0.0174, \chi^2 = 12.461$), and right saccades ($p_{corr} = 0.0036, \chi^2 = 15.384$) with dissenting stimuli than congruent stimuli. For image stimuli, we observed significantly more fixations ($p_{corr} = 0.0283, \chi^2 = 11.560$) when viewing dissenting stimuli compared to congruent stimuli.

4.3 Electrodermal Activity

4.3.1 Signal Preprocessing. EDA signals recorded from a wearable device may contain motion artefacts. We, therefore, applied a lowpass filter with a cutoff frequency of 3 Hz to remove potential high-frequency motion artefacts. Subsequently, we applied a highpass filter with a 0.05 Hz cutoff frequency to extract the skin conductance responses (SCR) and the skin conductance level (SCL). SCR peaks were then identified by applying a peak detection algorithm to the SCR signals. Lastly, we derived two EDA measures: the mean of SCL and the count of SCR peaks throughout the period of exposure to a stimulus.

4.3.2 Data Analysis. Similar to the eye tracking data analysis, a Friedman test was performed on the EDA features and a one-step Bonferroni correction was used to correct the p-values. For text stimuli, we found that participants exhibited significantly greater counts of SCR peaks on dissenting statements ($p_{corr} = 0.0052, \chi^2 = 15.695$) than congruent statements. For image stimuli, however, we did not detect any significant effects of ideological congruence on EDA statistics.

4.4 Brain Hemodynamic Responses

4.4.1 Signal Preprocessing. The fNIRS we used in this study measured the optical density in two near-infrared frequencies, 730nm and 850nm. However, these signals are susceptible to noise, such as motion and physiological artefacts. Thus, for each participant, we first identified and discarded data that were distorted because of bad optode placement, i.e., when they were obstructed by hair or interfered with ambient light. Bad optode placement was considered if either (1) 90% of the optode's raw optical density fell outside an acceptable range of [400 mV, 4000 mV]; or (2) the raw optical density's coefficient of variation (defined as the ratio of the signal's standard deviation and mean) exceeded 20%.

Subsequently, we corrected noise and motion artefacts in the signals using 10-second time epochs. This involved two steps; first, we applied a bandpass filter with cutoff frequencies between [0.001 Hz, 1 Hz] on the optical densities to filter out signals from irrelevant frequency bands. Subsequently, the Temporal Derivative Distribution Repair (TDDR) algorithm [32] was applied to the filtered optical densities to correct motion and physiological artefacts.

We also manually removed parts of the recordings that consisted of suspected motion artefacts, i.e., rapid spikes in the signal which were caused by participants' body movement. We then subtracted the optical densities with the initial 5-second baseline. The baseline was recorded before the data collection started when participants were sitting still for about 20 seconds. The baselined optical densities were converted to oxygenated haemoglobin and deoxygenated haemoglobin concentrations ([HbO] and [HbR]) using the modified Beer-Lambert law [6]. [HbO] and [HbR] were then standardised within each participant to mitigate the effects of individual differences. Subsequently, we subtracted [HbR] from [HbO] and obtained the brain oxygenation level, Δ[Hb] = [HbO] − [HbR]. This step was done in order to improve the signal strength. Then, for each participant, we obtained the overall oxygenation level, Δ_{all}[Hb], by averaging Δ[Hb] across all available optodes. We opted to use Δ_{all}[Hb] to represent the overall changes in the forehead hemodynamic activity. Lastly, for each stimulus exposure, we calculated the mean Δ_{all}[Hb] for the exposure period.

4.4.2 Data Analysis. We applied a one-way repeated measures ANOVA on the mean overall oxygenation level, Δ_{all}[Hb], for each time window of stimulus exposure. However, we found no significant effects of ideological congruence on the mean overall oxygenation levels.

4.5 Summary and Lessons Learned

Our findings indicate that participants tended to spend more time and exhibited more fixations when facing ideologically dissenting stimuli. This implies that dissenting information might hinder or disrupt the comprehension process. However, it was inconclusive whether cognitive biases did contribute to this phenomenon. One possible assumption could be that ideologically dissenting text stimuli (i.e., the *con* statements) were more cognitively demanding than ideologically congruent stimuli [94]. Alternatively, since the make-up of the study participants were predominantly aligned with the *pro* statements, the dissenting stimuli may also have systematically caused longer dwelling time.

Although we found a significant effect on the counts of SCR during exposure to text stimuli, it remained inconclusive whether physiological signals are reliable indicators of cognitive biases. As a potential explanation, the study design may have introduced confounding factors: we did not provide a time gap between two consecutive stimuli, so-called inter-stimuli intervals (ISI). Due to the lack of ISI, the stimulus-related physiological responses may not reflect the reactions induced by the stimulus itself but those induced by the preceding stimuli.

In addition, we observed that image stimuli yielded less expressiveness than text stimuli for two reasons. First, we found no significant effect of the image stimuli's ideological congruence on dwelling time. Secondly, we found that the self-report expressiveness ratings on image stimuli were significantly lower than on text stimuli (one-way repeated measures ANOVA: $\mathcal{F}(1, 31) = 5.808, \eta_p^2 = 0.16, p = 0.022$).

Visual information is one of the most prevalent media on the Internet and is highly contextual, usually presented together with text information [89]. In contrast to text stimuli, image stimuli can thus be ambiguous, leading to different interpretations in different individuals.

To eliminate possible confounds, we conducted a follow-up study, which (1) ensured that the polarising statements successfully induced biased information processing, (2) used a reliable ground truth for the induced biased information processing, and (3) allowed us to observe clearer changes in physiological signals in response to each stimulus.

5 STUDY 2: DESIGN

As Study 1 was inconclusive as to whether cognitive biases were induced, we cannot draw any conclusions as to what extent physiological signals can be used to infer the presence of cognitive biases yet. We, therefore, designed and conducted Study 2 to address the same question as Study 1 – are physiological signals reliable, objective measures of cognitive biases? Study 2 comprised a similar approach in that we exposed participants to a series of polarising statements, but revised the experimental design to account for potential confounding factors.

5.1 Stimuli Selection

We employed 62 text stimuli in Study 2. We decided to expose our participants to a wider range of opinion statements. Thus, we aimed to increase the external validity of Study 2 by diversifying our stimuli and obtaining more observations. We extended the number of topics to eight in Study 2: *progressivism*, *climate change*, *feminism*, *multiculturalism in Australia*, *vegetarianism*, *renewable energy*, *abortion*, and *same-sex marriage*. We used the 32 original text stimuli from Study 1 and introduced 30 additional stimuli for the four new topics. We provide details of each new topic in the following paragraph. Informed by Study 1, we opted for not using image stimuli, as they proved difficult to limit confounding factors like the expressiveness and ambiguity of the images.

We included **vegetarianism** as one of the new topics because of an increasing debate (about 12% of Australians identify as vegetarians [101]) between proponents of vegetarianism (i.e., those who do not eat meat) and its opponents (i.e., those who support meat consumption). Meanwhile, we selected **renewable energy** since it is contextually parallel to the topic of climate change. In Australia, there has been a growing political debate between supporters and opponents of renewable energy [23]. We also selected **abortion** and **same-sex marriage** because they are part of the discussions on feminism and progressivism. Although abortion has been legalised in Australia, a notable proportion of pro-life messages still exist on the Internet [1]. Similarly, same-sex marriage in Australia was a heated debate during the 2017 marriage law survey [84]. While the poll showed that the majority of Australians (61%) expressed support for same-sex marriage, there was a significant proportion of those who voted "no" [33].

The 30 new stimuli were gathered from the Procon.org website. The ideological stances were counterbalanced, i.e., there was an equal number of pro and con statements. We controlled the length of each text stimulus to be around 50 to 80 words. In addition, we ensured that no text stimulus had a score lower than 30 according to the Flesch reading ease score [34], which is equivalent to the university level. In Table 3, we summarise the pro and con ideologies for each of the four additional topics.

Similar to Study 1, we presented the stimuli on a computer monitor. Each stimulus was displayed with the same font (Verdana 27 pt, in dark grey colour), double line spacing, centred-justified alignment, and white background. Figure 2 gives an example of the text stimuli in Study 2.

5.2 Study Protocols

5.2.1 Experimental Design. We conducted the study with a within-subject design with the independent variable being participants' congruent or dissenting opinion (i.e., two levels). Our dependent variables consisted of behavioural measures (dwelling time and eye tracking data), physiological measures (EDA and brain hemodynamic responses), and self-report measures (stimulus-wise ideological alignment, likelihood to share the stimulus, and cognitive effort). We present a list of study variables of Study 2 in Table 4.

5.2.2 Participants. We invited 31 participants (16 female, 13 male, and 4 preferred not to disclose) to Study 2. The mean age was 29.41 (SD= 11.17) and ranged from 18 to 68 years old. Of those who disclosed their age, 6 were between 18 and 20 years old, 10 were in their 20s, 6 were in their 30s, 2 were in their 40s, and 6 were 50 years old or older. All participants reported that they were either native, bilingual, or professional users of English. For their highest level of education, 10 had year 12 education, 1 had certificate III/IV education, 6 had a bachelor's degree, 4 had a graduate diploma/certificate, and 10 had a postgraduate degree. We excluded two participants for fNIRS data analysis and one participant for EDA data analysis since their recordings were mostly corrupted or missing.

5.2.3 Procedure. Similar to Study 1, Study 2 took place in a quiet room where participants were seated in a comfortable position in front of a 24-inch monitor. We first informed each participant about the purpose and protocols of the study. After receiving their written consent, we asked the participants to answer a pre-study survey and calibrated the physiological sensors. After that, participants went through a warm-up round to familiarise themselves with the protocols. We presented participants with a series of four text stimuli on the topic "*Should zoos exist*?". These warm-up stimuli were sourced from Procon.org[7].

In this study, we exposed participants to stimuli differently from Study 1. For each stimulus, participants first read the stimulus statement. Once they finished reading it, they responded to an in-study survey, which asked participants three questions regarding the stimulus. After providing their responses, participants entered a 15-second resting period, where we asked them to close their eyes and count from 1 to 15. A 15-second timer was placed on a screen. Once the timer counted down to 0, participants proceeded to the next block by clicking on the "next" button. The presentation order of the stimuli was randomised.

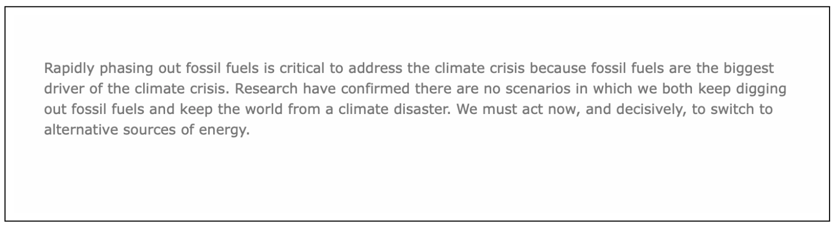
We introduced a 15-second resting period as an inter-stimulus interval (ISI) in order to observe clearer physiological changes. We decided that 15 seconds would be an appropriate ISI since it allowed sufficient time to observe hemodynamic responses, which typically take three to five seconds to reach a peak and a few seconds to decay [69, 107].

After participants finished the warm-up round, they entered the data collection round. In this round, we presented participants with a series of 62 text stimuli from the eight topics mentioned. Like the warm-up round, participants read the stimulus, responded to an in-study survey, and entered a 15-second resting period. Each stimulus was presented in a randomised order; each stimulus' topic and stance were also randomised. Upon completion, we engaged

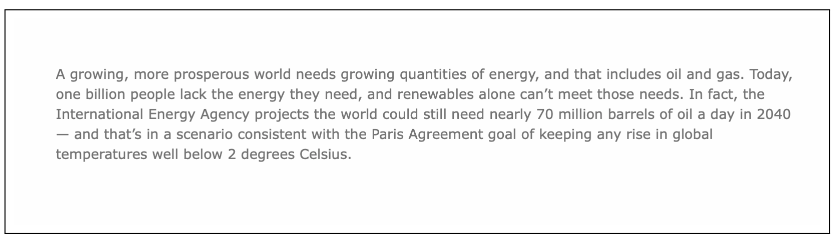
[7]https://www.procon.org/headlines/zoos-top-3-pros-and-cons/

Table 3: Additional topics in Study 2 and their ideological ends

Topic	Pro stance	Con stance
Vegetarianism	I support vegetarianism and oppose meat consumption	I support meat consumption and oppose vegetarianism
Renewable Energy	I believe renewable energy is necessary	I believe renewable energy is not necessary
Abortion	I think abortion should be legal	I think abortion should be prohibited
Same-sex marriage	I think same-sex marriage should be legal	I think same-sex marriage should be prohibited


Rapidly phasing out fossil fuels is critical to address the climate crisis because fossil fuels are the biggest driver of the climate crisis. Research have confirmed there are no scenarios in which we both keep digging out fossil fuels and keep the world from a climate disaster. We must act now, and decisively, to switch to alternative sources of energy.

(a) A pro stance on renewable energy


A growing, more prosperous world needs growing quantities of energy, and that includes oil and gas. Today, one billion people lack the energy they need, and renewables alone can't meet those needs. In fact, the International Energy Agency projects the world could still need nearly 70 million barrels of oil a day in 2040 — and that's in a scenario consistent with the Paris Agreement goal of keeping any rise in global temperatures well below 2 degrees Celsius.

(b) A con stance on renewable energy

Figure 2: Examples of stimuli presentation for Study 2

participants for a brief interview and compensated them with a $20 cash voucher. The study took approximately 90 minutes.

5.2.4 Sensors. Throughout the experiment, we recorded physiological data from the participants. In a similar fashion to Study 1, we employed the Empatica E4 wristband to record EDA and the BIOPAC fNIR Sensors 2000 to record brain oxygenation levels from the forehead. We used a Tobii Pro Nano[8] to record the participant's eye-tracking data with a sampling frequency of 120 Hz. We asked participants to refrain from moving their heads throughout the data collection period to prevent the occurrence of motion artefacts.

5.3 Ground Truth

5.3.1 Pre-study Survey. For each of the eight topics, we asked participants to rate their stance on the topic on a continuous slider scale of 0 (*I agree with the* ***con*** *stance*) to 100 (*I agree with the* ***pro*** *stance*). Unlike Study 1, we used a continuous scale for ideological stance since it provides more granularity for assessing the participants' stances on a spectrum. In addition, we asked participants to rate their interest and familiarity with each topic on a scale from 1 (*least interested/familiar*) to 5 for (*most interested/familiar*).

5.3.2 In-study Survey. For each stimulus, we asked participants to report the congruence of ideological stance between them and the statement, the likelihood to share it on their social media, and the cognitive effort spent reading it, by asking three questions: (Q1) *How much does the statement align with your beliefs?*; (Q2) *How likely are you to share this statement on your social media?*; and (Q3) *How much effort did you put into reading this statement?*. Participants gave their ratings using a 5-point Likert scale (1: *least aligning/likely/effortful* to 5: *most aligning/likely/effortful*). The in-study survey was triggered each time participants finished reading a stimulus.

5.3.3 Participant-stimulus Ideological Stance. On each topic, we determined the ideological alignment of each participant from their self-reported stance (from 0 to 100) in the pre-study survey. Similar

[8]https://www.tobiipro.com/product-listing/nano/

Table 4: Summary of Independent and Dependent Variables in Study 2

Variables	Measures	Scale
Independent Variables	Participant-Stimulus Ideological Congruency	2 levels (congruent and dissenting)
	Topic Interest	2 levels (high and low)
	Topic Familiarity	2 levels (high and low)
Dependent Variables	**Behavioural**	
	- Dwelling Time	Continuous
	- Number of Fixations	Number of occurrences
	- Number of up, down, left, and right saccades	Number of occurrences
	Self-report	
	- Q1: participant-stimulus ideological congruence	5-Likert scale
	- Q2: likelihood to share the stimulus on one's social media	5-Likert scale
	- Q3: effort spent reading the stimulus	5-Likert scale
	Physiological	
	(Time windows {2.5s, 5s, 10s} × {EXP1, EXP2, POST})	
	- SCL: Skin Conductance Level	Continuous
	- Frequency of Skin Conductance Response (SCR) peaks	Number of occurrences
	- Δ_{all}[Hb]: The Overall Brain Oxygenation Level	Continuous

to Study 1, we applied a threshold of 50 on the stance ratings. A rating of more than 50 represented an ideological stance that supports the pro stance. Conversely, ratings less than 50 were considered to support the con stance.

Using the abovementioned thresholds, our 31 participants identified their stances as follows: climate change (pro : con : neutral = 28 : 1 : 2), feminism (pro : con : neutral = 25 : 2 : 4), progressivism (pro : con : neutral = 24 : 5 : 2), multiculturalism in Australia (pro : con : neutral = 30 : 0 : 1), vegetarianism (pro : con : neutral = 12 : 12 : 6), renewable energy (pro : con : neutral = 29 : 1 : 1), same-sex marriage (pro : con : neutral = 26 : 4 : 1), and abortion (pro : con : neutral = 27 : 3 : 1).

Subsequently, we defined a score that describes the ideological congruency between the participant and the stimulus. The score was in a range between −50 (*the participant's stance is completely opposite of the stimulus*) and +50 (*the participant's stance completely aligns with the stimulus*). A positive score implied that the stances of the participant and the stimulus were in the same direction, and vice versa. The congruency score between participant p and stimulus s, $Congruence(p, s)$, can be derived by applying formula 1. We denoted $Pos(s)$ as the ideological stance of the stimulus s, which took a binary value of +1 if the stance was aligned with the pro opinion or −1 if the stance was aligned with the con opinion. $Stance(p)$ is the self-report stance of the participant p, ranging from 0 to 100.

$$Congruence(p, s) = (Stance(p) - 50) \times Pos(s) \tag{1}$$

For example, if a person rated themself with 80 out of 100 on the topic of abortion and a stimulus stated an anti-abortion statement, the congruency score between them would be $(80 - 50) \times (-1) = -30$.

In this study, we considered stimulus exposures with a congruency score greater than +20 and those with a score lower than −20 to be ideologically *congruent* (C) and *dissenting* (D) respectively. We discarded data points where the congruency score was between −20 and +20 as they were considered neutral or weak in inclination. The scale for the score was continuous with the mean of $M = 0$ and $SD = 37.32$.

6 STUDY 2: RESULTS

We analysed the effects of the congruency between the stimuli's ideologies and the participants' leanings on behavioural, physiological, and interaction measures collected during the study. The goal was to examine physiological expressions of cognitive biases that may be experienced when aligning with or distancing oneself from content items. In the following, we describe our analysis and findings along with the training of a classifier to detect the participant-stimulus ideological congruency from interaction and physiological data on whether participants encountered attitudinal information.

6.1 Effects of Ideological Congruency

We applied a one-way repeated measures ANOVA on the amount of time each participant spent with each stimulus. Similar to Study 1, the independent variable was the congruence of ideological stance between the participant and the stimulus, which had two levels: congruent (C) and dissenting (D). Accordingly, we examined the effects of opinion congruency on dwelling time, self-report measures, eye tracking measures, and physiological (electrodermal activity and hemodynamic responses). Table 5 reports statistical results of the self-report and behavioural measures. Table 6 reports the statistical results of the physiological measures.

6.1.1 Behavioural Measures. We found that participants spent significantly more time with ideologically dissenting stimuli than with congruent stimuli (C: 12.35 ± 7.46 seconds, D: 12.92 ± 7.26 seconds, $\mathcal{F}(1, 30) = 5.713, \eta_p^2 = 0.160, p = 0.023$)

For eye tracking measures, similar to Study 1, throughout the period of stimulus exposure, we calculated the number of fixations and the number of saccades in each of the four directions: up, down, left, and right. Since the length of each exposure was not identical, we normalised the measures by dividing each of them by the stimulus dwelling time. We found no significant effect of opinion congruency on the normalised eye-tracking measures.

6.1.2 Self-report Measures. We performed a similar analysis on the self-reported ratings for each stimulus. We examined (Q1) the ideological congruence between the participant and the stimulus, (Q2) participants' likelihood to share the stimulus on their social media, and (Q3) their cognitive effort spent reading it.

We found that Q1 and Q2 responses from congruent stimuli were significantly higher than those from dissenting stimuli (Q1: C: 3.90± 0.92, D: 2.29 ± 1.12, $\mathcal{F}(1, 30) = 203.481, \eta_p^2 = 0.871, p < 0.001$; Q2: C: 1.83 ± 1.03, D: 1.23 ± 0.51, $\mathcal{F}(1, 30) = 51.564, \eta_p^2 = 0.632, p < 0.001$). This confirms the internal validity of the stimulus materials as the participants' general tendency toward a topic ($Congruence(p, s)$) and their content-specific alignment (Q1) were congruent. Specifically, we found that $Congruence(p, s)$ and Q1 were strongly correlated (Pearson $r = 0.737, p < 0.001$). Moreover, participants with general tendencies in favour of a topic were more willing to share content that aligned with their views. Q3 responses for congruent and dissenting stimuli were not significantly different from each other (Q3: C:2.98 ± 1.20, D:2.76 ± 1.23, $\mathcal{F}(1, 30) = 3.375, \eta_p^2 = 0.101$, n.s.).

6.1.3 Physiological Measures. The task design in Study 2 allowed us to observe physiological changes both during and after stimulus exposure. Thus, we analysed the collected physiological signals in three different time windows: a period during the beginning of stimulus exposure (EXP1: the first 0 to w seconds), a period during the end of stimulus exposure (EXP2: the final w seconds), and a period after stimulus exposure (POST: the first 0 to w seconds after exposure). We analysed the data using three different window sizes (w): 2.5, 5, and 10 seconds. The choices of window size followed those commonly used in prior EDA [11, 22] and fNIRS studies [2, 40]. To ensure that our window analysis is valid, we discarded any exposure that lasted shorter than the defined window size.

Additionally, we corrected the temporal drift in SCL by subtracting the SCL values in the baseline window from the SCL values in the analysis window. For each stimulus, we used the final 2 seconds of the resting period (i.e., the ISI) before the participant started reading it as the baseline window.

We followed the same signal preprocessing pipeline as in Study 1. We examined 3 × 3 dependent variables. From EDA data, we calculated the mean of SCL and the frequency of SCR. For hemodynamic responses, we obtained the mean overall oxygenation levels, Δ_{all}[Hb]. Each of these measures was calculated in the three time windows: EXP1, EXP2, and POST.

In a similar manner, we ran a repeated measures ANOVA on each of the dependent variables. We did not detect a significant effect of opinion congruency on the mean of SCL in any time window. However, we observed a trend during the first 10 seconds of stimulus exposure (EXP1 period) that the mean SCL was higher when presented with dissenting stimuli (C: 0.000436 ± 0.143, D: 0.0203 ± 0.188, $\mathcal{F}(1, 29) = 4.242, \eta_p^2 = 0.127, p = 0.0502$).

We did not find a significant effect of opinion congruency on the overall oxygenation levels; yet, we found significant effects on the overall oxygenation levels of the subgroup of participants who reported low interest in a topic. We discuss this finding in detail in the following section.

6.2 Effects of Interest and Familiarity

We examined whether participants' interest in and familiarity with a topic influenced their self-report, behavioural, and physiological expressions. To do so, we considered subgroups of participants with high/moderate/low interest and familiarity with a topic.

For each topic, we set a threshold of 3 on the interest (1-5) ratings. We considered those who rated topic interest as 4 or 5 to have *high interest*. Participants who rated 3 were regarded as having *moderate interest*. Lastly, those who rated 1 or 2 on interest were deemed as *low interest*. We also applied the same threshold on the familiarity (1-5) ratings to form participant groups with high, moderate, and low familiarity.

There were a total of 1482 observations across 31 participants. When filtered by topic interest, there were 232 observations across 19 participants in the low-interest group and 900 observations across 29 participants in the high-interest group. When filtered by topic familiarity, there were 354 observations across 18 participants in the low-familiarity group and 522 observations across 25 participants in the high-familiarity group.

We analysed the high-interest and low-interest groups separately by employing a one-way repeat measures ANOVA on each of the measures. While we detected some effects of topic interest, we found no effect from familiarity; thus, we provide the analysis only for topic interest in the following. Table 5 gives the testing results, including the sample size for each subgroup.

6.2.1 Dwelling Time. In line with the general results, we found that participants with a higher interest in a topic spent significantly more time with dissenting information (C:11.96 ± 6.46, D: 13.20 ± 7.49, $\mathcal{F}(1, 28) = 13.470, \eta_p^2 = 0.829, p < 0.001$). We also detected a greater effect size in the high-interest group (compared to the general group, $\eta_p^2 = 0.160$), which indicated that the effect of ideological congruency was stronger in participants with high interest. Meanwhile, we found no significant effect when considering data from low-interest individuals.

6.2.2 Self-report Measures. We obtained consistent results from both subgroups: they tended to rate both Q1 and Q2 higher for congruent stimuli. We observed a greater effect size on Q2 in the high-interest group (high-interest group: $\eta_p^2 = 0.655$, general group: $\eta_p^2 = 0.632$), indicating that individuals were more likely to share attitude-confirming contents as they were more interested in the topic. In addition, we found that participants with low interest reported significantly higher effort (Q3) when reading congruent statements than dissenting ones (C: 2.98 ± 1.20, D:2.76 ± 1.23, $\mathcal{F}(1, 18) = 9.348, \eta_p^2 = 0.341, p = 0.006$).

6.2.3 Physiological Measures. When examining data from low-interest individuals, we detected significant effects of ideological congruency on the overall oxygenation levels during the EXP1 period. Our analysis showed that the effects were significant in window lengths of 2.5 and 5 seconds, where participants tended to exhibit higher oxygenation levels when facing dissenting information (2.5-second window: C: -0.18 ± 1.08, D: 0.10 ± 1.11, $\mathcal{F}(1, 16) = 5.352, \eta_p^2 = 0.250, p = 0.034$; 5-second window: C: -0.16 ± 0.99, D: 0.059 ± 1.07, $\mathcal{F}(1, 16) = 4.607, \eta_p^2 = 0.223, p = 0.048$). As higher oxygenation levels associate with more neural activation, our results suggested that ideologically diverging information induced higher neural activity than congruent information. We found no significant effect when considering high-interest individuals.

6.3 Building a Bias Classifier

To examine our measures as indicators of cognitive biases, we performed a binary classification on the collected data to detect and distinguish the exposure to ideologically congruent stimuli (C) from ideologically dissenting (D) ones. We extracted the input features of the classifiers from statistical values of the EDA and brain oxygenation levels. Each of the features was extracted in a 2-second time window in each observation period (EXP1, EXP2, and POST). Statistics include the mean, standard deviation, median, kurtosis, skewness, and slope. We also included eye-tracking features, which were the counts of fixations and saccades in different directions (up, down, left, and right). Due to counterbalancing in the study design, our dataset (4960 samples) was perfectly balanced between the congruent and dissenting conditions (class ratio C : D = 2480 : 2484).

We trained a model using the following classifiers: linear discriminant analysis (LDA), support vector machine (SVM) with an RBF kernel, random forest [13], and XGBoost [36]. We evaluated the models by using the average accuracy across a 5-fold cross-validation. The average accuracy was calculated from the mean of the validation accuracy for each fold. For tree-based models, we performed hyperparameter tuning using a randomised search for the number of trees and the maximum depth. The optimal parameters were 1600 trees and 30 levels for random forest, and 1500 trees and 6 levels for XGBoost.

We found that the highest accuracy achieved was 55.27% on average through the XGBoost algorithm. The result, however, indicated that our classifier performed barely above the performance of a ZeroR classifier, i.e., the level of chance (50.04% accuracy for our dataset). To ensure that the model performance scores were not obtained by chance, we performed a permutation test [78] on each of the classification algorithms. We found that all models except ZeroR achieved a p-value lower than 0.05, indicating that the employed models can give better predictions than the chance level with 95% confidence. Table 7 summarises each model's classification performance and the p-value of the permutation test.

7 DISCUSSION

To avoid information overload and effectively categorise the vast information available online, people often resort to mental shortcuts to make quick judgments about new information. These shortcuts can lead to a biased interpretation of that information and hence form what is called cognitive biases [102]. In the presented studies, we explored the indicators of cognitive biases in information consumption in two experiments, in which we exposed participants to ideologically polarising stimuli while collecting self-reports, behavioural, and physiological measures. Study 1 showed that some of our results were inconclusive in terms of physiological measures due to a lack of time gaps between subsequent stimuli. Hence, we were unable to isolate the effect of the stimuli on participants' opinion-related reactions. However, we found that participants spent more time with ideologically dissenting information but it was unclear whether this was due to the influence of their biased perception of the topic or whether some stimuli were more cognitively demanding in the way they were presented than others.

In Study 2, we addressed this limitation by redesigning the study and collecting not only behavioural (dwelling time and eye tracking) and physiological measures (EDA and hemodynamic responses) but also self-reports on topic interest and familiarity as they have been shown to influence the depth of information processing [63]. We ensured internal validity as participants demonstrated that their general tendency on a topic and their content-specific alignment (Q1) were consistent. Secondly, we introduced the use of inter-stimuli intervals (ISI) in Study 2. This allowed us to observe clearer physiological responses following stimulus exposure. Thirdly, we exposed participants to a greater range of opinion statements, thus increasing the external validity of the study.

In the remainder of this section, we discuss the outcomes of both studies focusing on the behavioural and physiological expressions of cognitive biases when viewing different opinions. We first discuss the effects of ideological congruency on dwelling time found in both studies. Subsequently, with Study 2 suggesting that topic interest influences the effects of ideological congruency, we discuss topic interest as a factor of biases. Lastly, we discuss the implications of building bias-aware systems, their feasibility, potential impact, as well as some ethical considerations.

7.1 Behavioural Expressions of Biases

In both studies, we observed that participants tended to spend more time with dissenting than opinion-confirming information. As in Study 2, the effects became stronger when considering participants with high interest in a specific topic. Our results support prior findings on selective exposure [37, 100]. Meanwhile, the results draw contrast to some prior works [68, 91], which stated that people tend to spend more time viewing confirmatory information.

Research on selective exposure has produced mixed results in terms of behavioural measures. With our studies showing different results from some of the existing literature, study designs may influence the behaviour of the participants and thus their behavioural expression of biases. Our study exposed participants to discrete pieces of information – i.e., participants read the stimulus contents one by one. Research by Garrett [37] and Taber and Lodge [100]

Table 5: Inferential statistics of Study 2's stimulus dwelling time and self-report measures. N and n denote the number of included participants and the number of included stimulus exposure, respectively. We denote **, *, and **** for significance levels of 0.05, 0.01, and 0.001, respectively.**

Measure Sample sizes	**General** ($N = 31, n = 1482$)	**Low interest** ($N = 19, n = 232$)	**High interest** ($N = 29, n = 900$)
Dwelling time	**D > C**** $\mathcal{F}(1, 30) = 5.713$ $\eta_p^2 = 0.160, p = 0.023$	n.s. $\mathcal{F}(1, 18) = 2.560$ $\eta_p^2 = 0.126, p = 0.0995$	**D > C***** $\mathcal{F}(1, 28) = 13.470$ $\eta_p^2 = 0.324, p = 0.001$
Q1	**C > D****** $\mathcal{F}(1, 30) = 203.481$ $\eta_p^2 = 0.871, p < 0.001$	**C > D****** $\mathcal{F}(1, 18) = 37.064$ $\eta_p^2 = 0.673, p < 0.001$	**C > D****** $\mathcal{F}(1, 28) = 135.994$ $\eta_p^2 = 0.829, p < 0.001$
Q2	**C > D****** $\mathcal{F}(1, 30) = 51.564$ $\eta_p^2 = 0.632, p < 0.001$	**C > D***** $\mathcal{F}(1, 18) = 11.628$ $\eta_p^2 = 0.392, p = 0.003$	**C > D****** $\mathcal{F}(1, 28) = 53.279$ $\eta_p^2 = 0.655, p < 0.001$
Q3	n.s. $\mathcal{F}(1, 30) = 3.375$ $\eta_p^2 = 0.101, p = 0.076$	**C > D***** $\mathcal{F}(1, 18) = 9.348$ $\eta_p^2 = 0.341, p = 0.006$	n.s. $\mathcal{F}(1, 28) = 3.493$ $\eta_p^2 = 0.0792, p = 0.131$

Table 6: Inferential statistics of Study 2's physiological measures. N, n, and w denote the number of included participants, the count of included stimulus exposure, and the window size, respectively. We denote **, *, and **** for significance levels of 0.05, 0.01, and 0.001, respectively.**

Measure	**General**	**Low interest**	**High interest**
SCL during EXP1 ($w = 10$ seconds)	n.s. ($N = 30, n = 754$) $\mathcal{F}(1, 29) = 4.242$ $\eta_p^2 = 0.127, p = 0.0502$	n.s. ($N = 18, n = 112$) $\mathcal{F}(1, 12) = 0.921$ $\eta_p^2 = 0.0713, p = 0.356$	n.s. ($N = 28, n = 448$) $\mathcal{F}(1, 26) = 0.987$ $\eta_p^2 = 0.0365, p = 0.329$
Δ_{all}[Hb] during EXP1 ($w = 2.5$ seconds)	n.s. ($N = 29, n = 1397$) $\mathcal{F}(1, 27) = 1.133$ $\eta_p^2 = 0.0402, p = 0.296$	**D > C**** ($\mathbf{N = 17, n = 200}$) $\mathcal{F}(1, 16) = 5.352$ $\eta_p^2 = 0.250, p = 0.034$	n.s. ($N = 27, n = 814$) $\mathcal{F}(1, 25) = 1.390$ $\eta_p^2 = 0.0526, p = 0.249$
Δ_{all}[Hb] during EXP1 ($w = 5$ seconds)	n.s. ($N = 29, n = 1228$) $\mathcal{F}(1, 27) = 1.865$ $\eta_p^2 = 0.0646, p = 0.183$	**D > C**** ($\mathbf{N = 17, n = 185}$) $\mathcal{F}(1, 16) = 4.607$ $\eta_p^2 = 0.223, p = 0.048$	n.s. ($N = 27, n = 777$) $\mathcal{F}(1, 25) = 2.605$ $\eta_p^2 = 0.0943, p = 0.119$

Table 7: The Evaluation Scores for Bias Classification.

	ZeroR	**LDA**	**SVM**	**Random Forest**	**XGBoost**
Mean Accuracy (SD)	50.04 (0)	50.96 (0.02)	50.20 (0.001)	54.39 (1.94)	55.27 (2.74)
p-value	1.00	0.047	0.047	0.047	0.047

followed a similar protocol to our studies and produced congruent results with our work. On the other hand, works by Marquart [68], for example, comprised a different study design where the participants freely navigated information on the screen while their dwelling time was tracked through area-specific fixation time.

Regarding reading effort (Q3), we find that individuals tended to spend more time but reported less effort reading information with dissenting stimuli. Our results align with the theory of epistemic monitoring by Richter [86, 87], which states that ideological dissenting information disrupts the fluency of information processing. As a result, individuals economise their cognitive resources by allocating them to attitude-consistent information. The theory may explain our findings that the prolonged reading time for dissenting statements resulted from the participant's reduced fluency in comprehending inconsistent information. Subsequently, less reading effort implies that individuals tend to save up their cognitive resources to process congruent information.

7.2 Physiological Expressions of Biases

We found that topic interest influenced the effects of opinion congruency on physiological responses. When considering individuals with low interest in a topic, we detected significant effects on the brain oxygenation levels during the start of the stimulus exposure. Our findings indicate that individuals tended to exhibit higher neural activation levels when processing ideologically dissenting information. This result is in line with prior research on cognitive dissonance [12, 49, 110], suggesting higher neural activation when facing attitude-challenging information.

Our results add to the existing literature on psychophysiology. To the best of our knowledge, this is the first study to obtain these findings using fNIRS sensors in the context of information exposure. While the physiological research on information consumption has been limited, it will be interesting to devise future studies that observe the interactions between individuals' involvement with a topic and their ideological tendency through more objective measures like physiological data.

In addition, we detected a trend that the skin conductance levels (SCL) were higher in dissenting stimuli. However, the result remained statistically inconclusive. Our results drew parallels to a study by Ploger et al. [82] which investigated cognitive dissonance through video media consumption. Similarly, albeit not statistically significant, Ploger et al. found that SCL tended to be higher when facing attitude-challenging information. We argue that our attitude-dissenting stimuli may induce dissonance arousal [111] – i.e., the physiological by-product of cognitive dissonance. Nonetheless, future research may focus on the potential of EDA in detecting the psycho-physiological effects of ideologically polarising information.

7.3 Topic Interest as a Factor of Bias

By varying the analysis on subgroups of high and low-interest individuals, we found that topic interest impacted the occurrence of cognitive biases. In sum, higher topic interest strengthened the effects of ideological congruency on dwelling time and the likelihood of sharing the stimulus content. Lower topic interest, on the other hand, positively influenced the effects of ideological congruency on the reading effort (Q3) and the physiological measures (skin conductance levels and brain oxygenation levels).

Our finding is in line with prior research on selective exposure [29, 53, 90, 95], which states that topic interest is one of the influencing factors for the selective exposure effect. Our result is also supported by the two-step model of processing conflicting information by Richter [86, 87]. The theory states that people tend to use the perceived plausibility of the information as a heuristic: they tend to save up their cognitive resources on attitude-consistent information and process the information based on their beliefs. On the other hand, individuals with relevant background knowledge tend to process it in an informed and balanced way.

Interestingly, we did not find significant effects of topic familiarity on the occurrence of biases. Instead, we detected such effects from topic *interest*. Since we did not explicitly assess prior knowledge, future studies should consider the effects of topic knowledge and familiarity on bias occurrence.

7.4 Towards Bias-Aware Systems

The studies presented are a first step to building *bias-aware systems*, i.e., computing systems that detect and take into account the presence of cognitive biases in users [24]. The notion of bias-aware systems parallels cognition-aware systems coined by Bulling and Zander [15] as they pick up and adjust to cognitive states but with a focus on biases and predispositions. Our results feed into system frameworks, such as Nussbaumer et al. [77], which collect user-system interaction data, learn to detect cognitive biases from such data, and help users reduce their biases by providing feedback from bias detection.

With multimodal data collected in our study, we employed a range of machine learning algorithms on the collected data to classify exposures that involved congruent information from those with dissenting information. As our models barely outperformed chance, the challenge remains to build a bias-aware system based on a well-performing classifier.

Our studies, however, show some promising results in connecting physiological and interaction data with users' innate opinions and attitudes. For the field of human-computer interaction, identifying these markers and designing experiments around eliciting and measuring cognitive biases is the first step towards researching and building bias-aware systems. In the context of recent societal impacts of computing systems, we envision more research in and broader use of measuring tools for the presence and effects of user biases in the evaluation of computing systems. To this end, we also release our study materials, including content and data collection apparatus as supplementary materials.

Being able to quantify the occurrence and the effect of cognitive biases will allow researchers to closely study the influence that user interfaces and algorithms have on opinion formation. Systems capable of identifying biases will subsequently enable work to address and mitigate their effects. Hence, interventions can be designed and tested to help users overcome cognitive fallacies as a result of their biases and encourage them to engage more critically with computing systems and information. In the current climate of misinformation, sensing users' attitudes and reactions towards potentially biased information can help design better information diets that help users break out of their filter bubbles, leave their silos of selected exposure and engage on a broader spectrum of ideas and opinions. Critical thinking and informed decision-making are critical for a healthy and diverse public discourse and have the potential to curtail the misinformation pandemic [60].

Finally, we would like to acknowledge the potential ethical implications of systems that sense biases, attitudes, and opinions. What can be used to identify and mitigate biases might as well be abused to reaffirm and steer people's beliefs, spread propaganda, and influence decision-making. The case of Cambridge Analytica has prominently demonstrated how people's attitudes can be derived from interaction data on social media platforms and used to influence opinion making [17]. Our research contributes to the systematic study of biases in the hopes that future work focuses on the demystification of how biases occur and what exacerbates or mitigates them.

8 LIMITATIONS

Despite Study 2 having addressed the main limitations of Study 1, there are a number of limitations we would like to discuss with regard to our study design and the interpretation of our findings.

First of all, our study design did not impose time constraints on each stimulus. Participants were free to spend as much time as they wanted with the stimulus until they clicked the next button. While this protocol allowed users to fully comprehend the stimuli materials, it introduced a number of limitations to the data analysis. The varying stimulus exposure time made it difficult to anticipate the temporal location of physiological reactions (i.e., the rise of oxygenation and skin conductance levels) regarding the stimuli. In addition, it was unclear how the reading motives of each participant affected their decision to end the stimulus exposure. Some participants may have tried to fully comprehend the material before clicking *next*, while others may have clicked next once they felt it unnecessary to further read the statement. We, however, took this into account by discarding data samples that lasted shorter than the analysis window size to mitigate the effects of shortened exposure.

Second, we were unable to assess the degree or strength of each stimulus's ideological tendency. As the primary source of our study stimuli, *ProCon.org* provides a collection of supporting and refuting information, which consists of opinions on different spectrums of attitude strength. In addition, our participants may perceive each stimulus individually, and therefore differently. During our post-study interviews, some participants reported they found some statements were not aligned with any particular ideological standpoint. The use of topic interest and familiarity as a subjective measure, however, helped us refine our analysis and isolate those cases where stronger tendencies may have been present.

The make-up of our study participants was also rather imbalanced in terms of ideologies. We found that most participants identified themselves as progressive, left-leaning, i.e., they mostly positioned themselves with the pro stances: they believed in man-made climate change and supported same-sex marriage. Since most participants were recruited from the university community, it was difficult for us to find people from conservative or right-leaning groups. While we tried to ensure that participants were exposed to an even number of congruent and dissenting stimuli, further studies with people with strong convictions and rather conservative and right-leaning attitudes are needed. For example, Knobloch-Westerwick et al. [55] found a greater leaning toward the US Republican party increased confirmation bias hinting that ideological alignment may have an effect on the creation and experience of biases.

In terms of physiological sensors, we used the Empatica E4 wristband to measure EDA. While the device is compact and unobtrusive, recent research has expressed concerns that E4-generated EDA signals are prone to motion artefacts and measurement noise and may not be as reliable as laboratory-grade devices [4, 10, 73]. Moreover, we only collected the hemodynamic responses using fNIRS from the forehead region. Although the placement of the sensor pad allowed unobtrusive data collection, it limited our observations to neural activities beyond the forehead area that is not covered by hair. Future works may investigate the hemodynamic activity from full-head fNIRS.

We used self-report ratings as pre-study questions to gauge participants' ideological inclination on, interest in, and familiarity with each topic, as well as in-study surveys to improve the internal validity of our study. While self-reports are convenient tools to collect information, they can be confounded by a range of factors, notably, memory, self-presentation [96] and preference falsification [59]. Moreover, written questions are susceptible to misinterpretation. For instance, some participants reported that they understood the question "*How much effort did you put in reading this statement*" (Q3) as the amount of cognitive resources spent on reading the statement, while some reported that they replied to Q3 by providing the degree of how well they understood the statement.

Moreover, we only used single-item questions to represent each of the self-report measures. This limits the external validity of our study since we did not use standardised, established subjective measures. For example, we did not examine the reading effort (Q3) using a well-established NASA-TLX questionnaire out of concern for participants' time and fatigue levels. Moreover, we did not assess the participant's general ideological alignment using, for instance, the Wilson-Patterson conservatism scales or performing an implicit association test [24]. Such tools may allow future research to uncover a more nuanced relationship between the strength of ideological conviction and individuals' experience of biases.

9 CONCLUSION

Biases as cognitive shortcuts help people cope with the vast amounts of online information but can also trap us in one-sided exposure and filter bubbles that reaffirm our existing beliefs. To study the occurrence and effects of biases on information consumption behaviour and decision-making, we set out to explore indicators for the occurrence of biases in physiological and interaction data. In this paper, we presented two experiments, in which we exposed users to opinionated statements on polarising topics while collecting physiological, behavioural, and interaction data. Our stimuli samples stated opinions that were either congruent or dissenting with participants' attitudes. We found that participants tended to generally spend more time processing statements that were incongruent with their own opinion. We further observed higher neural activity as indicated by certain brain regions' blood oxygenation levels when participants were facing ideologically dissenting attitudes while having expressed relatively low interest in that topic. Our results demonstrate the existence of behavioural and physiological differences in the expression of congruency between people's innate opinions and ideologically tainted information, a first step towards building classifiers to detect cognitive biases.

Our study design and findings pave the way for future research in understanding the occurrence of cognitive biases with the goal of detecting them and quantifying their effects. The ability to equip systems with bias-awareness allows HCI researchers to study the role that design, algorithms, and content elements play in mitigating or exacerbating user biases.

ACKNOWLEDGMENTS

We thank the participants of our studies.

REFERENCES

[1] Robert Ackland and Ann Evans. 2017. Using the web to examine the evolution of the abortion debate in Australia, 2005-2015. In *The Web as History*, Niels Brügger and Ralph Schroeder (Eds.). University College London, Gower Street, London WC1E 6BT, 159–189.

[2] Haleh Aghajani, Marc Garbey, and Ahmet Omurtag. 2017. Measuring Mental Workload with EEG+fNIRS. *Frontiers in Human Neuroscience* 11 (2017). https://doi.org/10.3389/fnhum.2017.00359

[3] Faisal Alatawi, Lu Cheng, Anique Tahir, Mansooreh Karami, Bohan Jiang, Tyler Black, and Huan Liu. 2021. A Survey on Echo Chambers on Social Media: Description, Detection and Mitigation. *arXiv preprint arXiv:2112.05084* (2021). https://arxiv.org/abs/2112.05084

[4] Ebrahim Babaei, Benjamin Tag, Tilman Dingler, and Eduardo Velloso. 2021. A Critique of Electrodermal Activity Practices at CHI. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 177, 14 pages. https://doi.org/10.1145/3411764.3445370

[5] Ricardo Baeza-Yates. 2018. Bias on the Web. *Commun. ACM* 61, 6 (May 2018), 54–61. https://doi.org/10.1145/3209581

[6] Wesley B. Baker, Ashwin B. Parthasarathy, David R. Busch, Rickson C. Mesquita, Joel H. Greenberg, and A. G. Yodh. 2014. Modified Beer-Lambert law for blood flow. *Biomed. Opt. Express* 5, 11 (Nov 2014), 4053–4075. https://doi.org/10.1364/BOE.5.004053

[7] Eytan Bakshy, Solomon Messing, and Lada A. Adamic. 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science* 348, 6239 (2015), 1130–1132. https://doi.org/10.1126/science.aaa1160 arXiv:https://www.science.org/doi/pdf/10.1126/science.aaa1160

[8] Mathias Benedek and Christian Kaernbach. 2010. A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods* 190, 1 (2010), 80–91. https://doi.org/10.1016/j.jneumeth.2010.04.028

[9] Alexander Benlian. 2015. Web Personalization Cues and Their Differential Effects on User Assessments of Website Value. *Journal of Management Information Systems* 32, 1 (2015), 225–260. https://doi.org/10.1080/07421222.2015.1029394 arXiv:https://doi.org/10.1080/07421222.2015.1029394

[10] Adrian Borrego, Jorge Latorre, Mariano Alcañiz, and Roberto Llorens. 2019. Reliability of the Empatica E4 wristband to measure electrodermal activity to emotional stimuli. In *2019 International Conference on Virtual Rehabilitation (ICVR)*. 1–2. https://doi.org/10.1109/ICVR46560.2019.8994546

[11] Wolfram Boucsein. 2012. *Electrodermal activity*. Springer Science & Business Media. https://doi.org/10.1007/978-1-4614-1126-0

[12] Ming M. Boyer. 2021. Aroused Argumentation: How the News Exacerbates Motivated Reasoning. *The International Journal of Press/Politics* (2021), 19401612211010577. https://doi.org/10.1177/19401612211010577 arXiv:https://doi.org/10.1177/19401612211010577

[13] Leo Breiman. 2001. Random forests. *Machine learning* 45, 1 (2001), 5–32.

[14] Andreas Bulling and Thorsten O. Zander. 2014. Cognition-Aware Computing. *IEEE Pervasive Computing* 13, 3 (2014), 80–83. https://doi.org/10.1109/MPRV.2014.42

[15] Andreas Bulling and Thorsten O. Zander. 2014. Cognition-Aware Computing. *IEEE Pervasive Computing* 13, 3 (2014), 80–83. https://doi.org/10.1109/MPRV.2014.42

[16] Richard B. Buxton, Kâmil Uludağ, David J. Dubowitz, and Thomas T. Liu. 2004. Modeling the hemodynamic response to brain activation. *NeuroImage* 23 (2004), S220–S233. https://doi.org/10.1016/j.neuroimage.2004.07.013 Mathematics in Brain Imaging.

[17] Carole Cadwalladr. 2017. The great British Brexit robbery: how our democracy was hijacked. *The Guardian* 7 (2017).

[18] Filipe R Campante and Daniel A Hojman. 2013. Media and polarization: Evidence from the introduction of broadcast TV in the United States. *Journal of Public Economics* 100 (2013), 79–92.

[19] Pew Reseach Center. 2014. Political polarization in the american public. *Retrieved September* 2 (2014), 2019.

[20] Russ Clay, Jessica M. Barber, and Natalie J. Shook. 2013. Techniques for Measuring Selective Exposure: A Critical Review. *Communication Methods and Measures* 7, 3-4 (2013), 147–171. https://doi.org/10.1080/19312458.2013.813925 arXiv:https://doi.org/10.1080/19312458.2013.813925

[21] Chad Cooper. 2012. The immigration debate in Australia: from federation to World War One. (2012).

[22] Michael E Dawson, Anne M Schell, and Diane L Filion. 2017. The electrodermal system. (2017).

[23] Kristin Demetrious. 2019. 'Energy wars': Global PR and public debate in the 21st century. *Public Relations Inquiry* 8, 1 (2019), 7–22. https://doi.org/10.1177/2046147X18804283 arXiv:https://doi.org/10.1177/2046147X18804283

[24] Tilman Dingler, Benjamin Tag, David A. Eccles, Niels van Berkel, and Vassilis Kostakos. 2022. Method for Appropriating the Brief Implicit Association Test to Elicit Biases in Users. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association

for Computing Machinery, New York, NY, USA, Article 243, 16 pages. https://doi.org/10.1145/3491102.3517570

[25] Tilman Dingler, Benjamin Tag, Evangelos Karapanos, Koichi Kise, and Andreas Dengel. 2020. Workshop on Detection and Design for Cognitive Biases in People and Computing Systems. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3334480.3375159

[26] Tilman Dingler, Benjamin Tag, Philipp Lorenz-Spreen, Andrew W. Vargo, Simon Knight, and Stephan Lewandowsky. 2021. Workshop on Technologies to Support Critical Thinking in an Age of Misinformation. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 105, 5 pages. https://doi.org/10.1145/3411763.3441350

[27] Tilman Dingler, Benjamin Tag, and Andrew Vargo. 2022. Technologies to Support Critical Thinking in an Age of Misinformation (Dagstuhl Seminar 22172). *Dagstuhl Reports* 12, 4 (2022), 72–95. https://doi.org/10.4230/DagRep.12.4.72

[28] Gabriele Donzelli, Giacomo Palomba, Ileana Federigi, Francesco Aquino, Lorenzo Cioni, Marco Verani, Annalaura Carducci, and Pierluigi Lopalco. 2018. Misinformation on vaccination: A quantitative analysis of YouTube videos. *Human Vaccines & Immunotherapeutics* 14, 7 (2018), 1654–1659. https://doi.org/10.1080/21645515.2018.1454572 arXiv:https://doi.org/10.1080/21645515.2018.1454572 PMID: 29553872.

[29] Lauren Feldman, Natalie Jomini Stroud, Bruce Bimber, and Magdalena Wojcieszak. 2013. Assessing Selective Exposure in Experiments: The Implications of Different Methodological Choices. *Communication Methods and Measures* 7, 3-4 (2013), 172–194. https://doi.org/10.1080/19312458.2013.813923 arXiv:https://doi.org/10.1080/19312458.2013.813923

[30] Leon Festinger. 1957. *A theory of cognitive dissonance*. Vol. 2. Stanford university press.

[31] Frank Fishburn, Megan Norr, Andrei Medvedev, and Chandan Vaidya. 2014. Sensitivity of fNIRS to cognitive state and load. *Frontiers in Human Neuroscience* 8 (2014). https://doi.org/10.3389/fnhum.2014.00076

[32] Frank A. Fishburn, Ruth S. Ludlum, Chandan J. Vaidya, and Andrei V. Medvedev. 2019. Temporal Derivative Distribution Repair (TDDR): A motion correction method for fNIRS. *NeuroImage* 184 (2019), 171–179. https://doi.org/10.1016/j.neuroimage.2018.09.025

[33] Ian Flaherty and Jennifer Wilkinson. 2020. Marriage equality in Australia: The 'no' vote and symbolic violence. *Journal of Sociology* 56, 4 (2020), 664–674. https://doi.org/10.1177/1440783320969882 arXiv:https://doi.org/10.1177/1440783320969882

[34] Rudolf Flesch. 1979. How to write plain English. *University of Canterbury. Available at http://www. mang. canterbury. ac. nz/writing_guide/writing/flesch. shtml.[Retrieved 5 February 2016]* (1979).

[35] Society for Psychophysiological Research Ad Hoc Committee on Electrodermal Measures. 2012. Publication recommendations for electrodermal measurements. *Psychophysiology* 49, 8 (2012), 1017–1034. https://doi.org/10.1111/j.1469-8986.2012.01384.x arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-8986.2012.01384.x

[36] Jerome H Friedman. 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics* (2001), 1189–1232.

[37] R. Kelly Garrett. 2009. Echo chambers online?: Politically motivated selective exposure among Internet news users. *Journal of Computer-Mediated Communication* 14, 2 (01 2009), 265–285. https://doi.org/10.1111/j.1083-6101.2009.01440.x arXiv:https://academic.oup.com/jcmc/article-pdf/14/2/265/21491614/jjcmcom0265.pdf

[38] Eija Haapalainen, SeungJun Kim, Jodi F. Forlizzi, and Anind K. Dey. 2010. Psycho-Physiological Measures for Assessing Cognitive Load. In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing* (Copenhagen, Denmark) *(UbiComp '10)*. Association for Computing Machinery, New York, NY, USA, 301–310. https://doi.org/10.1145/1864349.1864395

[39] Martie G Haselton, Daniel Nettle, and Damian R Murray. 2015. The evolution of cognitive bias. *The handbook of evolutionary psychology* (2015), 1–20.

[40] Dominic Heger, Reinhard Mutter, Christian Herff, Felix Putze, and Tanja Schultz. 2013. Continuous Recognition of Affective States by Functional Near Infrared Spectroscopy Signals. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. 832–837. https://doi.org/10.1109/ACII.2013.156

[41] Vincent F. Hendricks and Mads Vestergaard. 2019. *Reality Lost*. Number 978-3-030-00813-0 in Springer Books. Springer. https://doi.org/10.1007/978-3-030-00813-0

[42] Martin Hilbert. 2012. Toward a synthesis of cognitive biases: how noisy information processing can bias human decision making. *Psychological bulletin* 138, 2 (March 2012), 211—237. https://doi.org/10.1037/a0025940

[43] Leanne Hirshfield, Phil Bobko, Alex Barelka, Natalie Sommer, and Senem Velipasalar. 2019. Toward Interfaces that Help Users Identify Misinformation Online: Using fNIRS to Measure Suspicion. *Augmented Human Research* 4, 1 (Dec. 2019), 1. https://doi.org/10.1007/s41133-019-0011-8

[44] J.M. Inhofe. 2012. *The Greatest Hoax: How the Global Warming Conspiracy Threatens Your Future*. WND Books. https://books.google.co.th/books?id=Ry7SygAACAAJ

[45] Kathleen Hall Jamieson and Joseph N Cappella. 2008. *Echo chamber: Rush Limbaugh and the conservative media establishment*. Oxford University Press.

[46] Dietmar Jannach and Michael Jugovac. 2019. Measuring the Business Value of Recommender Systems. *ACM Trans. Manage. Inf. Syst.* 10, 4, Article 16 (Dec 2019), 23 pages. https://doi.org/10.1145/3370082

[47] E Jonas, S Schulz-Hardt, D Frey, and N Thelen. 2001. Confirmation bias in sequential information search after preliminary decisions: an expansion of dissonance theoretical research on selective exposure to information. *Journal of personality and social psychology* 80, 4 (April 2001), 557—571. https://doi.org/10.1037//0022-3514.80.4.557

[48] Daniel Kahneman. 2011. *Thinking, fast and slow*. Macmillan.

[49] Jonas T Kaplan, Sarah I Gimbel, and Sam Harris. 2016. Neural correlates of maintaining one's political beliefs in the face of counterevidence. *Scientific reports* 6 (December 2016), 39589. https://doi.org/10.1038/srep39589

[50] Toshinori Kato, Atsushi Kamei, Sachio Takashima, and Takeo Ozaki. 1993. Human Visual Cortical Function during Photic Stimulation Monitoring by Means of near-Infrared Spectroscopy. *Journal of Cerebral Blood Flow & Metabolism* 13, 3 (1993), 516–520. https://doi.org/10.1038/jcbfm.1993.66 arXiv:https://doi.org/10.1038/jcbfm.1993.66 PMID: 8478409.

[51] Joseph T Klapper. 1960. The effects of mass communication. (1960).

[52] Jan Kleinnijenhuis. 2008. *Negativity*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781405186407.wbiecn005 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781405186407.wbiecn005

[53] Silvia Knobloch-Westerwick and Jingbo Meng. 2009. Looking the Other Way: Selective Exposure to Attitude-Consistent and Counterattitudinal Political Information. *Communication Research* 36, 3 (2009), 426–448. https://doi.org/10.1177/0093650209333030 arXiv:https://doi.org/10.1177/0093650209333030

[54] Silvia Knobloch-Westerwick and Jingbo Meng. 2011. Reinforcement of the Political Self Through Selective Exposure to Political Messages. *Journal of Communication* 61, 2 (04 2011), 349–368. https://doi.org/10.1111/j.1460-2466.2011.01543.x arXiv:https://academic.oup.com/joc/article-pdf/61/2/349/22324723/jjnlcom0349.pdf

[55] Silvia Knobloch-Westerwick, Cornelia Mothes, and Nick Polavin. 2020. Confirmation Bias, Ingroup Bias, and Negativity Bias in Selective Exposure to Political Information. *Communication Research* 47, 1 (2020), 104–124. https://doi.org/10.1177/0093650217719596 arXiv:https://doi.org/10.1177/0093650217719596

[56] Asher Koriat, Sarah Lichtenstein, and Baruch Fischhoff. 1980. Reasons for confidence. *Journal of Experimental Psychology: Human learning and memory* 6, 2 (1980), 107.

[57] Arthur E Kramer. 1990. Physiological Metrics of Mental Workload: A Review of Recent Progress.

[58] Katherine Kricorian, Rachel Civen, and Ozlem Equils. 2022. COVID-19 vaccine hesitancy: misinformation and perceptions of vaccine safety. *Human Vaccines & Immunotherapeutics* 18, 1 (2022), 1950504. https://doi.org/10.1080/21645515.2021.1950504 arXiv:https://doi.org/10.1080/21645515.2021.1950504 PMID: 34325612.

[59] Timur Kuran. 1997. *Private truths, public lies: The social consequences of preference falsification*. Harvard University Press.

[60] David MJ Lazer, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer, Miriam J Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, et al. 2018. The science of fake news. *Science* 359, 6380 (2018), 1094–1096.

[61] Thomas J. Leeper and Rune Slothuus. 2014. Political Parties, Motivated Reasoning, and Public Opinion Formation. *Political Psychology* 35, S1 (2014), 129–156. https://doi.org/10.1111/pops.12164 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/pops.12164

[62] Gilat Levy and Ronny Razin. 2019. Echo Chambers and Their Effects on Economic and Political Outcomes. *Annual Review of Economics* 11, 1 (2019), 303–328. https://doi.org/10.1146/annurev-economics-080218-030343 arXiv:https://doi.org/10.1146/annurev-economics-080218-030343

[63] Stephan Lewandowsky, Ullrich K. H. Ecker, Colleen M. Seifert, Norbert Schwarz, and John Cook. 2012. Misinformation and Its Correction: Continued Influence and Successful Debiasing. *Psychological Science in the Public Interest* 13, 3 (2012), 106–131. https://doi.org/10.1177/1529100612451018 arXiv:https://doi.org/10.1177/1529100612451018 PMID: 26173286.

[64] Stephan Lewandowsky, Gilles E Gignac, and Klaus Oberauer. 2013. The role of conspiracist ideation and worldviews in predicting rejection of science. *PloS one* 8, 10 (2013), e75637.

[65] Scott O. Lilienfeld, Rachel Ammirati, and Kristin Landfield. 2009. Giving Debiasing Away: Can Psychological Research on Correcting Cognitive Errors Promote Human Welfare? *Perspectives on Psychological Science* 4, 4 (2009), 390–398. https://doi.org/10.1111/j.1745-6924.2009.01144.x arXiv:https://doi.org/10.1111/j.1745-6924.2009.01144.x PMID: 26158987.

[66] S. Lloyd-Fox, A. Blasi, and C.E. Elwell. 2010. Illuminating the developing brain: The past, present and future of functional near infrared spectroscopy. *Neuroscience & Biobehavioral Reviews* 34, 3 (2010), 269–284. https://doi.org/10.1016/j.

neubiorev.2009.07.008
[67] Regan L. Mandryk, M. Stella Atkins, and Kori M. Inkpen. 2006. A Continuous and Objective Evaluation of Emotional Experience with Interactive Play Environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Montréal, Québec, Canada) *(CHI '06)*. Association for Computing Machinery, New York, NY, USA, 1027–1036. https://doi.org/10.1145/1124772.1124926
[68] Franziska Marquart. 2016. Selective Exposure in the Context of Political Advertising: A Behavioral Approach Using Eye-Tracking Methodology. (2016), 20. https://ijoc.org/index.php/ijoc/article/view/4415
[69] John Martindale, John Mayhew, Jason Berwick, Myles Jones, Chris Martin, Dave Johnston, Peter Redgrave, and Ying Zheng. 2003. The Hemodynamic Impulse Response to a Single Neural Event. *Journal of Cerebral Blood Flow & Metabolism* 23, 5 (2003), 546–555. https://doi.org/10.1097/01.WCB.0000058871.46954.2B arXiv:https://doi.org/10.1097/01.WCB.0000058871.46954.2B PMID: 12771569.
[70] Daina Mazutis and Anna Eckardt. 2017. Sleepwalking into Catastrophe: Cognitive Biases and Corporate Climate Change Inertia. *California Management Review* 59, 3 (2017), 74–108. https://doi.org/10.1177/0008125617707974 arXiv:https://doi.org/10.1177/0008125617707974
[71] Jennifer McCoy, Tahmina Rahman, and Murat Somer. 2018. Polarization and the global crisis of democracy: Common patterns, dynamics, and pernicious consequences for democratic polities. *American Behavioral Scientist* 62, 1 (2018), 16–42.
[72] Miriam J. Metzger, Ethan H. Hartsell, and Andrew J. Flanagin. 2020. Cognitive Dissonance or Credibility? A Comparison of Two Theoretical Explanations for Selective Exposure to Partisan News. *Communication Research* 47, 1 (2020), 3–28. https://doi.org/10.1177/0093650215613136 arXiv:https://doi.org/10.1177/0093650215613136
[73] Nir Milstein and Ilanit Gordon. 2020. Validating measures of electrodermal activity and heart rate variability derived from the empatica E4 utilized in research settings that involve interactive dyadic states. *Frontiers in Behavioral Neuroscience* 14 (2020), 148. https://doi.org/10.3389/fnbeh.2020.00148
[74] Randall K. Minas, Robert F. Potter, Alan R. Dennis, Valerie Bartelt, and Soyoung Bae. 2014. Putting on the Thinking Cap: Using NeuroIS to Understand Information Processing Biases in Virtual Teams. *Journal of Management Information Systems* 30, 4 (2014), 49–82. https://doi.org/10.2753/MIS0742-1222300403 arXiv:https://doi.org/10.2753/MIS0742-1222300403
[75] Raymond S. Nickerson. 1998. Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology* 2, 2 (1998), 175–220. https://doi.org/10.1037/1089-2680.2.2.175 arXiv:https://doi.org/10.1037/1089-2680.2.2.175
[76] Fateme Nikseresht, Runze Yan, Rachel Lew, Yingzheng Liu, Rose M. Sebastian, and Afsaneh Doryab. 2021. Detection of Racial Bias from Physiological Responses. In *Advances in Usability, User Experience, Wearable and Assistive Technology*, Tareq Z. Ahram and Christianne S. Falcão (Eds.). Vol. 275. Springer International Publishing, Cham, 59–66. https://doi.org/10.1007/978-3-030-80091-8_8 Series Title: Lecture Notes in Networks and Systems.
[77] Alexander Nussbaumer, Katrien Verbert, Eva-Catherine Hillemann, Michael A. Bedek, and Dietrich Albert. 2016. A Framework for Cognitive Bias Detection and Feedback in a Visual Analytics Environment. In *2016 European Intelligence and Security Informatics Conference (EISIC)*. IEEE, Uppsala, Sweden, 148–151. https://doi.org/10.1109/EISIC.2016.038
[78] Markus Ojala and Gemma C Garriga. 2010. Permutation tests for studying classifier performance. *Journal of machine learning research* 11, 6 (2010).
[79] Anneli Olsen. 2012. The Tobii I-VT fixation filter. *Tobii Technology* 21 (2012), 4–19. https://www.tobiipro.com/siteassets/tobii-pro/learn-and-support/analyze/how-do-we-classify-eye-movements/tobii-pro-i-vt-fixation-filter.pdf
[80] Myrto Pantazi, Scott Hale, and Olivier Klein. 2021. Social and Cognitive Aspects of the Vulnerability to Political Misinformation. *Political Psychology* (2021). https://doi.org/10.1111/pops.12797
[81] Eli Pariser. 2011. *The filter bubble: What the Internet is hiding from you*. Penguin UK.
[82] Gavin W. Ploger, Johnanna Dunaway, Patrick Fournier, and Stuart Soroka. 2021. The psychophysiological correlates of cognitive dissonance. *Politics and the Life Sciences* 40, 2 (2021), 202–212. https://doi.org/10.1017/pls.2021.15
[83] Gregory A Poland, Robert M Jacobson, et al. 2011. The age-old struggle against the antivaccinationists. *N Engl J Med* 364, 2 (2011), 97–9.
[84] Louise Richardson-Self. 2018. Same-Sex Marriage and the "No" Campaign. *Humanities Australia* 9 (2018), 32–39.
[85] Tobias Richter. 2011. *Cognitive Flexibility and Epistemic Validation in Learning from Multiple Texts*. Springer Netherlands, Dordrecht, 125–140. https://doi.org/10.1007/978-94-007-1793-0_7
[86] Tobias Richter. 2015. Validation and Comprehension of Text Information: Two Sides of the Same Coin. *Discourse Processes* 52, 5-6 (2015), 337–355. https://doi.org/10.1080/0163853X.2015.1025665 arXiv:https://doi.org/10.1080/0163853X.2015.1025665
[87] Tobias Richter and Johanna Maier. 2017. Comprehension of Multiple Documents With Conflicting Information: A Two-Step Model of Validation. *Educational Psychologist* 52, 3 (2017), 148–166. https://doi.org/10.1080/00461520.2017.1322968 arXiv:https://doi.org/10.1080/00461520.2017.1322968
[88] Michael Salter. 2016. Men's Rights or Men's Needs? Anti-Feminism in Australian Men's Health Promotion. *Canadian Journal of Women and the Law* 28, 1 (2016), 69–90.
[89] Emily Saltz, Claire R Leibowicz, and Claire Wardle. 2021. *Encounters with Visual Misinformation and Labels Across Platforms: An Interview and Diary Study to Inform Ecosystem Approaches to Misinformation Interventions*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411763.3451807
[90] Josephine B. Schmitt, Christina A. Debbelt, and Frank M. Schneider. 2018. Too much information? Predictors of information overload in the context of online news exposure. *Information, Communication & Society* 21, 8 (2018), 1151–1167. https://doi.org/10.1080/1369118X.2017.1305427 arXiv:https://doi.org/10.1080/1369118X.2017.1305427
[91] Desirée Schmuck, Miriam Tribastone, Jörg Matthes, Franziska Marquart, and Eva Maria Bergel. 2020. Avoiding the Other Side?: An Eye-Tracking Study of Selective Exposure and Selective Avoidance Effects in Response to Political Advertising. *Journal of Media Psychology* 32, 3 (July 2020), 158–164. https://doi.org/10.1027/1864-1105/a000265
[92] Felix Scholkmann, Stefan Kleiser, Andreas Jaakko Metz, Raphael Zimmermann, Juan Mata Pavia, Ursula Wolf, and Martin Wolf. 2014. A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. *NeuroImage* 85 (2014), 6–27. https://doi.org/10.1016/j.neuroimage.2013.05.004 Celebrating 20 Years of Functional Near Infrared Spectroscopy (fNIRS).
[93] Cornelia Setz, Bert Arnrich, Johannes Schumm, Roberto La Marca, Gerhard Tröster, and Ulrike Ehlert. 2010. Discriminating Stress From Cognitive Load Using a Wearable EDA Device. *IEEE Transactions on Information Technology in Biomedicine* 14, 2 (2010), 410–417. https://doi.org/10.1109/TITB.2009.2036164
[94] Murray Singer. 2013. Validation in Reading Comprehension. *Current Directions in Psychological Science* 22, 5 (2013), 361–366. https://doi.org/10.1177/0963721413495236 arXiv:https://doi.org/10.1177/0963721413495236
[95] Steven M. Smith, Leandre R. Fabrigar, and Meghan E. Norris. 2008. Reflecting on Six Decades of Selective Exposure Research: Progress, Challenges, and Opportunities. *Social and Personality Psychology Compass* 2, 1 (2008), 464–493. https://doi.org/10.1111/j.1751-9004.2007.00060.x arXiv:https://compass.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1751-9004.2007.00060.x
[96] Natalie Jomini Stroud. 2008. Media Use and Political Predispositions: Revisiting the Concept of Selective Exposure. *Political Behavior* 30, 3 (2008), 341–366. http://www.jstor.org/stable/40213321
[97] Natalie Jomini Stroud. 2010. Polarization and Partisan Selective Exposure. *Journal of Communication* 60, 3 (08 2010), 556–576. https://doi.org/10.1111/j.1460-2466.2010.01497.x arXiv:https://academic.oup.com/joc/article-pdf/60/3/556/22324536/jjnlcom0556.pdf
[98] Jill M Swirsky and David Jason Angelone. 2016. Equality, empowerment, and choice: what does feminism mean to contemporary women? *Journal of Gender Studies* 25, 4 (2016), 445–460.
[99] Michael Sülflow, Svenja Schäfer, and Stephan Winter. 2019. Selective attention in the news feed: An eye-tracking study on the perception and selection of political news posts on Facebook. *New Media & Society* 21, 1 (Jan. 2019), 168–190. https://doi.org/10.1177/1461444818791520
[100] Charles S. Taber and Milton Lodge. 2006. Motivated Skepticism in the Evaluation of Political Beliefs. *American Journal of Political Science* 50, 3 (2006), 755–769. https://doi.org/10.1111/j.1540-5907.2006.00214.x arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-5907.2006.00214.x
[101] Animals Australia Team. 2019. Surge in Aussies eating vegetarian continues. https://animalsaustralia.org/latest-news/study-shows-surge-in-aussies-eating-veg/
[102] Amos Tversky and Daniel Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *science* 185, 4157 (1974), 1124–1131.
[103] Nicholas A. Valentino, Antoine J. Banks, Vincent L. Hutchings, and Anne K. Davis. 2009. Selective Exposure in the Internet Age: The Interaction between Anxiety and Information Utility. *Political Psychology* 30, 4 (2009), 591–613. http://www.jstor.org/stable/25655419
[104] Boris M. Velichkovsky and John Paulin Hansen. 1996. New Technological Windows into Mind: There is More in Eyes and Brains for Human-Computer Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, British Columbia, Canada) *(CHI '96)*. Association for Computing Machinery, New York, NY, USA, 496–503. https://doi.org/10.1145/238386.238619
[105] Micah N. Villarreal, Alexander J. Kamrud, and Brett J. Borghetti. 2019. Confirmation Bias Estimation from Electroencephalography with Machine Learning. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 63, 1 (2019), 73–77. https://doi.org/10.1177/1071181319631208 arXiv:https://doi.org/10.1177/1071181319631208

[106] Nico Voigtländer and Hans-Joachim Voth. 2015. Nazi indoctrination and anti-Semitic beliefs in Germany. *Proceedings of the National Academy of Sciences* 112, 26 (2015), 7931–7936. https://doi.org/10.1073/pnas.1414822112 arXiv:https://www.pnas.org/doi/pdf/10.1073/pnas.1414822112

[107] Michelle W. Voss. 2016. Chapter 9 - The Chronic Exercise–Cognition Interaction: fMRI Research. In *Exercise-Cognition Interaction*, Terry McMorris (Ed.). Academic Press, San Diego, 187–209. https://doi.org/10.1016/B978-0-12-800778-5.00009-8

[108] Kimberlee Weaver, Stephen M Garcia, Norbert Schwarz, and Dale T Miller. 2007. Inferring the popularity of an opinion from its familiarity: a repetitive voice can sound like a chorus. *Journal of personality and social psychology* 92, 5 (2007), 821.

[109] Chris Wells, Katherine J Cramer, Michael W Wagner, German Alvarez, Lewis A Friedland, Dhavan V Shah, Leticia Bode, Stephanie Edgerly, Itay Gabay, and Charles Franklin. 2017. When we stop talking politics: The maintenance and closing of conversation in contentious times. *Journal of Communication* 67, 1 (2017), 131–157.

[110] Drew Westen, Pavel S. Blagov, Keith Harenski, Clint Kilts, and Stephan Hamann. 2006. Neural Bases of Motivated Reasoning: An fMRI Study of Emotional Constraints on Partisan Political Judgment in the 2004 U.S. Presidential Election. *Journal of Cognitive Neuroscience* 18, 11 (11 2006), 1947–1958. https://doi.org/10.1162/jocn.2006.18.11.1947 arXiv:https://direct.mit.edu/jocn/article-pdf/18/11/1947/1935646/jocn.2006.18.11.1947.pdf

[111] Mark P Zanna and Joel Cooper. 1974. Dissonance and the pill: an attribution approach to studying the arousal properties of dissonance. *Journal of personality and social psychology* 29, 5 (1974), 703. https://doi.org/10.1037/h0036651

[112] Arne Zillich and Lars Guenther. 2021. Selective Exposure to Information on the Internet: Measuring Cognitive Dissonance and Selective Exposure with Eye-Tracking. *International Journal of Communication* 15, 0 (2021). https://ijoc.org/index.php/ijoc/article/view/14184

[113] Fabiana Zollo and Walter Quattrociocchi. 2018. *Misinformation Spreading on Facebook*. Springer International Publishing, Cham, 177–196. https://doi.org/10.1007/978-3-319-77332-2_10