Knowledge-Seeking Reflects and Shapes Well-Being

Christopher A. Kelly[^123] & Tali Sharot[^123]

Affective Brain Lab
[1]Department of Experimental Psychology, University College London, London, UK
[2]Max Planck University College London Centre for Computational Psychiatry and Ageing Research, London, UK
[3]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, USA

[^]Corresponding authors christopher.kelly.16@ucl.ac.uk, t.sharot@ucl.ac.uk

Abstract

Humans are spending an increasing amount of time searching for knowledge online. It is thus imperative to examine whether and how this activity impacts well-being. Here, we test the hypothesis that the affective properties of the knowledge sought effect well-being, which in turn alters knowledge-seeking, forming a self-reinforcing loop. To that end, we quantified the affective properties of text in webpages participants (N = 947) chose to browse and related these to their well-being. We find that browsing more negative information was associated with worse mental-health and mood. By manipulating the webpages browsed and measuring mood and vice versa, we reveal that the relationship is causal and bi-directional. Moreover, when participants were made aware of the affective nature of webpages before browsing, they choose to access more positive and less negative webpages. These findings provide a potential method for assessing and enhancing human welfare in the digital age.

Introduction

Determining which factors are associated with well-being has been a key pursuit of scientists, policymakers, and the general public. Research has linked well-being to various elements such as social relationships (Pieh et al., 2020; Ertel et al., 2009; Robles & Kiecolt-Glaser, 2003), exercise (Marconcin et al., 2022; Peluso et al., 2005), and wealth (Ettman et al., 2022; Pollack et al., 2007). In recent years, as people spend more time online, the need to investigate the relationship between online activity and well-being has become imperative (DataReportal, 2022). This will inform the development of online tools to enhance well-being and could provide real-time assessment of it.

One of the most frequent online activities is knowledge-seeking. Interestingly, what people choose to know varies vastly from one individual to the next (Kelly & Sharot, 2021; Kobayashi et al., 2019; Sunstein, 2019). These variations may provide important clues about an individual's inner cognitive and affective state. In particular, we have theorized that the affective properties of the knowledge people consume from self-guided searches may reflect their well-being (Sharot & Sunstein, 2020). For instance, experiencing a negative affective state may drive individuals to search for knowledge with a similar sentiment, resulting in the consumption of negatively charged information, which could in turn exasperate one's negative affective state. Hence, the relationship between well-being and knowledge-seeking may be reciprocal and form a self-reinforcing loop.

If indeed a bi-directional relationship exists between the type of information that is consumed from self-guided online searches and well-being, it would have significant theoretical and practical implications. As humans constantly engage in knowledge-seeking online, there is a unique opportunity to harness this data to detect mental health issues and help guide information-seeking patterns. Given this rich potential, it is surprising how limited our knowledge is of the links between well-being and the affective properties information browsed online.

According to the American Psychological Association (APA) dictionary, well-being is defined as "a state of happiness and contentment, with low levels of distress, overall good physical and mental health and outlook, or good quality of life" (VandenBos, 2007). In this study, we focus on psychological and emotional subjective well-being, which we assess through self-reported questionnaires evaluating mental health and/or mood. Over four studies (total N = 947) we test the hypothesis that the affective characteristics of the information people expose themselves to online reflect and shape their psychological and emotional subjective well-being. To quantify the affective properties of the information people expose themselves to, we asked participants to share their web-browsing history and then used a natural language processing (NLP) approach to quantify the valence of the text on webpages that participants browsed. We first related these affective characteristics to participants' emotional and psychological well-being (Study 1 and 2) and then manipulated these factors to examine for causal relationships (Study 3). Finally, we examined whether providing cues about the potential emotional impact of webpages on well-being would influence participants' web-browsing behavior, in a way that was consistent with improvements in well-being (Study 4).

Results

Below we first detail our methods of quantifying the factors of interest (subjective well-being; affective properties of information browsed online). Next, we report observed associations (Study 1 and 2) and then manipulate the factors of interest to test for causation (Study 3). Finally using insight from Study 1, 2 and 3, we develop an intervention to alter information-seeking patterns (Study 4). Note, that Study 2 is a replication of Study 1 except that Study 2 includes less observations per participant than Study 1.

Knowledge-Seeking is Associated with Well-Being (Study 1 &2)

Participants in Study 1 (N = 289) browsed the web for 20-30 minutes a day for five days, and in Study 2 (N = 447) for 30 minutes on one day. They then submitted their web-browsing history. We used this web browsing history to access the web pages visited and extracted the text of these websites (see methods). We then scored the text on affective properties (positive and negative valence, and specific emotions; see **Figure 1**).

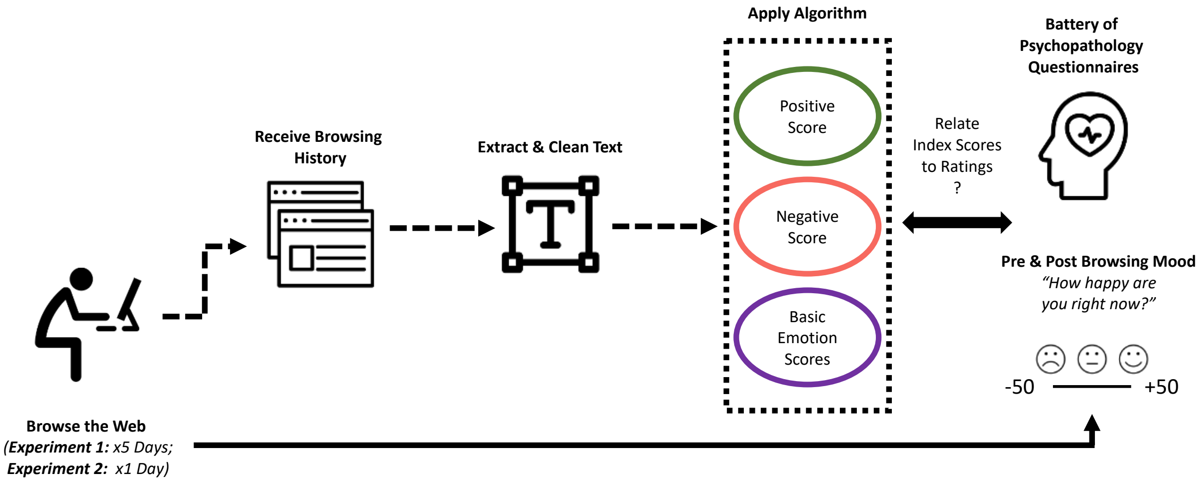

Figure 1. Data collection and pre-processing pipeline. Participants browsed the internet for 20-30 minutes a day for 1-5 days using Mozilla Firefox and then submitted their internet search history for this period. We extracted the paragraph text from each webpage, denoted by <p> in the webpage's html code and cleaned it (see methods). The text was then submitted to an algorithm that calculated a Negative score and a Positive score for each webpage (see methods) as well as scores for Anger, Fear, Anticipation, Trust, Surprise, Sadness, Joy and Disgust. On day one participants completed self-report questionnaires which assess mental health. On days 1-5, participants also indicated their mood directly before and after the web-browsing session. Participates' scores were then related to self-reported psychopathology symptoms and mood.

Quantifying the affective properties of web pages. There are many validated methods to score text on sentiment (valence). These include machine-learning methods (Devlin et al., 2008; Goldberg et al., 2014; Liu et al., 2019) and 'bag of words' (lexicon) approaches which are developed by asking large groups of people to rate words on specific dimensions (Hu & Liu, 2004; Mohammad, 2018). We first tested whether these different methods provide consistent scores for participants. We selected two popular lexicons - the NRC VAD lexicon (Mohammad, 2018) and the Hu and Liu Opinion lexicon (Hu & Liu, 2004) - and a state-of-the-art large language machine learning model, the distilbert-base-uncased-finetuned-sst-2-english (i.e., Distilbert; HuggingFace, 2022; see methods for details). We used each method separately to score all webpages visited by 100 participants from our study and averaged the webpage scores for each participant. We used an intra-class correlation coefficient (ICC) analysis to examine how consistent the scores were across different scoring methods, separately for positive and negative scores. We observed good reliability between all three methods: (i) the NRC VAD lexicon and the Hu and Liu Opinion lexicon (Positive score: ICC = 0.784, $p < 0.001$**;** Negative score: ICC = 0.943, $p < 0.001$); (ii) the NRC VAD lexicon and the Distilbert algorithm (Positive score: ICC = 0.783, $p < 0.001$**;** Negative score: ICC = 0.751, $p < 0.001$); and (iii) the Distilbert algorithm and the Hu and Liu Opinion lexicon (Positive score: ICC = 0.655, $p < 0.001$**;** Negative score: ICC = 0.759, $p < 0.001$). This suggests that these different methods measure the same construct.

As the NRC lexicons performed equivalently to machine learning algorithms but required significantly less computational resources, we opted to use it. We first checked that the ratings computed by the NRC VAD lexicon were reflective of human assessment. One hundred participants, each rated 10 webpages on how positive and negative they were. These scores were significantly related to the NRC Valence scores (Negative score: ICC = 0.707, $p < 0.001$, Positive score: ICC = 0.499, $p < 0.001$), suggesting that the NRC VAD Lexicon scores reflect human subjective assessment of webpages well.

As the method we used scores entire webpages rather than the sub-text participants consume, we tested whether the former was likely to be good indicator of the latter. To that end we examined if the valence of a whole webpage is a relatively good indicator of the valence of part of it. To test this, we randomly extracted segments from webpages (N = 100) with a minimum word count of 200 words (see Yazman, 2017). We observed a strong correlation between the NRC Valence scores of randomly sampled segments and the scores of their respective webpages whole texts (Negative score: $r(98) = 0.936$, $p < 0.001$; Positive score: $r(98) = 0.828$, $p < 0.001$). This result suggests that by analysing the whole text of a webpage, we can reliably compute the sentiment of a random section of a webpage.

Quantifying mental health. To assess mental health, we adopted a dimensionality approach, which considers the possibility that a specific psychopathology symptom is predictive of several conditions and allows an investigation that cuts through classic clinical psychopathology boundaries (Gillan et al., 2016; Rouault et al., 2018; Seow & Gillan, 2020). In particular, previous work used a factor analysis across items in a large battery of traditional psychopathology questionnaires and identified three psychopathology dimensions across those items: '*Anxious-Depression*', '*Social-Withdrawal*' and '*Compulsive-Behaviour and Intrusive Thought'* (Gillian et al., 2016). The factor analysis provided a weight for each item in relation to each dimension.. Thus, a person's symptom severity for each dimension can be quantified by having an individual complete a battery of traditional psychopathology questionnaires and then calculating a weighted average across items' ratings. Indeed, this is what we did for each participant; we Z-scored the ratings of each questionnaire item separately across participants and then for each participant we calculated the three-dimension scores as explained above (as done in Kelly & Sharot, 2021, see methods for more details).

Affective properties of webpages visited provides a marker of mental health. We first examined whether the tendency to browse content with a specific valence was stable over time. To that end, we calculated the Intraclass Correlation Coefficient (ICC) of the Negative and Positive Valence of webpages visited by each participant over the five days. The ICC of both the Negative score (ICC = 0.554, $p < 0.001$) and Positive score (ICC = 0.626, $p < 0.001$) all indicate moderate stability, which was statistically significant. This suggests that the tendency is likely impacted both by 'trait-like' and 'state-like' tendencies.

We next examined if there is a relationship between mental health and the affective properties of pages participants browsed. For each participant, we calculated the three-psychopathology dimension scores ('*Anxious-Depression*', '*Social-Withdrawal*', '*Compulsive-Behaviour and Intrusive Thought*'), which we submitted to a within-subjects factors mixed ANOVAs. In the first mixed ANOVA, the Negative Valence score of the webpages that participants browsed (Z-scored) was input as a within-subject modulating factor. Participants' age and gender were entered as between-subject modulating covariates (both Z-scored). We observed a significant main effect of the Negative Valence score of webpages participants browsed on psychopathology scores (**Study 1:** $F(1,284) = 4.464$, $p = 0.035$, partial eta square = 0.015; **Study 2:** $F(1,442) = 8.303$, $p = 0.004$, partial eta square = 0.018). These results suggest that individuals who browse webpages that are more negatively valenced experience poorer mental health across the three mental health dimensions. The valance of webpages browsed is thus a fingerprint of mental health, rather than associated with a specific condition.

To show this result in a more intuitive manner, we conducted a linear regression with psychopathology as the dependent measure (quantified as the average psychopathology score across the three dimensions) and the Negative Valence score of the webpages that participants browsed, age and gender as the predictor variables, all Z-scored. In line with the results above, we observed a significant positive relationship between psychopathology and the Negative Valence score of webpages participants browsed (**Study 1:** $\beta = 0.087 \pm 0.042$ (SE), $t(288)= 2.069$, $p = 0.039$, $R = 0.122$, **Figure 2a**; **Study 2:** $\beta = 0.099 \pm 0.034$ (SE), $t(446) = 2.930$, $p = 0.004$, $R = 0.138$, **Figure 2b**), suggesting that participants who browsed more negatively valenced webpages reported worse mental health.

The second mixed ANOVA was identical to the first except that the Positive Valence score was input as a within-subject modulating factor instead of the Negative Valence score. We did not observe a significant main effect of Positive Valence score of webpages on psychopathology scores in Study 1, (**Study 1:** $F(1,284) = 0.000$, $p = 0.997$, partial eta square = 0.000), although there was a significant effect in Study 2 (**Study 2:** $F(1,442) = 8.149$, $p = 0.005$, partial eta square = 0.018) with participants reporting higher psychopathology symptoms browsing less positively valanced text.

We implemented the exact same method described above using a second valence lexicon (Hu & Liu, 2004). All results were replicated (see Supplementary Analysis), suggesting that the results are not restricted to a specific method.

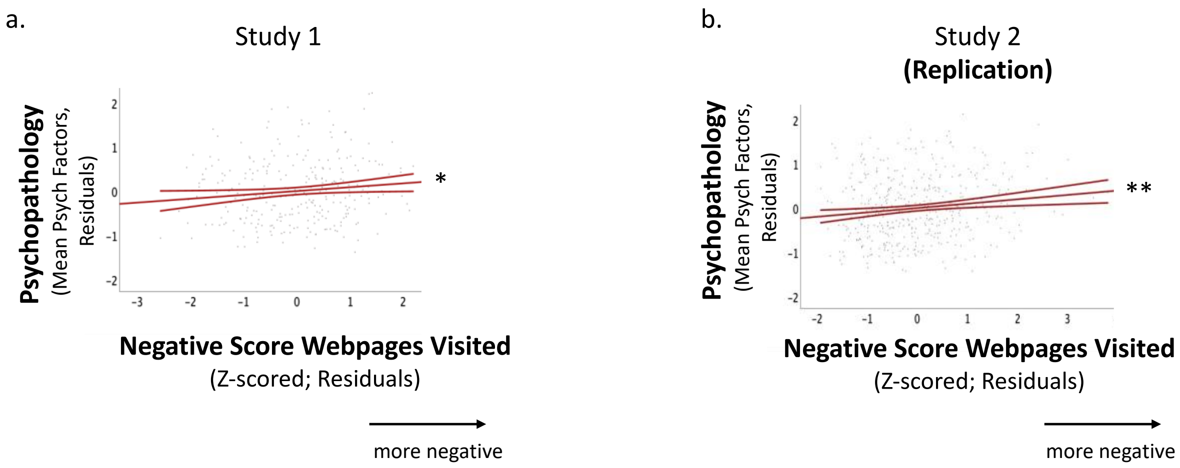

Figure 2. Self-guided browsing of negative content online is associated with poorer mental health. (a&b) Greater psychopathology symptoms (the average score across the three dimensions) are associated with higher Negative Valence score in **(a)** Study 1 and **(b)** Study 2. Dots represent the residual values from the model for individual participants. The outer lines represent confidence intervals. The inner line represents the relationship between the abscissa and ordinate controlling for the effect of age and gender. $^{**}P<0.01$, $^{*}P<0.05$ (two-sided).

Because participants knew they would submit their browsing history, it is possible they may browse differently than if 'no one was watching', despite anonymity. This would induce noise that may make the relationship between information-seeking and well-being more difficult to detect and thus likely larger than reported here. While participants were explicitly asked to browse the internet during the study session, not all of them did. We suspected as much from the time stamps and thus asked participants after completing the study, whether they indeed submitted data that was browsed in-session or from their archived browsing history. Thirty-nine participants in Study 1 and 7 in Study 2 admitted they submitted archived data (due to this small N the following analysis was conducted across Study 1 and 2 together). We tested whether the average Valence scores of webpages browsed from this group was different than for those who browsed in-session – it was not for Positive scores (browsed in study data M = 0.015, SD = 0.032, archived data M = 0.015, SD = 0.030, t(514) = 0.243, p = 0.808, Cohen's d = 0.107) nor for Negative scores (browsed in study data M = 0.030, SD = 0.014, archived data M = 030, SD = 0.014, t(514) = 0.112, p = 0.928, Cohen's d = 0.081) This suggests that it is unlikely that participants are browsing more positive or negative webpages on average due to the study set-up.

A Bidirectional Association Between Knowledge-Seeking and Mood (Study 1 & 2)

Thus far, we observed that the valence of information consumed from self-guided searches provides a fingerprint of mental health. Next, we ask whether it is also associated with mood, which is a feature of well-being, and if so whether this association is bidirectional.

To that end, we asked participants (**Study 1**: N = 164; **Study 2**: N = 400) to indicate their current mood directly before their web-browsing session and directly afterwards, on scales from -50 (very unhappy) to +50 (very happy). We tested whether participants pre-browsing mood and post-browsing mood was related to the valence of information they browsed.

First, we examined the relationship between valence of information consumed and pre-browsing mood. We ran separate linear mixed effect models - one predicting the Negative Valence score and the other the Positive Valence score of webpages visited - from participants' pre-browsing happiness ratings in Study 1 (fixed and random effects) along with age and gender as fixed effects. In Study 2, as we only had one observation per subject for each variable of interest (compared to five in Study 1), we ran two simple linear regressions predicting the Negative Valence score and Positive Valence score from pre-browsing mood ratings, controlling for age and gender (**see Supplementary Table** for control variable statistics). We found that participants who reported better mood prior to browsing the internet, exposed themselves to less negatively valenced webpages (**Study 1:** β -0.082 ± 0.041 (SE), t(380.29) = -1.981, p = 0.048, **Figure 3a**; **Study 2:** β = -0.096 ± 0.049 (SE), t(399) = -1.974, p = 0.049, R = -0.099, **Figure 3a**), with no significant relationship observed for Positive Valence score (**Study 1:** β = -0.001 ± 0.002 (SE), t(104.88) = -0.493, p = 0.623; **Study 2:** β = 0.088 ± 0.048 (SE), t(399) = 1.830, p = 0.068, R = 0.092).

Next, we ran a similar analysis as above to predict post-browsing mood from the Negative score of webpages participants visited, while controlling for pre-browsing mood in all models. We found that participants expressed better mood after browsing less negatively valenced webpages, controlling for mood pre-browsing age and gender (**Study 1:** β = -0.044 ± 0.019 (SE), t(58.12) = -2.338, p = 0.023, **Figure 3b**; **Study 2:** β = -0.093 ± 0.035 (SE), t(399) = -2.686, p = 0.008, R = -0.134, **Figure 3b**). Participants also reported better mood after browsing more Positive Valence webpages in Study 1 (**Study 1:** β = 0.037 ± 0.019 (SE), t(82.62) = 2.013, p = 0.047) but this effect was not significant in Study 2 (β = 0.063 ± 0.035 (SE), t(399) = 1.770, p = 0.077, R = 0.089). Together, these results suggest a bi-directional relationship between mood and the Negative Valence of webpages participants consume from self-guided se arches. Specifically, individuals that were happier directly before browsing the internet, browsed less negatively valenced information, and individuals who browsed less negatively valenced information reported being happier after browsing the internet. As these results are still correlational, we next ran a study to test for causation.

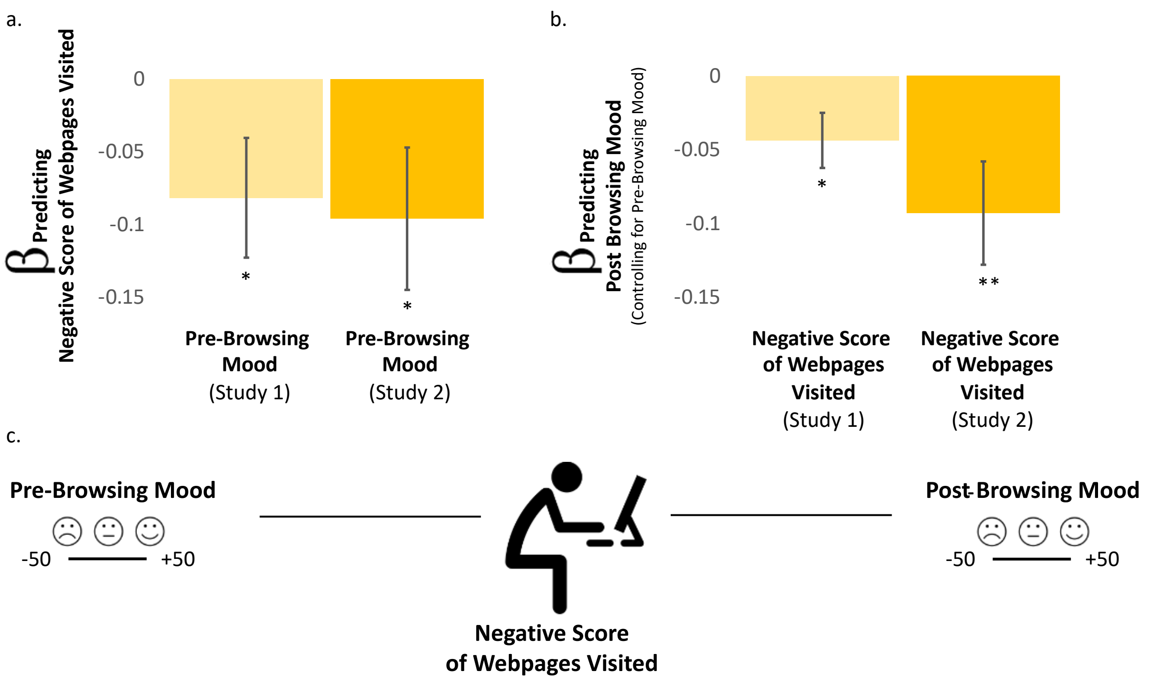

Figure 3. Browsing more negatively valanced webpages is associated with worse mood before and after browsing. (a) Plotted on the y-axis is the beta coefficient predicting the Negative score of webpages visited by participants from their pre-browsing mood in Study 1 (light yellow) and Study 2 (dark yellow). Participants with worse pre-browsing mood tend to browse more negatively valenced webpages controlling for age and gender in both studies. **(b)** Plotted on the y-axis is the beta coefficient predicting participants post browsing mood from the Negative score of webpages they visited in Study 1 (light yellow) and Study 2 (dark yellow), controlling for pre-browsing mood, age and gender. Participants who browsed more negatively valenced webpages reported worse post-browsing mood. Error bars = standard error (SEM). ** = P <0.01, * = P <0.05 (two-sided).

The Bidirectional Association Between Knowledge-Seeking and Mood is Causal (Study 3)

In Study 3 (N = 102) we first manipulated the webpages participants were exposed to and then tested for mood. Specifically, participants were asked to read information from two webpages randomly selected either from six negative webpages (i.e., negative valence condition; N = 55) or six neutral pages (i.e., control condition; N = 47). The negative pages were randomly selected from all webpages browsed in Study 1 that were +2.5 SD from the mean Negative score. The neutral webpages were randomly selected from webpages browsed in Study 1 that were between -1 and +1 SD from the mean. Participants indicated their mood levels on a scale ranging from *very unhappy* (-50) to *very happy* (+50) before and after being exposed to the webpages.

A 2 (condition: negative valence, control) by 2 (time: pre-manipulation, post manipulation) ANOVA on self-reported mood revealed a significant interaction (F(1, 97) = 15.922, p < 0.001, partial eta squared = 0.141). The interaction was characterised by participants in the negative valence condition reporting feeling unhappier post manipulation (M = -1.93, SD = 23.69) compared to pre-manipulation (M = 9.47, SD = 24.29, t(54) = -5.031, p < 0.001, Cohen's d = -0.678), with no difference in the control condition (post manipulation: M = 9.31, SD = 19.31, pre-manipulation: M = 9.53, SD = 19.42, t(46) = 0.131, p = 0.896, Cohen's d = 0.019). Importantly, participants in the negative valence condition reported feeling unhappier post manipulation relative to controls (negative valence condition: M = -1.93, SD = 23.69; control condition: M = 9.53, SD = 19.43, t(100) = 2.242, p = 0.010, Cohen's d = 0.525), with no difference pre-manipulation (negative valence condition: M = 9.47, SD = 24.29; control condition: M = 9.32, SD = 19.32, t(100) = -0.035, p = 0.972, Cohen's d = -0.007; **see Figure 4a**). This suggests that being exposed to negatively valence webpages results in worse mood.

Now that the negative valence group reported worse mood than the control group, we asked whether this group of participants would go on to consume more negatively valenced webpages than the control group from self-guided searches. To that end, participants were asked to browse the internet for 10-minutes and then submit their internet search history for this period. The negative valence of webpages participants exposed themselves to was quantified as in Studies 1 and 2 (**see Methods**). Results show that participants in the negative valence condition subsequently browsed significantly more negatively valenced webpages (M = 0.034, SD = 0.020) than those in the control condition (M = 0.026, SD = 0.014, t(96.04) = -2.259, p = 0.026; Cohen's d = -0.436; **see Figure 4b**). These results suggest a causal bi-directional relationship between participants' mood and web-browsing patterns (**see Figure 4c**). All results remain the same when removing participants that have a values plus/minus 3 standard deviations from the mean.

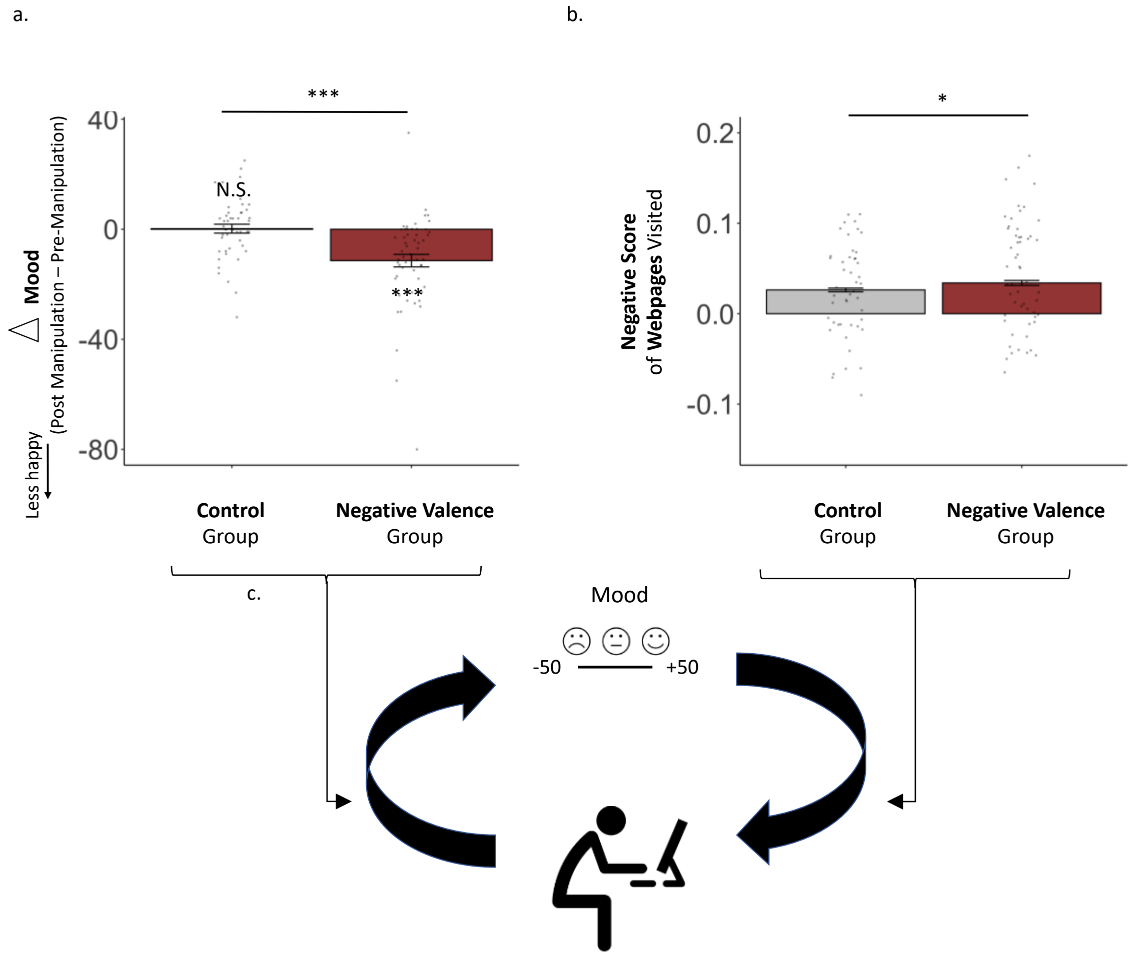

Figure 4. Bi-directional relationship between mood and the valence of information consumed. (a) Participants were asked to browse two webpages, randomly selected from either six very negative webpages or six neutral webpages (control). Participants reported their mood on a scale ranging from very unhappy (-50) to very happy (+50) before and after the manipulation. Plotted on the y-axis are participants' post manipulation mood rating minus their pre-manipulation mood rating for the negative valence condition (grey) and control condition

(red). Participants in the negative valence group reported worse mood after browsing compared to before, while participants in the control condition reported no difference in their mood after browsing compared to before. Moreover, participants in the negative valence condition reported worse mood after browsing than those in the neutral condition. **(b)** After browsing the webpages selected by us participants had the opportunity to freely browse the web. Those in the negative valence condition browsed significantly more negatively valenced webpages than those in the control condition. Individual scores are shown as dots. **(c)** The results suggest a bi-directional relationship between mood and valence of webpages browsed. Specifically, **(b)** worse mood leads to browsing more negatively valenced information, and **(a)** browsing more negatively valenced information leads to worse mood. Error bars = standard error (SEM). *** = P <0.001, * = P <0.05, N.S. = not significant (two-sided).

An Intervention to Alter Patterns of Knowledge-Seeking (Study 4).

Studies 1-3 show that browsing negatively valanced information is associated with negative features of psychological and emotional well-being. We thus pondered whether people would select to expose themselves to more positive and less negative information if they had advance knowledge of the affective properties of webpages. That is, would providing people with cues about the valence of webpages alter their knowledge-seeking patterns, resulting in less consumption of negative, and more consumption of positive, knowledge?

To answer this question, we conducted Study 4. Participants were assigned to either a label condition or no label condition. In the **no label condition** participants were randomly presented with three Google search result pages from a set of 18 (**Figure 5a)**. Each page contained three possible webpage links participants could click on. They simply had to click on one of the three on each trial. They would then spend at least 90 seconds browsing that webpage.
These 18 pages were selected from Google's list of frequent queries, for which Google results contained varying levels of valence scores (i.e., positive, neutral and negative).

Participants in the **label condition** did the same, except that next to each link there was a label indicating the sentiment of that webpage (**Figure 5a**). The label was assigned based on valence scores calculated as in Studies 1-2. If the Positive score of the page was >2.5 SD from the mean of webpages browsed in Studies 1-2, the webpage was given the label *"feel better"*; If the negative score of the page was >2.5 SD from the mean of webpages browsed in Studies 1-2, the webpage was given the label *"feel worse";* if neither was neither it was given the label *"neutral"*. The labels indicate whether *on average* this website makes people feel worse/better.

The question of interest was if participants would use the labels to alter the information they exposed themselves to. The results suggest they did. A 2 (condition: label, no label) by 3 (valence: positive, neutral, negative) ANOVA on webpage choices revealed a significant interaction between condition and valence ($F(1, 107) = 7.695$, $p = 0.007$, partial eta squared = 0.067; **see Figure 5b**). The interaction was characterised by participants in the label condition selecting more webpages with the positive label ($M = 1.444$, $SD = 1.04$) than the no label condition ($M = 1.055$, $SD = 0.68$, $t(90.93) = -2.314$, $p = 0.023$, Cohen's d = -0.445; **see Figure 5b**) and less webpages with the negative label ($M = 0.630$, $SD = 0.73$) than the no label condition ($M = 1.000$, $SD = 0.839$, $t(107) = 2.251$, $p = 0.016$, Cohen's d = 0.469; **see Figure 5b**). There was no difference in the number of neutral webpages selected between the label condition ($M = 0.910$, $SD = 0.88$) and no label condition ($M = 0.910$, $SD = 0.85$, $t(107) = 0.010$, $p = 0.992$, Cohen's d = 0.002; **see Figure 5b**).

Additionally, within the label condition webpages with the positive label were selected more than neutral webpages (Mean Difference = 0.537, $SD = 1.77$, $t(53) = 2.234$, $p = 0.030$, Cohen's d = 0.304) and negative label webpages (Mean Difference = 0.815, $SD = 1.58$, $t(53) = 3.792$, $p < 0.001$, Cohen's d = 0.516), with the latter two not different (Mean Difference = 0.278, $SD = 1.23$, ($t(53) = 1.653$, $p = 0.104$, Cohen's d = 0.225).

In contrast, in the no label condition none of the webpages were labelled, thus there was no difference in the likelihood of selecting webpages which should have been labelled as positive and neutral (Mean Difference = 0.145, $SD = 1.28$, $t(54) = 0.841$, $p = 0.404$, Cohen's d = 0.113), or should have been labelled as positive and negative (Mean Difference = 0.055, $SD = 1.27$, $t(54) = 0.319$, $p = 0.751$, Cohen's d = 0.043), nor between those which should have been labelled as neutral and negative (Mean Difference = -0.091, $SD = 1.53$, $t(54) = -0.440$, $p = 0.661$, Cohen's d = -0.059).

Together, the results suggest that emphasizing the affective properties of webpages decreases the number of negative webpages, and increases the number of positive webpages, participants expose themselves to. Clearly, we are not suggesting that one should make information consumption decisions based only on affective properties.

To the contrary, we have written extensively about the multi-features of information critical in making information-consumption decisions, of which affect is only one (e.g., Sunstein & Sharot, 2020; Kelly & Sharot, 2021; Cogliati-Dezza et al., 2022; Charpentier et al., 2018; Vellani et al., 2020; Vellani et al., 2022). Instrumental utility of information and uncertainty reduction, for example, do and should drive information-seeking. What we envision is that affective labels could be used in the future together with other labels (such as the instrumental utility of information and its reliability) to empower users to make better information-consumption decisions that align with their goals.

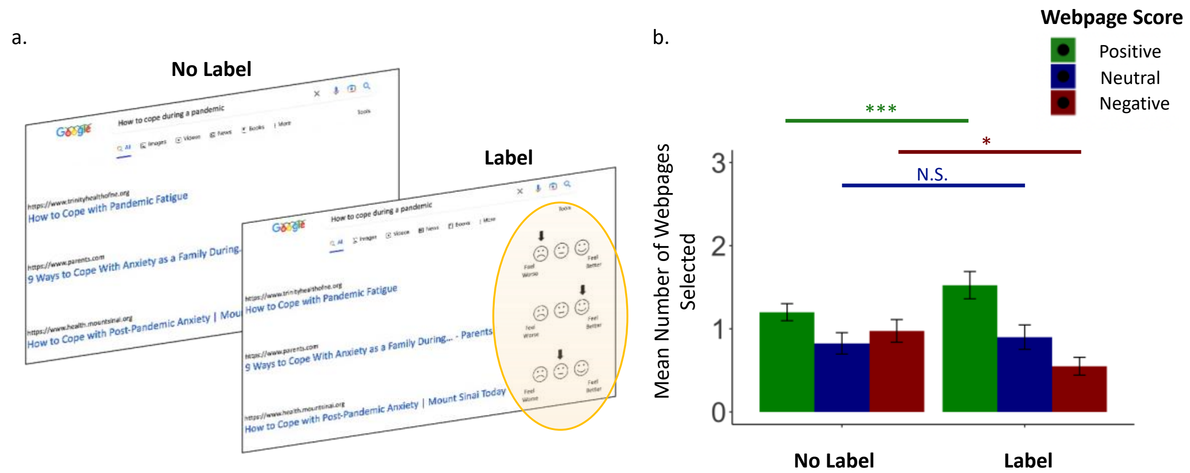

Figure 5. Novel online intervention decreases the amount of negative information browsed online. (a) Participants were assigned to either a label or no label condition. In the label condition they were presented with three Google search results pages from a set of 18. Each included three possible webpage links. Participants were asked to select the webpage they wanted to visit. In the label condition they also observed a label next to each link: either positive (*"feel better"*; green), neutral (blue), or negative *("feel worse"*; red). The yellow oval is for illustrations purposes only and *was not* present in the actual experiment. **(b)** Participants in the label condition clicked on more webpages with the positive label and less webpages with the negative label than the no label condition. There was no difference in the number of neutral webpages selected. This suggests that cues indicating the effective properties of webpages alters participants web browsing patterns, such that they expose themselves to less negative and more positive information. Error bars = standard error (SEM). *** = P <0.001, * = P <0.05, N.S. = not significant (two-sided).

Discussion

Our findings reveal that knowledge-seeking online both reflects and shapes well-being. The valence of the information people browse online was associated with their psychological and emotional well-being, with those consuming more negative information tending to report lower well-being as measured by mood and self-reported psychiatric symptoms. A central question is whether browsing patterns alter well-being or vice versa. Our results supports a reciprocal causal relationship between the affective properties of information consumed from self-guided searches and mood.

In particular, we show that participants who reported worse mood prior to browsing tended to access more negative online content. Exposure to negative content was in turn associated with worse mood (controlling for pre-browsing mood). We established the causality of this relationship by exposing participants to either negative or neutral webpages. We found that exposure to negative webpages resulted in worse mood, and this change in mood then led to more browsing of negatively valence information. Together, these findings reveal a feedback loop; low mood leads to the consumption of more negative information online which in turn leads to worse mood and so on.

We found that participants' tendency to browse webpages of specific valence, either more negative or positive, was moderately stable over time, indicating that web-browsing behavior has both trait-like and state-like characteristics. The actual relationship between online information consumption tendencies and well-being is likely to be stronger than observed here, as our analysis only accounted for the sentiment of text and further studies could potentially include images and videos. Additionally, future studies could measure the exact amount of time users spend on each piece of content, providing a more precise estimate of the relationship.

Our study is innovative in its approach of examining the link between the information browsed and psychological well-being. Previous research in this area has primarily focused on analysing search engine queries rather than the actual text on webpages visited. This traditional approach monitors certain keywords, such as "therapist" or "Prozac," to infer changes in population mental health. This method may be limited in its ability to assess an individual's well-being, as it only provides a limited dataset based on a few keywords which would be used by individuals who are already aware of their symptoms and seek help.

The rational of our approach (namely, that affective properties of text is more informative then pre-selected terms) is consistent with studies showing a relationship between the affective properties of generate content and mental health (i.e., such as posting on social media; De Choudhury et al., 2017; Chancellor et al., 2019; Kelley et al, 2021; Eichstaedt et al., 2019). This may indicate an intriguing overlap between the mechanism governing information-seeking and those governing information-sharing (e.g., Vellani et al., 2022). An advantage of the current approach, is that it does not necessitate that people share information online, an activity that is clearly prevalent, but less so than online information-seeking.

Given the established bi-directional relationship between exposure to negative information and affective well-being, we examined whether individuals would choose to access more positive and less negative information if they were made aware of the affective nature of webpages before browsing. Indeed, our results showed that providing individuals with cues about the valence of webpages effectively changed their browsing patterns, leading to a decrease in exposure to negative content and an increase in exposure to positive content. This result suggests that a simple intervention may be effective in reducing exposure to negative information and potentially improving mood.

In many cases it would be obviously suboptimal to solely base information-consumption decisions on the affective properties of information. For example, if someone searches for information on whether smoking causes cancer, the most positive link may not necessarily be the wise choice. Thus, we do not envision the intervention described here as a stand-alone tool. Rather, by providing users with affective labels in addition to other labels, such as the reliability and instrumental utility of information, they can make more informed decisions that align with their current goals. For instance, users may want to prioritize the instrumental utility of information in one situation and prioritize their mood in another by focusing on affective labels. As such, our study not only provides evidence for the relationship between knowledge-seeking and psychological and emotional well-being, but can inform the development of tools aimed at enhancing well-being by improving information consumption decisions.

Methods

Study 1

Participants. Three hundred and twelve participants completed a study online via Prolific's recruitment platform. Data from 23 participants whose searches did not result in at least 1KB of text from at least 3 webpages each day was not analyzed further. Thus, data of 289 participants were analyzed (age = 33.17, SD =11.71; females = 50.5%, males = 48.1%, other = 1.4%). Out of those, 171 participants also completed state mood ratings. Data of five participants who indicated that contrary to the instructions they submitted archived browsing history was not included in mood analysis, as their current mood ratings obviously could not be temporally associated with their submitted browsing data, leaving for mood analysis N = 164 (age = 33.23, SD =11.62; females = 52.4%, males = 47.6%, other = 0%)All participants received £7.50 for their participation on day 1 and £3.25 for days 2-5. Ethical approval was provided by the Research Ethics Committee at University College London.

Procedure

Data collection. Participants were asked to browse the internet for 20-30 minutes a day for 5 days using Mozilla Firefox and then submit their internet search history for this period. We extracted the paragraph text from each webpage, denoted by <p> in the webpage's html code, using the '*rvest*' package in RStudio. We then cleaned the text by removing extraneous information such as punctuation, symbols (e.g., @, #), emojis, links (URLs), and all other non-alphanumeric characters (similar to Kelley et al., 2021). Participants were instructed not to visit any sites which require a password nor to watch videos as we were not able to access, extract, or analyze such content. In addition, participants were asked to browse the internet during non-work hours so that their web-browsing behavior would not reflect mandatory work-related tasks. All consecutive duplicate webpages were removed from analysis. Participants for whom we had less than three webpages from which we could extract at least 1KB of data per day were excluded from analysis.

Text valence analysis. To quantify the valence of webpages visited, we used the NRC VAD lexicon (Mohammad, 2018), which categorizes the valence of terms on a scale from 0 (most negative) to 1 (most positive). In line with Kiritchenko and colleagues (2020), we computed the percentage of words with a Positive valence score greater or equal to 0.75 (2668 terms, e.g., "*delicious*" and "*admire*") and percentage of words with a Negative valence score less or equal to 0.25 (3081 terms, e.g., "*despise*" and "*danger*"), out of all words contained in the extracted text of each webpage visited for each of the five days. We then averaged these positive valence and negative valence scores separately across all webpages visited on each day and then averaged the daily scores across the five days to create a Positive Valence score and Negative Valence score, respectively. We also quantified separately the percentage of *Anger, Fear, Anticipation, Trust, Surprise, Sadness, Joy* and *Disgust* associated words greater or equal to 0.75, as defined by the NRC Emotion Lexicon (Mohammad and Turney, 2013), out of all words on each webpage visited by participants for each day and then across days (i.e., Emotion scores).

To assess whether the NRC VAD lexicon scores of webpages was related to alternative sentiment analysis approaches, we computed the Intraclass Correlation Coefficient (ICC) for the mean NRC Positive and Negative Valence scores of webpages visited by participants (N = 100) separately with the same from another widely used lexicon, the Hu and Liu Opinion lexicon (Hu & Liu, 2004), which categorizes 2006 words as positive and 4783 words as negative. Next, we calculated the ICC for the mean NRC Positive and Negative scores with a state-of-the-art large language machine learning model, the distilbert-base-uncased-finetuned-sst-2-english (i.e., Distilbert; HuggingFace, 2022). Finally, we calculated the ICC for the Hu and Liu Opinion lexicon with the distilbert-base-uncased-finetuned-sst-2-english.

To examine if the NRC VAD lexicon scores corresponded to human ratings, we asked participants (N = 100) to rate the positive (0 (not at all) to 6 (very positive)) and negative valence (0 (not at all) to 6 (negative)) of 10 randomly assigned webpages from a corpus of 48 webpages. We then computed the NRC VAD Lexicon's Positive and Negative score for each webpage and correlated them with their respective human rating for that webpage.

Finally, we were interested whether the valence of webpages' whole text was associated with a sample of its text. To test this, we randomly extracted segments from webpages (N = 100) with a minimum word count of 200 words and then computed their NRC Positive and Negative Valence scores as well as for the whole text on webpages. We then correlated the Positive and Negative scores for the random samples text with its whole text.

Mental Health and mood. On day one, participants completed self-report questionnaires which assess psychopathology symptoms (the list is adopted from Gillan et al., 2016) These included: Obsessive-Compulsive Inventory – Revised (OCI-R; Foa et al., 2005), Self-Rating Depression Scale (SDS; Zung, 1965), State-Trait Anxiety Inventory (STAI; Spielberger et al., 1983), Alcohol Use Disorder Identification Test (AUDIT; Saunders et al., 1993), Apathy Evaluation Scale (AES; Marin et al., 1991), Eating Attitudes Test (EAT-26; Garner et al., 1982), Barratt Impulsivity Scale (BIS-11; Patton et al., 1995), Short Scales for Measuring Schizotypy (Mason et al., 2005), Liebowitz Social Anxiety Scale (LSAS; Fresco et al., 2001). On days 1-5, participants indicated their current mood directly before their web-browsing session and directly afterwards, on scales from -50 (very unhappy) to + 50 (very happy). The task was coded using the Qualtrics online platform (https://www.qualtrics.com).

Analysis

Relating the valence of webpages to mental health. Each participant was scored on the three psychopathology dimensions identified by Gillan and colleagues (2016) and replicated by Rouault and colleagues (2018) ('*Anxious-Depression*', '*Social-Withdrawal*' and '*Compulsive-Behaviour and Intrusive Thought*'. To generate these scores, we followed Kelly & Sharot (2021) - we first Z-scored the ratings for each questionnaire item separately across participants. Next, we multiplied each Z-scored item by its factor weight as identified earlier (Gillan et al., 2016). Then for each subject the three-psychopathology dimension scores were calculated by summing all of the weighted items assigned to each dimension.

The valence of webpages visited by participants were then related to the psychopathology dimensions by submitting the three-psychopathology dimension scores into a mixed ANOVA with psychopathology dimension as a within-subject factor and valence as within subject modulating covariates as well as participants' age and gender as between-subjects modulating covariates (similar to Kelly & Sharot, 2021). This analysis was then followed up with a simplified analysis in which the average of the three-psychopathology dimension scores of each individual were entered as a dependent measure in a linear regression with valence entered as an independent measure as well as age and gender.

Relating the valence of webpages to mood. To investigate the relationship between web-browsing patterns and mood, we asked participants to indicate their current mood directly before their web-browsing session and directly afterwards, on scales from -50 (*very unhappy*) to +50 (*very happy*). We first assessed whether participants pre-browsing mood was related to the valence of information they browsed. To that end, we ran two separate mixed effect models each including participants pre-browsing mood ratings as fixed and random effects along with age and gender as fixed effect predicting the Negative Valence score and Positive Valence score of webpages visited, separately. Next, we were interested in whether the valence of the webpages that participants browsed had an impact on their mood directly after browsing the internet. To test this, we once again ran two mixed effect models, each predicting post browsing mood ratings from either the Negative and Positive Valence score of webpages visited (input as a fixed and random effect), controlling for pre-browsing mood (fixed and random effect) as well as age and gender (fixed effect).

Assessing the stability of the valence of web-browsing across time. To assess the within-subject stability of the valence of webpages visited across the five days, we calculated an intraclass correlation coefficient (ICC). Specifically, we submitted separately the Negative and Positive Valence score and the scores for the specific emotions of webpages visited by each participant for each of the five days into ICC analysis.

Study 2: Replication of Study 1

Participants. Five hundred participants completed a study online via Prolific's online recruitment system. Data of 53 participants from whom we could not obtain at least 1KB of text from a minimum of 3 webpages a day was not analyzed. Thus, data of 447 participants were analyzed (age = 33.85, SD =12.58; females = 56.4%, males = 41.8%, other = 1.8%). For the mood analysis, we only included those participants that submitted data that was browsed during the study session (N = 400, age = 33.23, SD =11.62; females = 52.4%, males = 47.6%, other = 0%), as otherwise their reported mood ratings would not be temporally reflective of their submitted browsing data. Participants received £7.50 for their participation. Ethical approval was provided by the Research Ethics Committee at University College London.

Procedure

The procedure was exactly as in Study 1 except that all participants were asked to browse the internet for 30-minutes for one day.

Analysis

Relating the valence score of webpages to psychopathology. This analysis was conducted as described in Study 1.

Relating the valence of webpages to mood. We first tested whether participants pre-browsing mood was related to the valence of information they browsed. As we only had one observation per participant for each variable of interest, (compared to five observations in Study 1), we ran two simple linear regressions predicting the Negative Valence score and Positive Valence score, separately, from pre-browsing mood ratings, controlling for age and gender. Next, we were interested in whether the valence of the webpages that participants browsed had an impact on their mood directly after browsing the internet. To test this, we ran two simple linear regressions, both predicting participants post browsing mood ratings from either the Negative or Positive Valence score of webpages visited. Both models controlled for participants pre-browsing mood ratings, age and gender.

Study 3

Participants. One hundred and thirty-nine participants completed the study on Qualtrics (www.qualtrics.com) and were recruited via Prolific's online recruitment platform (www.prolific.co). Participants received £7.50 per hour for their participation. Thirty-seven participants were excluded for not providing at least 3 webpages from which we could extract at least 1KB of data, leaving 102 participants (negative valence condition: N = 55, age = 33.96, SD = 9.68; females = 45.5%, males = 49.1%, other = 5.5%; control condition: N = 47, age = 34.72, SD = 12.14; females = 46.8%, males = 51.1%, other = 2.1%).

Procedure

Data collection. To assess the directionality of the relationship between mood and web-browsing patterns, we first conducted a manipulation of webpages that participants were exposed to. Specifically, we asked participants to browse two webpages, randomly selected from either six very negative (i.e., negative valence manipulation) or six neutral (i.e., control manipulation) webpages. The stimuli were selected from webpages that participants browsed in Studies 1-2. The negative webpages had a Negative Score of >2.5 SD from the mean of webpages browsed in Studies 1-2, while the neutral webpages had a Negative Score of between -1 and 1 SD from the mean of webpages browsed in Studies 1-2. The valence of the webpages was quantified using the exact method as outlined in Study 1 and pages were included if they had a Negative score greater or equal to 3 standard deviations from the mean (i.e., negative webpages) or a Negative score between 0 and 1 standard deviations from the mean (i.e., neutral pages). Participants indicated their happiness levels on a scale ranging from very unhappy (-50) to very happy (+50) before and after the manipulation.

Next, participants were asked to browse the internet for 10-minutes using Mozilla Firefox and then submit their internet search history for this period. We then extracted the paragraph text from each webpage, denoted by <p> in the webpage's html code, using the '*rvest*' package in RStudio. Participants were instructed not to visit any sites which require a password, or to watch videos, as we were not able to access, extract or analyze this content. All consecutive duplicate webpages were removed from analysis.

Analysis

To assess whether the mood manipulation was successful, a 2x2 ANOVA with condition (negative manipulation, control) as a between-subject factor, and time (pre-manipulation, post manipulation) as the within-subject factor was conducted. Follow up pair-wise t-tests were also conducted. Next, for each participant, we computed the Negative Valence score of the webpages browsed. Finally, we tested for a difference in the Negative Valence score of the webpages browsed between the negative valence manipulation group and control group.

Study 4

Participants. One hundred and nine participants (label condition: N = 55; no label condition: N = 54) completed the study on Qualtrics (www.qualtrics.com) and were recruited via Prolific's online recruitment platform (www.prolific.co). Participants received £7.50 per hour for their participation.

Procedure

Data collection. Participants were assigned to either a label condition or no label condition. In the **no label condition** participants were randomly presented with three Google search result pages from a set of 18. Each page contained three possible webpage links they could click on. They simply had to click on one of the three on each trial. They would then spend 90 seconds browsing that webpage.

Participants in the **label-condition** did the same, except that next to each link there was a label indicating the affective label of that webpage. The label was assigned based on valence scores calculated as in Studies 1-2. If the Positive score of the page was >2.5 SD from the mean of webpages browsed in Studies 1-2, the webpage was given the label *"feel better"*; If the negative score of the page was >2.5 SD from the mean of webpages browsed in Studies 1-2, the webpage was given the label *"feel worse";* if neither was true it was given the label neutral. The labels indicate that *on average* this website makes people feel worse/better.

Analysis

To assess whether the manipulation was successful, we used a 2x3 ANOVA with condition (label vs. no label) as a between-subject factor, and label valence (positive, negative, neutral) as the within-subject factor. Follow up pair-wise t-tests were conducted.

Acknowledgments
We thank Bastien Blain, Irene Cogliati Dezza, Moshe Glickman, Laura Globig, Valentina Vellani, Sarah Zheng, India Pinhorn, Hadeel Haj-Ali, Jade Serfaty and Daniel Rebbin for comments on previous versions of the manuscript. The work was funded by a Wellcome Trust Senior Research Fellowship to TS.

Competing interests
The authors declare no competing interests.

References

Pieh, C., O´ Rourke, T., Budimir, S., & Probst, T. (2020). Relationship quality and mental health during COVID-19 lockdown. *Plos one*, 15(9), e0238906.

Ertel, K. A., Glymour, M. M., & Berkman, L. F. (2009). Social networks and health: A life course perspective integrating observational and experimental evidence. *Journal of Social and Personal Relationships*, 26(1), 73-92.

Robles, T. F., & Kiecolt-Glaser, J. K. (2003). The physiology of marriage: Pathways to health. *Physiology & behavior*, 79(3), 409-416.

Marconcin, P., Werneck, A. O., Peralta, M., Ihle, A., Gouveia, É. R., Ferrari, G., ... & Marques, A. (2022). The association between physical activity and mental health during the first year of the COVID-19 pandemic: a systematic review. *BMC Public Health*, 22(1), 209.

Peluso, M. A. M., & De Andrade, L. H. S. G. (2005). Physical activity and mental health: the association between exercise and mood. *Clinics*, 60(1), 61-70.

Ettman, C. K., Adam, G. P., Clark, M. A., Wilson, I. B., Vivier, P. M., & Galea, S. (2022). Wealth and depression: A scoping review. *Brain and Behavior*, 12(3), e2486.

Pollack, C. E., Chideya, S., Cubbin, C., Williams, B., Dekker, M., & Braveman, P. (2007). Should health studies measure wealth?: A systematic review. *American journal of preventive medicine*, 33(3), 250-264.

DataReportal (2022), "Digital 2022 Global Digital Overview," retrieved from https://datareportal.com/reports/digital-2022-global-overview-report

Kelly, C. A., & Sharot, T. (2021). Individual differences in information-seeking. *Nature communications*, 12(1), 7062.

Sharot, T., & Sunstein, C. R. (2020). How people decide what they want to know. *Nature Human Behaviour*, 4(1), 14-19.

Kobayashi, K., Ravaioli, S., Baranès, A., Woodford, M. & Gottlieb, J. Diverse motives for human curiosity. *Nature Human Behaviour*. 3, 587–595. https://doi.org/10.1038/s41562-019-0589-3 (2019).

Sunstein, C. R. (2019). Ruining popcorn? The welfare effects of information. *Journal of Risk and Uncertainty*, *58*, 121-142.

VandenBos, G. R. (2007). *APA dictionary of psychology*. American Psychological Association.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* preprint arXiv:1810.04805.

Goldberg, Y., & Levy, O. (2014). word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method. *arXiv* preprint arXiv:1402.3722.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv* preprint arXiv:1907.11692.

Hu, M., & Liu, B. (2004, August). Mining and summarizing customer reviews. *In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 168-177).

Mohammad, S. (2018, July). Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. *In Proceedings of the 56th annual meeting of the association for computational linguistics* (volume 1: Long papers) (pp. 174-184).

Yazman, J. (2017). A pair of text analysis explorations. Retrieved August 17, 2022, from https://joshyazman.github.io/pair-of-text-analysis-explorations/

Gillan, C. M., Kosinski, M., Whelan, R., Phelps, E. A., & Daw, N. D. (2016). Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. *elife*, 5, e11305.

Rouault, M., Seow, T., Gillan, C. M., & Fleming, S. M. (2018). Psychiatric symptom dimensions are associated with dissociable shifts in metacognition but not task performance. *Biological psychiatry*, 84(6), 443-451.

Seow, T. X., & Gillan, C. M. (2020). Transdiagnostic phenotyping reveals a host of metacognitive deficits implicated in compulsivity. *Scientific reports*, 10(1), 1-11.

Dezza, I. C., Maher, C., & Sharot, T. (2022). People adaptively use information to improve their internal states and external outcomes. *Cognition*, 228, 105224.

Charpentier, C. J., Bromberg-Martin, E. S., & Sharot, T. (2018). Valuation of knowledge and ignorance in mesolimbic reward circuitry. *Proceedings of the National Academy of Sciences*, 115(31), E7255-E7264.

Vellani, V., de Vries, L. P., Gaule, A., & Sharot, T. (2020). A selective effect of dopamine on information-seeking. *elife*, 9, e59152.

Vellani, V., & Sharot, T. (2022). How people decide when to inform others.

De Choudhury, M., Gamon, M., Counts, S. & Horvitz, E. Predicting depression via social media. *ICWSM 2*, 128–137 (AAAI, 2013).

Chancellor, S., Birnbaum, M., Caine, E., Silenzio, V. & De Choudhury, M. A taxonomy of ethical tensions in inferring mental health states from social media. *in Proc. Conference on Fairness, Accountability, and Transparency* (FAT*) (ACM, 2019).

Kelley, S. W., & Gillan, C. M. (2022). Using language in social media posts to study the network dynamics of depression longitudinally. *Nature communications*, 13(1), 870.

Eichstaedt, J. C. et al. Facebook language predicts depression in medical records. Proc. Natl Acad. Sci. USA 115, 11203–11208 (2018).

Kiritchenko, S., Hipson, W. E., Coplan, R. J., & Mohammad, S. M. (2020). SOLO: A corpus of tweets for examining the state of being alone. arXiv preprint arXiv:2006.03096.

Supplementary

Replication of results using the Hu and Liu Valence lexicon (2004).

To validate the findings in the main article, we implemented the exact same approach described within but this time with a different valence lexicon (Hu & Liu, 2004), which categorizes 2006 words as positive and 4783 as negative. Other than that, the score for positive words and negative words were calculated exactly as described in the main text. Note, the Hu and Liu lexicon does not categorize specific emotions, so analysis was only conducted using its Negative and Positive scores.

Valence of webpages provides a marker of mental health. We again conducted two separate mixed ANOVA's for the Positive and Negative score of webpages visited, but using the Hu and Liu Valence Lexicon (described above). In the first mixed ANOVA, psychopathology scores ("Anxious-Depression", "Social-Withdrawal", "Compulsive-Behaviour and Intrusive Thought") was indicated as a within-subjects factor and the Negative Valence score of the webpages that participants browsed (Z-scored) was input as a within-subject modulating factor. Participants' age and gender were also indicated between-subject modulating covariates (both Z-scored). We observed a significant main effect of the Negative Score of webpages that participants browsed on psychopathology scores (**Study 1:** $F(1,284) = 4.083$, $p = 0.044$, partial eta square = 0.014; **Study 2:** $F(1,442) = 6.462$, $p = 0.011$, partial eta square = 0.014). The second mixed ANOVA was identical to the first except that the Positive Valence score was input as a within-subject modulating factor instead of the Negative score. We did not observe a significant main effect of the Positive score of webpages that participants browsed on psychopathology scores (**Study 1:** $F(1,284) = 0.155$, $p = 0.695$ partial eta square = 0.001; **Study 2:** $F(1,442) = 4.168$, $p = 0.042$, partial eta square = 0.009). Together, these results indicate that Negative valence of webpages that people expose themselves to online is indicative of their mental health.

Mood is bi-directionally related to browsing negatively valenced webpages. We again tested whether participants pre-browsing mood was related to the valence of information they browsed but using the Hu and Liu Valence Lexicon (described above). To do this, in Study 1 we ran two separate mixed effect models each including participants pre-browsing mood ratings as fixed and random effects along with age and gender as fixed effect predicting the Negative score and Positive score of webpages visited separately. In Study 2, as we only have 1 observation per participant for each variable of interest, (compared to 5 in Study 1), we ran two simple linear regressions predicting the Negative score and Positive score, separately, from pre-browsing mood ratings, controlling for age and gender. Participants pre-browsing mood was associated with the Negative score of webpages visited (**Study 1:** $\beta = -0.071 \pm 0.041$ (SE), $t(364.49) = -1.720$, trend $p = 0.086$; **Study 2:** $\beta = -0.001 \pm 0.000$ (SE), $t(399) = -2.293$, $p = 0.022$). With regard to participants pre-browsing mood predicting the Positive score of webpages visited, we once again did not observe a significant effect (**Study 1:** $\beta = -0.029 \pm 0.041$ (SE), $t(94.11) = -0.723$, $p = 0.471$; **Study 2:** $\beta = 0.000 \pm 0.001$ (SE), $t(399) = 0.861$, $p = 0.390$).

Next, we tested whether the valence of the webpages that participants browsed had an impact on their mood directly after browsing the internet. To test this, in Study 1, we once again ran two mixed effect models, each predicting post browsing mood ratings from either the Hu and Liu (2004) Negative or Positive score of webpages visited (input as a fixed and random effect). Both models included participants pre-browsing mood ratings as a fixed and random effect, along with age and gender as a fixed effect. In Study 2, we ran two simple linear regressions, both predicting participants post browsing mood ratings from either the Negative or Positive score of webpages visited. Both models included participants pre-browsing mood ratings, age and gender as control variables. We observed that Negative score of webpages visited was related to participants post browsing mood ratings controlling for pre-browsing mood, age and gender (**Study 1:** $\beta = -0.035 \pm 0.018$ (SE), $t(715.58) = -1.977$, $p = 0.048$; **Study 2:** $\beta = -1.413 \pm 0.630$ (SE), $t(399) = -2.244$, $p = 0.025$). In particular, participants expressed worse mood post browsing when they browsed more negatively valenced webpages. We also observed a significant effect of the Positive score of webpages visited on participants post browsing mood ratings using the Hu and Liu Valence Lexicon (2004), which we did not see when using the NRC lexicon (Mohammad, 2018): (**Study 1:** $\beta = 0.036 \pm 0.018$ (SE), $t(604.29) = 1.990$, $p = 0.047$; **Study 2:** $\beta = 1.670 \pm 0.626$ (SE), $t(399) = 2.669$, $p = 0.008$).

Supplementary Table. *Model results for Age and Gender (i.e., controls).*

Full Model	Age ß	Gender ß
Study 1: (Mean Psychopathology ~ Negative score + Age + Gender)	-0.179*** (i.e., younger participants report more psychopathology symptoms).	0.132** (i.e., females report more psychopathology symptoms).
Study 2: (Mean Psychopathology ~ Negative score + Age + Gender)	-0.135*** (i.e., younger participants report more psychopathology symptoms).	0.122*** (i.e., females report more psychopathology symptoms).
Study 1: (Negative score ~ Pre-Mood + Age + Gender)	0.008* (i.e., older participants browse more negative webpages).	-0.141(not significant).
Study 2: (Negative score ~ Pre-Mood + Age + Gender)	-0.071 (not significant).	-0.177*** (i.e., males browse more negative webpages).
Study 1: (Post Mood ~ Negative score + Pre-Mood + Age + Gender)	0.000 (not significant).	-0.080** (i.e., males report better mood after browsing the web).
Study 2: (Post Mood ~ Negative score + Pre-Mood + Age + Gender)	-0.029 (not significant).	-0.070* (i.e., males report better mood after browsing the web).
*** = P <0.001, ** = P <0.01 * = P <0.05		

Fear sentiment of webpages browsed associated with mental health.

To examined if distinct emotions of webpages browsed was associated with mental health we quantified the percentage of *Anger, Fear*, *Anticipation, Trust, Surprise, Sadness, Joy* and *Disgust* associated words (as defined by the NRC Emotion Lexicon; Mohammad and Turney, 2013) out of all words on each webpage participants browsed. For each day separately, we then calculated the average emotion score of the webpages visited by each participant and then averaged these scores across the five days. We then input the eight Emotion scores, along with age and gender into a stepwise regression predicting the mean of the three psychopathology factors (i.e, '*Anxious-Depression*', '*Social-Withdrawal*' and '*Compulsive-Behaviour and Intrusive Thought*). A stepwise regression was ideal in this case as the variance inflation factor (VIF) of some predictor variables was high (e.g., greater than 3). The winning model included the Fear score of webpages (**Study 1:** $\beta = 0.105 \pm 0.042$ (SE), $t(288) = 2.500$, $p = 0.013$, $R = 0.146$, **Figure 3d**; **Study 2:** $\beta = 0.087 \pm 0.034$ (SE), $t(446) = 2.569$, $p = 0.011$, $R = 0.121$, **Figure 3e**) as well as age and gender, suggesting that those with poorer mental health browse more fear related webpages. The Intraclass Correlation Coefficient (ICC) of the Fear scores across the 5 days revealed statistically significant moderate stability (ICC = 0.505, $p < 0.001$), indicating that the tendency to consume text high in fear words is likely due both to trait and state.