Information-Seeking Under Threat: How the Characteristics of Web Searches Changed During the Pandemic

Christopher Kelly (✉ christopher.kelly.16@ucl.ac.uk)

UCL <https://orcid.org/0000-0003-0572-2136>

Bastien Blain

University College London

Tali Sharot

University College London <https://orcid.org/0000-0002-8384-6292>

Article

Keywords:

Posted Date: December 3rd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1089747/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

To adjust to novel and threatening environments people seek information. Here, we examine whether and how a threatening global event --the pandemic-- altered the characteristics of the information people sought out online. An analysis of queries submitted to Google search engine revealed that people were more likely to submit queries for information that could guide action (i.e., *"How to"* and *"How do"* searches) during the pandemic relative to before, controlling for total search volume. This tendency may have contributed to the rapid adaptation observed in response to the pandemic. Indeed, stress levels reported weekly by 17K individuals predicted the proportion of *"How to"* and *"How do"* searches, controlling for COVID-19 related confinement. Markedly, population stress levels were more strongly associated with this high-level feature of web searches than they were with searches for specific terms such as *"anxiety"* or *"stress"*. In contrast, COVID-19 related confinement, but not stress levels, was associated with the proportion of *"What"* and *"Why"* questions submitted to Google, suggesting that the confinement was related to increased desire for general knowledge. Key results were replicated across two countries (UK and US). The study suggests that in situations of high stress people ask questions that can guide action. An intriguing possibility is that tracking of this feature could be used to monitor population stress levels beyond the pandemic.

Main Text

Every person will face unexpected adversities during their lifetime. These events may be global (e.g., war, pandemic) or unique to the individual (e.g., being diagnosed with cancer, losing one's job, divorce). Such events often lead to stress, anxiety, confusion, and a reduced sense of control (Finlay-Jones & Brown, 1981; Francis et al., 2012; McLaughlin et al., 2010; Miloyan et al., 2018; Suls & Mullen, 1981). One available adaptive reaction is to seek information. Information can serve several important functions in such situations; first, information can help guide action to promote adaptation (Hirshleifer & Riley, 1979; Sharot & Sunstein, 2020; Stigler, 1961). For example, during wartime people may search for information on how to secure windows from being shuttered by rockets. Second, information can help people fill "information gaps" that have suddenly become salient (Golman & Loewenstein, 2018). For example, a person diagnosed with cancer may want to learn about the biological mechanism of the disease. While such information may not necessarily be used by the patient to guide action, it can increase their sense of comprehension and in turn reduce uncertainty. Third, people may be inclined to search for positively valenced information (Charpentier et al., 2018; Golman et al., 2017; Karlsson et al., 2009; Kobayashi & Hsu, 2019) to improve mood (Sharot & Sunstein, 2021). For example, recent divorcees may seek data about successful second marriages. Accordingly, by examining the features of the information that people seek, we may gain insight about the problems they are trying to solve, such as planning action, reducing uncertainty and improving mood (Hirshleifer & Riley, 1979; Sharot & Sunstein, 2020; Stigler, 1961).

Here, we examined how an abrupt aversive life event - the COVID-19 pandemic - led to changes in the features of information that people sought online. In March 2020, governments around the world declared a "National Emergency" due to the spread of the COVID-19 virus. Lockdown orders were put in place, restricting people to their homes, and changing the daily lives of most. In the following 12 months, approximately 145 million individuals were infected by the virus and 3 million perished (WHO, 2021). In addition, unemployment rates climbed, and the financial markets experienced an acute dive (Falcettoni et al., 2021; Falk, 2021; Zhang et al., 2020). These factors combined with high uncertainty regarding how the pandemic would unfold, contributed to a population-level increase in stress and anxiety (Globig et al., 2020; Fancourt et al., 2021; Varga et al., 2021).

To examine if and how the type of information people sought was altered in response to this threatening event, we analyzed queries submitted to Google search engine before and during the COVID-19 pandemic in the UK and US. Our investigation was guided by our theory of the key motives of information-seeking. According to this theory, people search for information to fulfill three main aims (Sharot & Sunstein, 2020; Kelly & Sharot, 2021); (i) guide action, (ii) increase comprehension of the world around them, and (iii) experience positive affect. These motives may be weighted differently in different situations. For example, when experiencing threat people may be more inclined to search for information that can guide adaptive action.

We quantified proxies of these features by examining the questions people submitted to the Google search engine. In particular, we first calculated the percentage of Google queries containing the question-words “*How to*” and “*How do*” submitted in each geographical region out of all searches submitted in that region, every week from the date the “National Emergency” was declared (UK: March 23rd, 2020; and US: March 13th, 2020) through March 21st, 2021, as well as every week from January 1st, 2017 until National Emergency was declared. The rationale was that asking “How to” and “How do” will likely result in information that can directly guide action (e.g., “How do you install Zoom?”). In a controlled experiment, we confirmed that participants believe that when asking “*How to*” and “*How do*” questions, people want information that can guide their action. We did the same for “*What*” and “*Why*” questions. The rationale was that such questions are likely to fill information-gaps that can increase people’s general sense of comprehension (e.g., “What is Zoom?”). That rationale was also tested and confirmed in a controlled experiment. Changes to the proportion of these queries cannot be explained by changes in the volume of Google searches during the pandemic, as we examined the change in the *percentage* of these queries out of all searches. Third, we quantified the valence of the most frequent questions submitted to the Google search engine by calculating the difference in the percentage of positive and negative words out of all words entered for these queries each week in the UK and US. This was done by matching the search words to an emotion lexicon database that includes 6,789 words scored on valence (Hu & Liu, 2004).

We then examined (i) how these features changed following the declaration of a “National Emergency” in the UK and US, and (ii) how these changes related to population stress levels in the UK. The latter was done by relating the three measures quantified weekly during the pandemic to weekly stress reports of approximately 17,468 individuals in the UK. In addition, we related the three features of information-seeking to COVID-19 related confinement in each geographical region. This allowed us to dissociate the effects of stress on information-seeking from the effect of COVID-19 related confinement.

Our approach differs substantially from extant research focusing on relating web searches for specific content terms (e.g., “suicide,” “anxiety”) with mental health indicators of a population, a method that has resulted in mixed findings (Ayers et al., 2021; Gunnell et al., 2015; Sueki, 2011; Hoerger et al., 2020; Barros et al., 2019; Ayers et al., 2012; Knipe et al., 2020; McCarthy et al., 2020; McCarthy, 2010; Misiak et al., 2020; Rana, 2020; Sinyor et al., 2014; Tran et al., 2017; Arora et al., 2019; Yang et al., 2010). If successful, our novel approach of quantifying high-level features of web searches to stress, may provide a new avenue for monitoring population-level stress during times of crisis.

Results

The pandemic resulted in significant changes to high-level features of web queries. To assess the high-level features of web searches we first extracted the Google search volume index of “*How to*”, “*How do*”, “*What*” and

“Why” questions separately. A Google search volume index is equal to the number of searches for the specific term of interest in a given week and region divided by the total number of searches for that same week and region. These percentages are then normalized to represent search interest relative to the highest percent for that region for the entire time frame (i.e., January 1st, 2017 - March 21st, 2021; **see Method for details**). A Google search volume index of 100 denotes the peak popularity for the term, while an index of 0 means there was not enough search data to calculate popularity of this term. Note, that weekly changes to the Google search volume index cannot be explained by weekly changes in the total volume of Google searches, as the index reflects the *percent* of specific queries out of all searches that week. We then averaged separately the (i) *“How to”* and *“How do”* scores and the (ii) *“What”* and *“Why”* scores. We also calculated a third feature - a Valence index - which indicates the valence of the most frequent questions submitted to Google search engine. To calculate this index, we extracted the most frequent search queries that included the words *“How”*, *“What”* or *“Why”* and calculated the difference in the proportion of positive words to negative words from the terms submitted using a library of words that have been categorized as positive or negative (Hu & Liu, 2004). We transformed this number to be on a scale from 0 to 100, such that it would be easily comparable to the other two indexes. A score of 100 denotes the most positively valenced score and a score of 0 the most negatively valenced score.

In a separate study (**see Supplementary Experiment**), we validated the proposition that when people submit *“How to”* and *“How do”* search queries they are primarily motivated to find information that can guide action and when submitting *“What”* and *“Why”* questions they are primarily motivated to increase their general understanding.

Analyses were conducted separately in the UK and the US. In each country we quantified the measures above for every week from the date the “National Emergency” was declared (UK: March 23rd, 2020; and US: March 13th, 2020) through March 21st, 2021, as well as every week from January 1st, 2017 until National Emergency was declared. There was a significant increase in the Google search volume index of (i) *“How to”* and *“How do”* questions and (ii) *“What”* and *“Why”* questions following the declaration of a “National Emergency” relative to the three years previous (*“How to”* and *“How do”* UK: before “National Emergency” declared: $M = 61.63$, $SD = 2.46$, after “National Emergency” declared: $M = 79.83$, $SD = 8.49$, $t(53.67) = -15.257$, $p = 0.0001$; *“How to”* and *“How do”* US: before “National Emergency” declared: $M = 63.67$, $SD = 1.72$, after “National Emergency” declared: $M = 76.59$, $SD = 5.46$, $t(56.42) = -17.109$, $p = 0.0001$; *“What”* and *“Why”* UK: before “National Emergency” declared: $M = 60.03$, $SD = 4.25$, after “National Emergency” declared: $M = 81.26$, $SD = 8.61$, $t(58.90) = -17.146$, $p = 0.0001$; *“What”* and *“Why”* US: before “National Emergency” declared: $M = 70.84$, $SD = 5.81$, after “National Emergency” declared: $M = 88.25$, $SD = 6.26$, $t(219) = -18.775$, $p = 0.0001$). In contrast, there was no difference in the Valence index before and after National Emergency was declared. (Valence index UK: before “National Emergency” declared: $M = 57.80$, $SD = 12.17$, after the “National Emergency” declared: $M = 57.38$, $SD = 9.03$, $t(218) = 0.231$, $p = 0.818$; Valence index US: before “National Emergency” declared: $M = 41.74$, $SD = 12.42$, after “National Emergency” declared: $M = 39.75$, $SD = 12.93$, $t(218) = 1.002$, $p = 0.318$; **Figure 1a-d**).

Plotting the three measures over time (**Figure 2**) reveals a sharp increase in *“How to”* and *“How do”* questions and *“What”* and *“Why”* questions submitted following the declaration of a “National Emergency”. In the US a second peak is observed around October 2020, a time in which COVID cases were again rising (Dong et al, 2021). In contrast, there was no change to the valence of searches in both countries.

Population stress-levels are selectively associated with asking *“How to”* and *“How do”*. Thus far, we have shown that there is an increase in the proportion of *“How to”* and *“How do”* and *“What”* and *“Why”* searches submitted to

the Google search engine during the pandemic relative to before. We next examined whether these were related to population stress levels. We had access to self-report stress levels collected every week in the UK between March 21st, 2020 and March 21st, 2021. Approximately 70,000 unique individuals completed the survey, on average 17,468 individuals a week in the UK (Fancourt et al., 2021). Specifically, participants were asked to indicate if over the previous week they felt worried and/or stressed about any of the following factors: (i) catching Covid-19, (ii) becoming seriously ill from Covid-19, (iii) finance, (iv) unemployment and (v) getting food. We computed the mean proportion of individuals who reported stress or worry over these factors. We then conducted three linear models predicting on a weekly basis the Google search volume index in the UK of “*How to*” and “*How do*” questions, “*What*” and “*Why*” questions, and the valence of submitted questions from stress levels. We also included weekly UK COVID-19 confinement scores in the models to disentangle the effects of emotional stress on web searches from the effects of confinement due to restrictions placed by the Government. Covid-19 related confinement data for each week in the UK was obtained from the Oxford University COVID-19 Government Response Tracker (Webster et al., 2021). The data includes ordinal variables coded by severity/intensity of confinement, on a daily basis (from January 1st, 2020 and March 21st, 2021), due to the following: (i) school and university closures, (ii) workplace closures, (iii) public event cancelations, (iv) restrictions on gatherings, (v) public transport restrictions, (vi) stay at home requirements, (vii) restrictions on domestic travel, and (viii) restrictions on international travel; **see Table 3 for coding**. To obtain weekly values, we computed weekly averages of the daily ratings. Finally, to quantify an overall COVID-19 related confinement score, we normalised all variables to range between 0 and 1 and then averaged the 8 transformed variables together.

Importantly, to account for simple temporal trends we first removed linear trend from the dependent variables (“*How to*” and “*How do*” questions, “*What*” and “*Why*” questions, and the Valence Index) and predictor variables (stress scores and COVID-19 related confinement), using the `detrend` function in the ‘`pracma`’ R package. The detrended dependent and predictor variables were then Z-scored before being entered in the linear models.

The linear model predicting “*How to*” and “*How do*” questions from stress levels and COVID-19 related confinement scores, revealed that both high stress ($\beta = 0.211 \pm 0.075$ (SE), $t(49) = 2.817$, $p = 0.007$) and greater COVID-19 related confinement ($\beta = 0.776 \pm 0.075$ (SE), $t(49) = 10.346$, $p = 0.0001$ **Figure 3a**) predicted proportion of “*How to*” and “*How do*” queries (R^2 of the model = 0.742). In other words, the relationship between stress levels and “*How to*” and “*How do*” searches cannot be solely explained by increased restrictions on movement during the pandemic, as our model controls for COVID-19 related confinement. In contrast, the proportion of “*What*” and “*Why*” questions was selectively predicted by COVID-19 related confinement ($\beta = 0.676 \pm 0.105$ (SE), $t(49) = 6.405$, $p = 0.0001$), but not stress levels ($\beta = 0.106 \pm 0.105$ (SE), $t(49) = 1.005$, $p = 0.320$), overall $R^2 = 0.492$ (**Figure 3b**). Valence of searches was not predicted by either variables (stress: $\beta = 0.238 \pm 0.147$ (SE), $t(49) = 1.621$, $p = 0.112$, COVID-19 related confinement: $\beta = -0.044 \pm 0.147$ (SE), $t(49) = -0.300$, $p = 0.766$, $R^2 = 0.013$; **Figure 3c**).

Thus far, we have shown that the relative volume of queries that can direct action are tightly related to stress levels. Next, we wanted to test the predictive validity of this simple model. Specifically, we used the proportion of UK sample reporting stress to predict the proportion of “*How to*” and “*How do*” searches using a leave one out analysis. To account for a simple temporal trend, we first removed the linear trend from the dependent variable (“*how to*” and “*how do*” questions) and the predictor variable (stress levels). The detrended predictor variables were then Z-scored before being entered in the simple linear model. The simple model was then run on all the data save for one time point which was held out from the analysis. We then used the regression beta to predict the proportion of “*How to*” and “*How do*” searches of the left-out time point. This process was repeated so that each

week's proportion of *"How to"* and *"How do"* searches was estimated from the simple model parameters generated without using that week to fit the data. The actual proportion of *"How to"* and *"How do"* searches of a week (data) and the predicted proportion of *"How to"* and *"How do"* searches (estimation) were then correlated and also compared using a paired sample t-test. This analysis indicates whether the stress levels in the UK is a good predictor of the proportion of *"How to"* and *"How do"* searches. We observed a correlation between the predicted proportion of *"How to"* and *"How do"* searches (estimate) and the actual proportion of *"How to"* and *"How do"* searches (data) ($r(50) = 0.401$, $p = 0.003$). The means of the two sets of values were not significantly different from one another ($t = 0.098$, $p = 0.922$). This analysis suggests that stress levels in a population is a good predictor of the proportion of *"How to"* and *"How do"* searches during the pandemic.

Finally, we tested whether stress was better predicted by *"How to"* and *"How do"* Google searches than searches for specific content terms (e.g., "stress" and "anxiety"), which are often used in attempt to predict population well-being (Yang et al., 2015; Barros et al., 2019). To test this, we ran a model predicting the proportion of UK sample reporting stress from Google search index of *"How to"* and *"How do"* questions as well as Google search index for "stress" and closely associated terms (i.e., "anxiety" and "mental health"). Once again, the dependent and predictor variables were first detrended and then Z-scored. The volume of *"How to"* and *"How do"* questions was the only significant predictor of stress ($\beta = 1.751 \pm 0.497$ (SE), $t(47) = 3.520$, $p = 0.001$; **Figure 4**). All other predictors were not significant (p -value > 0.404). Similar results are observed if multiple linear models are run to predict stress each time from only one term. Once again, *"How to"* and *"How do"* Google search index was the strongest predictor of stress ($\beta = 1.748 \pm 0.478$ (SE), $t(50) = 3.658$, $p = 0.0006$), while no other predictor was significant (all $p > 0.271$).

Note, that in all linear models we control for temporal trends by detrending the dependent and predictive variables. The same results are observed when detrending is not employed (**see Supplementary Table**).

Discussion

The global pandemic generated a new set of practical and mental challenges. To overcome these challenges, people turned to technology. On average, people spent almost 7 hours a day online in 2020, up 7.3% from the previous year (Kemp, 2021). A large fraction of this time was dedicated to searching for and consuming information (Kemp, 2021). Here, we examined how the high-level features of questions that people submitted to the Google search engine changed in response to the pandemic. Our results reveal two significant changes to web searches in response to the pandemic.

First, we observed a sharp increase in searches for information that can guide action. In particular, both in the US and UK, the proportion of queries that included the question-words *"How to"* and *"How do"* were greater during the 12 months following the declaration of "National Emergency" than in the years prior. This rise likely reflects an adaptive human tendency to ask questions that can facilitate rapid adjustment to new and potentially aversive environments. Our supplementary experiment indeed shows that participants believe that when people submit *"How to"* and *"How do"* search queries, they are primarily motivated to find information that can guide action (**see Supplementary Experiment**).

The fluctuation in proportion of *"How to"* and *"How do"* questions submitted to Google was associated with the proportion of individuals who reported experiencing COVID-related stress in the UK. Specifically, the greater the proportion of the sample who reported stress, the greater the proportion of *"How to"* and *"How do"* questions

submitted out of all searches, controlling for temporal trends in the data. This was observed using a simple model relating the weekly proportion of *“How to”* and *“How do”* questions submitted to Google search engine to the proportion of individuals in the UK reporting COVID- related stress out of a sample of over 17K residences. Importantly, we were able to disentangle the effects of stress on web searches from the effects of COVID-related confinement, by controlling for those in the model. Thus, the relationship between stress levels and the volume of *“How to”* and *“How do”* queries cannot be explained simply by increased confinement. The findings show that in the face of a novel threatening situation and high stress people search for information that can help guide action. While in the past such information may have been sought directly from other people, with the development of the internet, individuals are now able to turn to the web for answers. This ability may have contributed to the high resilience and quick adaptation observed in response to the pandemic (Aknin et al., 2021.; Globig et al., 2020).

Second, we observed a sharp increase in the proportion of queries that included the question-words *“What”* and *“Why”* during the global pandemic. Our supplementary study shows that participants believed that when people submit *“What”* and *“Why”* questions to the Google search engine they are primarily motivated to increase their general understanding, rather than have an intention to guide action (**see Supplementary Experiment**). Interestingly, this spike was not associated with stress levels. Instead, fluctuations in the proportion of *“What”* and *“Why”* questions were associated only with COVID-19 related confinement. Specifically, the more severe the levels of confinement imposed on people the more they asked *“What”* and *“Why”* questions relative to other searches. This rise may thus be due to an increase in down-time due to restrictions on activities outside the home. It is important to emphasize that any change in the proportion of *“What”* and *“Why”* searches and *“How to”* and *“How do”* searches cannot simply be explained by a general increase in number of Google searches, as the former are calculated as proportion of the latter. Neither can it be explained by temporal linear trends, as the data was detrended. We did not observe a change in the number of negative words used in questions submitted to the Google search engine during the pandemic relative to the number of positive words used.

Markedly, we show that the type of questions people asked, predicted the proportion of UK sample reporting COVID-related stress during the pandemic better than measuring the frequency of searches that include stress related content (i.e., *“anxiety”*, *“stress”*, and *“mental health”*). An interesting question is whether tracking the frequency of *“How to”* and *“How do”* questions, or other high-level features of web queries, can (i) predict population-level stress beyond the time of a pandemic and (ii) predict individual differences in stress. If affirmative, quantifying such search features can prove extremely valuable for monitoring mental health on both an individual and populating level. Future studies are needed to explore these intriguing possibilities.

Methods

Data Extraction

Web Search Data. Weekly search data was extracted from Google Trends (www.google.com/trends) for 220 weeks (January 1st, 2017 through March 21st, 2021). This was done separately for the UK and the US. We extracted the Google search volume index for four question-words (*“How to”*, *“How do”*, *“What”*, and *“Why”*) separately. A Google search volume index value is equal to the number of searches for the specific term of interest in a given week and region (for example total number of searches that include the question-words *“How to”* in the UK the first week of 2020) divided by the total number of searches in that same time and region (for example the total number of Google searches submitted in UK the first week of 2020). These values are normalized to represent

search interest relative to the highest value for that region for the entire time frame (i.e., January 1st, 2017 – March 21st, 2021). A value of 100 is the peak popularity for the term, a value of 50 means that the term is half as popular as the peak, while a score of 0 means there was not enough data to calculate the terms' popularity. We then averaged separately the (i) "*How to*" and "*How do*" scores and the (ii) "*What*" and "*Why*" scores.

To quantify valence, we extracted the 25 most popular search queries for each week and region for questions including "*How*", "*Why*" and "*What*". That is, for each week we extract 25 search queries per type of question (i.e., 75 total search queries for each week), as this is the maximum Google Trends reports. We then matched the words on the lists with a lexicon database (Hu & Liu, 2004) that includes 2006 positive words and 4783 negative words and count the number of positive and negative words for each week and region. Finally, we subtracted the proportion of negative words from the proportion of positive words. We transformed this number to be on a scale from 0 to 100, such that it would be easily comparable to the other two indexes. A score of 100 denotes the most positively valenced value and a score of 0 the most negatively valenced value.

Self-Reported Stress. Data was extracted, with permission, from the UK COVID-19 Social Study (Fancourt et al., 2021). The study is a panel study of over 70,000 UK citizens which aims to characterize the psychological and social experience of adults living in the UK during the Covid-19 pandemic (**see Table 1 for demographics**). The study commenced as a weekly survey, with participants receiving an invitation to the next wave of data collection 7 days following their last completion. All participants received up to 2 reminders (24 and 48 hours following their initial weekly invitation). The link to their last reminder remained live so they could return to the study a few days later if they chose to. Following week 22 of the study, monthly follow-ups rather than weekly follow-ups were sent. To attain an equal number of responses across time, participants were randomized to receive their monthly invitation on either week 1,2,3 or 4 of the month, with subsequent invitations following 28 days after they completed the survey. An average of 17,468 individuals submitted data each week (**see Table 2 for response frequency for each week**). For full methods and demographics for the sample see www.COVIDSocialStudy.org. The UK COVID-19 Social Study was approved by the UCL Research Ethics Committee and all participants gave written informed consent.

Participants were asked: "*over the past week, have any of the following been worrying you at all, even if only in a minor way?*" They were presented with 18 factors that may cause worry (for example internet access, boredom, neighbours) and were to pick any that they were worried about. Five of these factors were a-priori categorized by the authors of the survey (Fancourt et al., 2021) as ones that have been impacted by COVID. These were (i) catching Covid-19 (ii) becoming seriously ill from Covid-19, (iii) finances, (iv) losing your job/unemployment and (v) getting food. Second, they were asked "*have any of these things been causing you significant stress? (e.g., they have been constantly on your mind or have been keeping you awake at night)*". They were presented with the same 18 factors as above and were asked to tick any of those causing significant stress. For each week and factor, Fancourt et al., 2021 calculated the proportion of respondents that ticked that factor either in response to question 1 and/or question 2. Factors i and ii were a-priori combined by Fancourt and colleagues to make one factor, leaving us with four factors. For each week the proportion of people ticking 1 and/or 2 were averaged across the four factors to produce one indicator of "stress levels" for that week.

Table 1 shows the demographic of respondents to the UK COVID-19 Social Study. Importantly, data points reported by Fancourt et al., (2021) were weighted using auxiliary weights to the national census and Office for

National Statistics (ONS) data. We used these weighted data points in our study. Thus, reported stress levels are representative of the UK population.

Table 1. *Demographics of respondents in the UK COVID-19 Social Study (adapted from Fancourt et al., 2021). Data in the Fancourt et al., (2021) study and in our study are weighted using auxiliary weights to the national census and Office for National Statistics (ONS) data.*

		Number of observations	%		Number of observations	%
Age		Education levels				
	18-29	51,858	5.77	GCSE or below	126,427	14.1
	30-59	493,016	54.9	A-levels of equivalent	154,954	17.3
	60+	353,559	39.4	Degree or above	617,052	68.7
Gender		Any diagnosed mental health conditions				
	Male	225,578	25.2	No	748,416	83.3
	Female	669,279	74.8	Yes	150,017	16.7
Ethnicity		Any diagnosed physical health conditions				
	White	860,157	96.0	No	516,884	57.5
	Ethnic minority	35,455	3.96	Yes	381,549	42.5
UK nations		Keyworker				
	England	725,705	81.6	No	711,201	79.2
	Wales	108,598	12.2	Yes	187,232	20.8
	Scotland	55,416	6.23	Living with children		
Living arrangement				No (excluding those who live alone)	510,650	72.0
	Not living alone	709,289	79.0	Yes	198,639	28.0
	Living alone	189,144	21.1	Living area		
Annual household income				Village/hamlet/isolated dwelling	225,022	25.1
	>30k	482,268	59.6	City/large town/small town	673,411	75.0
	<30k	327,187	40.4			

Table 2. *The total number of participants providing data during each calendar week in the UK COVID-19 Social Study (adapted from Fancourt et al., 2021).*

Date	Week	Freq	Date	Week	Freq
21/03-27/03/2020	1	28,929	19/09/20-25/09/20	27	8,318
28/03-03/04/2020	2	27,873	26/09/20-02/10/20	28	8,366
04/04-10/04/2020	3	38,151	03/10/20-09/10/20	29	8,501
11/04-17/04/2020	4	38,453	10/10/20-16/10/20	30	8,072
18/04-24/04/2020	5	38,504	17/10/20-23/10/20	31	7,495
25/04-01/05/2020	6	36,513	24/10/20-30/10/20	32	7,612
02/05-08/05/2020	7	36,651	31/10/20-06/11/20	33	7,830
09/05-15/05/2020	8	37,549	07/11/20-13/11/20	34	7,443
16/05-22/05/2020	9	35,702	14/11/20-20/11/20	35	6,995
23/05-29/05/2020	10	33,293	21/11/20-27/11/20	36	7,078
30/05-05/06/2020	11	32,196	28/11/20-04/12/20	37	7,190
06/06-12/06/2020	12	31,304	05/12/20-11/12/20	38	6,947
13/06-19/06/2020	13	30,229	12/12/20-18/12/20	39	6,473
20/06-26/06/2020	14	29,153	19/12/20-25/12/20	40	6,240
27/06-03/07/2020	15	28,534	26/12/20-01/01/21	41	6,966
04/07-10/07/2020	16	27,552	02/01/21-08/01/21	42	7,038
11/07-17/07/2020	17	26,737	09/01/21-15/01/21	43	6,274
18/07-24/07/2020	18	25,983	16/01/21-15/01/21	44	6,219
25/07-31/07/2020	19	25,005	23/01/21-29/01/21	45	6,540
01/08-07/08/2020	20	24,530	30/01/21-05/02/21	46	6,831
08/08-14/08/2020	21	23,851	06/02/21-12/02/21	47	6,048
15/08-21/08/2020	22	23,120	13/02/21-19/02/21	48	6,217
22/08-28/08/2020	23	11,373	20/02/21-26/02/21	49	6,111
29/08-20-04/09/20	24	10,025	27/02/21-05/03/21	50	6,574
1205/09/20-11/09/20	25	9,916	06/03/21-12/03/21	51	8,683
12/09/20-18/09/20	26	10,009	13/03/21-19/03/21	52	9,128

COVID-19 Confinement Score. To measure COVID-19 related confinement, we extracted eight confinement variables from a publicly available dataset (The Oxford University COVID-19 Government Response Tracker; Webster et al., 2021). All variables are ordinal coded by severity/intensity of confinement, on a daily basis (from January 1st, 2020 to March 21st, 2021), for the following: (i) *school and university closures*, (ii) *workplace closures*, (iii) *public event cancelations*, (iv) *restrictions on gatherings*, (v) *public transport restrictions*, (vi) *stay at home requirements*, (vii) *restrictions on domestic travel*, and (viii) *restrictions on international travel*; **see Table 3 for coding**. To obtain weekly values, we computed weekly averages of the daily ratings. To quantify an overall COVID-19 related confinement score, we transformed all variables to range between 0 and 1 using the R function *scaler* from the R package, *bruceR*. Finally, we averaged the 8 transformed variables together.

Table 3. *Coding of COVID-19 related confinement variables (adapted from Webster et al, 2021).*

Variable Description	Coding instructions
<i>Closings of schools and universities</i>	<p>0 - No measures</p> <p>1 – Recommend closing, or all schools open with alterations resulting in significant differences compared to usual, non-Covid-19 operations</p> <p>2 - Require closing (only some levels or categories, e.g. just high school, or just public schools)</p> <p>3 - Require closing all levels</p>
<i>Closings of workplaces</i>	<p>0 - No measures</p> <p>1 - recommend closing (or work from home)</p> <p>2 - require closing (or work from home) for some sectors or categories of workers</p> <p>3 - require closing (or work from home) all-but-essential workplaces (E.g., grocery stores, doctors)</p>
<i>Cancelling public events</i>	<p>0 – No measures</p> <p>1 - Recommend cancelling</p> <p>2 - Require cancelling</p>
<i>Cut-off size for bans on gatherings</i>	<p>0 - No restrictions</p> <p>1 - Restrictions on very large gatherings (the limit is above 1000 people)</p> <p>2 - Restrictions on gatherings between 101-1000 people</p> <p>3 - Restrictions on gatherings between 11-100 people</p> <p>4 - Restrictions on gatherings of 10 people or less</p>
<i>Closing of public transport</i>	<p>0 - No measures</p> <p>1 - Recommend closing (or significantly reduce volume/route/means of transport available)</p> <p>2 - Require closing (or prohibit most citizens from using it)</p>
<i>Orders to “shelter-in-place” and otherwise confine to home.</i>	<p>0 - No measures</p> <p>1 - recommend not leaving house</p> <p>2 - require not leaving house with exceptions for daily exercise, grocery shopping, and ‘essential’ trips</p> <p>3 - Require not leaving house with minimal exceptions (E.g., allowed to leave only once a week, or only one person can leave at a time, etc.)</p>
<i>Restrictions on internal movement</i>	<p>0 - No measures</p> <p>1 - Recommend not to travel between regions/cities</p> <p>2 – internal movement restrictions in place</p>
<i>Restrictions on international travel</i>	<p>0 - No measures</p> <p>1 - Screening</p>

- 2 - Quarantine arrivals from high-risk regions
- 3 - Ban on arrivals from some regions
- 4 - Ban on all regions or total border closure

Analysis

Analysis was conducted separately for data from the UK and the US. In each region we quantified the measures above every week from the date the “National Emergency” was declared due to the COVID-19 pandemic, (March 23rd, 2020 in the UK and March 13th, 2020 in the US) through March 21st, 2021, as well as every week dating back to January 1st, 2017. We then compared the weekly scores before the “National Emergency” to that after using an independent samples t-test.

To assess whether our measures were related to the proportion of UK sample reporting stress, we conducted three linear models predicting on a weekly basis Google’s search volume index of “How to” and “How do” questions, “What” and “Why” questions, and the Valence index of questions submitted to Google in the UK, from UK stress levels. We also included weekly COVID-related confinement scores in the models to disentangle the effects of stress from the effects of confinement due to restrictions placed by the Government. To account for simple temporal trends, we removed the linear trend from the dependent and predictor variables first, using the *detrend* function in the *pracma* R package. The detrended dependent and predictor variables were then Z-scored. Finally, we were interested in whether stress was better predicted by “How to” and “How do” Google searches than searches for stress related terms. To test this, we first removed the linear trend from the dependent and predictor variables, and then Z-scored the dependent and predictor variables. We then ran a model predicting the proportion of UK sample reporting COVID-related stress from the “How to” and “How do” Google search index as well as Google search index for the words “stress”, “anxiety”, and “mental health” in the UK. In addition, we ran multiple linear models to predict the proportion of UK sample reporting COVID-related stress each time from only one of the terms above.

Next, we tested the predictive validity of a simple model using stress levels to predict the proportion of “How to” and “How do” searches using a leave one out analysis. Once again, we removed the linear trend from the dependent and predictor variables first, and then the dependent and predictor variables were Z-scored. Specifically, the simple model was run on all the data save for one time point which was held out from the analysis. We then used the regression beta to predict the proportion of “How to” and “How do” searches of the left-out time point. This process was repeated so that each week’s proportion of “How to” and “How do” searches was estimated from the simple model parameters generated without using that week to fit the data. This resulted in two values for the proportion of “How to” and “How do” searches for each week: the actual proportion of “How to” and “How do” searches (data) and the predicted value from the leave-one-out validation (estimate). The actual proportion of “How to” and “How do” searches of a week (data) and the predicted proportion of “How to” and “How do” searches (estimation) were then correlated and also compared using a paired sample t-test. This analysis indicates whether the population stress levels is a good predictor of the proportion of “How to” and “How do” searches.

Declarations

Data availability

Data and code are available at a dedicated Github repository [github.com/affective-brain-lab/information_seeking_under_threat]. The source data underlying Figs. 1-4 and Supplementary Fig 2 are provided as part of this repository. A reporting summary for this Article is available as a Supplementary Information file.

Code availability

Code supporting this study are available at a dedicated Github repository [github.com/affective-brain-lab/information_seeking_under_threat].

Acknowledgments

We thank Irene Cogliati Dezza, Liron Rozenkrantz, Moshe Glickman, Laura Globig, Valentina Vellani, Sarah Zheng, Gaia Molinaro and Christina Maher 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

- Aroua, V. S., McKee, M., & Stuckler, D.(2019). Google Trends: Opportunities and limitations in health and health policy research. *Health Policy (Amsterdam, Netherlands)*, 123(3), 338–341.
<https://doi.org/10.1016/J.HEALTHPOL.2019.01.001>
- Aknin, L., Neve, J.-E. De, Dunn, E., Fancourt, D., Goldberg, E., Helliwell, J., Jones, S. P., Karam, E., Layard, R., Lyubomirsky, S., Rzepa, A., Saxena, S., Thornton, E., VanderWeele, T., Whillans, A., zaki, jamil, Caman, O. K., & Amour, Y. Ben. (n.d.). *Mental Health During the First Year of the COVID-19 Pandemic: A Review and Recommendations for Moving Forward*. <https://doi.org/10.31234/OSF.IO/ZW93G>
- Ayers, J. W., Althouse, B. M., Allem, J. P., Childers, M. A., Zafar, W., Latkin, C., ... & Brownstein, J. S. (2012). Novel surveillance of psychological distress during the great recession. *Journal of Affective Disorders*, 142(1–3), 323–330. <https://doi.org/10.1016/J.JAD.2012.05.005>
- Ayers, J. W., Poliak, A., Johnson, D. C., Leas, E. C., Dredze, M., Caputi, T., & Nobles, A. L. (2021). Suicide-Related Internet Searches During the Early Stages of the COVID-19 Pandemic in the US. *JAMA Network Open*, 4(1). <https://doi.org/10.1001/JAMANETWORKOPEN.2020.34261>
- Barros, J. M., Melia, R., Francis, K., Bogue, J., O’Sullivan, M., Young, K., ... & Duggan, J. (2019). The Validity of Google Trends Search Volumes for Behavioral Forecasting of National Suicide Rates in Ireland. *International Journal of Environmental Research and Public Health*, 16(17). <https://doi.org/10.3390/IJERPH16173201>
- Blain, B., Globig, L. K., & Sharot, T. (n.d.). *Income shock increases delay discounting independently of emotion*. <https://doi.org/10.31234/OSF.IO/YHM3B>

- 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 of the United States of America*. <https://doi.org/10.1073/pnas.1800547115>
- Falcettoni, E., & Nygaard, V. M. (2021). A literature review on the impact of increased unemployment insurance benefits and stimulus checks in the United States. *Covid Economics*, 64, 186-201.
- Falk, G. (2020). *Unemployment Rates During the COVID-19 Pandemic*. Congressional Research Service.
- Fancourt, D., Bu, F., Mak, H. W., & Steptoe, A. (2021). COVID-19 social study. *Results release*, 25.
- Fancourt, D., Steptoe, A., & Bu, F. (2021). Trajectories of anxiety and depressive symptoms during enforced isolation due to COVID-19 in England: a longitudinal observational study. *The Lancet Psychiatry*, 8(2), 141-149.
- Finlay-Jones, R., & Brown, G. W. (1981). Types of Stressful Life Event and the Onset of Anxiety and Depressive Disorders. *Psychological Medicine*, 11(4), 803–815. <https://doi.org/10.1017/S0033291700041301>
- Francis, J. L., Moitra, E., Dyck, I., & Keller, M. B. (2012). The impact of stressful life events on relapse of generalized anxiety disorder. *Depression and Anxiety*, 29(5), 386–391. <https://doi.org/10.1002/da.20919>
- Globig, L., Blain, B., & Sharot, T. (2020). *When Private Optimism meets Public Despair: Dissociable effects on behavior and well-being*. <https://doi.org/10.31234/osf.io/gbdn8>
- Golman, R., Hagmann, D., & Loewenstein, G. (2017). Information avoidance. In *Journal of Economic Literature*. <https://doi.org/10.1257/jel.20151245>
- Golman, R., & Loewenstein, G. (2018). Information gaps: A theory of preferences regarding the presence and absence of information. *Decision*. <https://doi.org/10.1037/dec0000068>
- Gunnell, D., Derges, J., Chang, S. Sen, & Biddle, L. (2015). Searching for suicide methods. *Crisis*, 36(5), 325–331. <https://doi.org/10.1027/0227-5910/A000326>
- Sueki, H. (2011). Does the volume of Internet searches using suicide-related search terms influence the suicide death rate: data from 2004 to 2009 in Japan. *Psychiatry and Clinical Neurosciences*, 65(4), 392–394. <https://doi.org/10.1111/J.1440-1819.2011.02216.X>
- Hirshleifer, J., & Riley, J. G. (1979). The Analytics of Uncertainty and Information- An Expository Survey. *Source Journal of Economic Literature Journal of Economic Literature*.
- Hoerger, M., Alonzi, S., Perry, L. M., Voss, H. M., Easwar, S., & Gerhart, J. I. (2020). Impact of the COVID-19 pandemic on mental health: Real-time surveillance using google trends. *Psychological Trauma: Theory, Research, Practice, and Policy*, 12(6), 1–2. <https://doi.org/10.1037/TRA0000872>
- Hu, M., & Liu, B. (2004). *Mining and Summarizing Customer Reviews*.

- Karlsson, N., Loewenstein, G., & Seppi, D. (2009). The ostrich effect: Selective attention to information. *Journal of Risk and Uncertainty*. <https://doi.org/10.1007/s11166-009-9060-6>
- Kelly, C. A., & Sharot, T. (in press). Individual differences in information-seeking. *Nature Communications*.
- Kemp, S. (2021). *Digital 2021: Global Overview Report – DataReportal – Global Digital Insights*. <https://datareportal.com/reports/digital-2021-global-overview-report>
- Kobayashi, K., & Hsu, M. (2019). Common neural code for reward and information value. *Proceedings of the National Academy of Sciences*, 116(26), 13061-13066.
- Knipe, D., Evans, H., Marchant, A., Gunnell, D., & John, A. (2020). Mapping population mental health concerns related to COVID-19 and the consequences of physical distancing: a Google trends analysis. *Wellcome Open Research* 2020 5:82, 5, 82. <https://doi.org/10.12688/wellcomeopenres.15870.2>
- McCarthy, M. J. (2010). Internet Monitoring of Suicide Risk in the Population. *Journal of Affective Disorders*, 122(3), 277. <https://doi.org/10.1016/J.JAD.2009.08.015>
- McLaughlin, K. A., Conron, K. J., Koenen, K. C., & Gilman, S. E. (2010). Childhood adversity, adult stressful life events, and risk of past-year psychiatric disorder: A test of the stress sensitization hypothesis in a population-based sample of adults. *Psychological Medicine*, 40(10), 1647–1658. <https://doi.org/10.1017/S0033291709992121>
- Miloyan, B., Joseph Bienvenu, O., Brilot, B., & Eaton, W. W. (2018). Adverse life events and the onset of anxiety disorders. *Psychiatry Research*, 259, 488–492. <https://doi.org/10.1016/j.psychres.2017.11.027>
- Misiak, B., Szcześniak, D., Koczanowicz, L., & Rymaszewska, J. (2020). The COVID-19 outbreak and Google searches: Is it really the time to worry about global mental health? *Brain, Behavior, and Immunity*, 87, 126. <https://doi.org/10.1016/J.BBI.2020.04.083>
- Nuti, S. V., Wayda, B., Ranasinghe, I., Wang, S., Dreyer, R. P., Chen, S. I., & Murugiah, K. (2014). The use of google trends in health care research: a systematic review. *PloS One*, 9(10). <https://doi.org/10.1371/JOURNAL.PONE.0109583>
- Rana, U. (2020). Elderly suicides in India: An emerging concern during COVID-19 pandemic. *International Psychogeriatrics*, 32(10), 1251–1252. <https://doi.org/10.1017/S1041610220001052>
- Sharot, T., & Sunstein, C. R. (2020). How people decide what they want to know. *Nature Human Behaviour*. <https://doi.org/10.1038/s41562-019-0793-1>
- Sinyor, M., Spittal, M. J., & Niederkrotenthaler, T. (2020). Changes in Suicide and Resilience-related Google Searches during the Early Stages of the COVID-19 Pandemic. *Canadian Journal of Psychiatry. Revue Canadienne de Psychiatrie*, 65(10), 741–743. <https://doi.org/10.1177/0706743720933426>
- Stigler, G. J. (1961). The Economics of Information. *Journal of Political Economy*. <https://doi.org/10.1086/258464>

- Suls, J., & Mullen, B. (1981). Life Events, Perceived Control and Illness: The Role of Uncertainty. *Journal of Human Stress*, 7(2), 30–34. <https://doi.org/10.1080/0097840X.1981.9936824>
- Tran, U. S., Andel, R., Niederkrotenthaler, T., Till, B., Ajdacic-Gross, V., & Voracek, M. (2017). Low validity of Google Trends for behavioral forecasting of national suicide rates. *PLOS ONE*, 12(8), e0183149. <https://doi.org/10.1371/JOURNAL.PONE.0183149>
- Varga, T. V., Bu, F., Dissing, A. S., Elsenburg, L. K., Bustamante, J. J. H., Matta, J., ... & Rod, N. H. (2021). Loneliness, worries, anxiety, and precautionary behaviours in response to the COVID-19 pandemic: a longitudinal analysis of 200,000 Western and Northern Europeans. *The Lancet Regional Health-Europe*, 2, 100020
- WHO. (2021). *WHO Coronavirus (COVID-19) Dashboard With Vaccination Data*. <https://covid19.who.int/>
- Yang, A. C., Huang, N. E., Peng, C.-K., & Tsai, S.-J. (2010). Do Seasons Have an Influence on the Incidence of Depression? The Use of an Internet Search Engine Query Data as a Proxy of Human Affect. *PLOS ONE*, 5(10), e13728. <https://doi.org/10.1371/JOURNAL.PONE.0013728>
- Zhang, D., Hu, M., & Ji, Q. (2020). Financial markets under the global pandemic of COVID-19. *Finance Research Letters*, 36. <https://doi.org/10.1016/j.frl.2020.101528>

Figures

Before “National Emergency” Declared
 UK: January 1st, 2017 – March 22nd, 2020
 US: January 1st, 2017 – March 12th, 2020
 After “National Emergency” Declared
 UK: March 23rd, 2020 – March 21st, 2021
 US: March 13th, 2020 – March 21st, 2021

Figure 1

High-level features of web queries changed during the pandemic. The Google search volume index for (a&e) “How to” and “How do” questions and (b&f) “What” and “Why” questions were greater after the COVID-19 “National Emergency” declaration relative to before in (a&b) the UK and (e&f) the US. The Valence Index [0 (most negative valenced) to 100 (most positive valenced)] reveals that there was no difference in the valence of questions submitted to Google search engine in the UK and (g) the US after the COVID-19 “National Emergency” declaration relative to before. The period assessed prior to the “National Emergency” was from January 1st, 2017 to the declaration of each country’s “National Emergency”. The “National Emergency” was assessed from March 23rd, 2020 to March 21st, 2021, in the UK and from March 13th, 2020 to March 21st, 2021 in the US. (d&h) Plotted are the scores for each week from January 1st, 2017 to March 21st, 2021 for Google’s search volume index of “How to” and “How do” questions (Y-axis), Google’s search volume index of “What” and “Why” questions (X-axis) and the Valence index (Z-axis) for submitted questions in the (d) UK and (h) US. Grey circles represent the weeks prior to the “National Emergency”. Black diamonds represent the weeks after the “National Emergency”. Error bars = standard error (SE). *** = P < 0.001 (two-sided).

Figure 2

Temporal evolution of high-level features of web queries before and during the pandemic. (a&b) In the (a) UK and (b) US there was a sharp increase in the Google search volume index of “How to” and “How do” questions, and (c&d) “What” and “Why” questions after a “National Emergency” was declared. (e&f) There was no change in the Valence index [0 (most negatively valenced) and 100 (most positively valenced)]. The X-axis indicates the weeks ranging from January 1st, 2019 to March 21st, 2021. The solid black line indicates the time “National Emergency” was declared in the (a, c, &e) UK (March 23rd, 2020) and (b, d, &f) US (March 13th, 2020).

Figure 3

Stress is selectively associated with greater percentage of “How to” and “How do” questions submitted to Google search engine. Presented are the results of three separate linear regressions predicting the UK Google search volume index of (a) “How to” and “How do” questions, (b) “What” and “Why” questions, and (c) Valence Index [0 (most negative valanced) and 100 (most positive valanced)], from (i) the proportion of UK sample reporting COVID-related stress (detrended and Z scored) and (ii) UK COVID-19 related confinement score (detrended and Z scored). The X and Y values are the residuals (regressing out the respective control variable). The fine line represents the confidence interval. As can be observed, (a) increased stress was associated with increased “How to” and “How do” questions controlling for COVID-19 related confinement, while (b) increased COVID-19 related confinement was associated both with increased proportion of “What” and “Why” questions, and “How to” and “How do” questions. *** $P < 0.001$, ** $P < 0.01$ (two-sided).

Figure 4

"How to" and "How do" Google searches best predict population stress levels in the UK. Plotted on the y-axis are the beta coefficients from a linear regression predicting the proportion of UK sample reporting COVID related stress. On the x-axis are the predictor variables submitted into the linear model. We observe that the UK's Google search volume index for "How to" and "How do" questions predicted the proportion of UK sample reporting COVID related stress best. Error bars = standard error (SE). *** = $P < 0.001$ (two-sided).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SUPPLEMENTARY MATERIAL.docx](#)