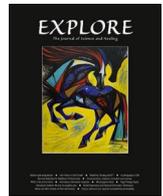


Contents lists available at [ScienceDirect](#)

EXPLORE

journal homepage: www.elsevier.com/locate/jsch

Validating the GCP data hypothesis using internet search data

Ulf Holmberg^{*,+}

Kammakargatan 50, 11160, Stockholm, Sweden

ARTICLE INFO

Keywords:

Consciousness
Global consciousness project
Internet searches
Google trends

ABSTRACT

The Global Consciousness Project (GCP) operates under the hypothesis that events that elicit widespread emotion or draw the simultaneous attention of a large number of people could affect the output of hardware-generated random numbers. The hypothesis thus suggests that the mind, in some sense, can interact with matter at a distance, a controversial suggestion because such a mechanism could challenge some current understandings. Testing the validity of the hypothesis thus carries substantial merit as negative results would reinforce already established scientific perceptions, whereas positive results would point in the direction of a needed update. In this paper, it is hypothesized that events inflicting a strong emotional response should also trigger the need for information. As such, global internet search trends should correlate with the GCP data, allowing for the hypothesis to be objectively tested. In practice, Google Trends search data is used to construct several search indexes that are correlated with GCP data aggregates using time series statistics. It is found that the GCP data significantly correlates with the indexes and can be used to improve the statistical model's in-sample fit. Furthermore, it is found that out-of-sample forecasts can be made more accurate if the GCP data is used. The results thus point toward the validity of the GCP data hypothesis and that the data produced by the GCP can be put to practical use by, for example, forecasters.

Brief introduction

The Global Consciousness Project (GCP) is an international and multidisciplinary collaboration project that generates and collects random number data continuously from a network of physical random number generators (RNGs) at several different locations around the world.¹ The random numbers are generated using quantum tunneling techniques, and the hypothesis underlying the GCP is that events that elicit widespread emotion or draw the simultaneous attention of large numbers of people could affect the output of hardware-generated random numbers in a statistically significant way. This is a controversial hypothesis as it suggests that the mind, in some sense, can affect matter at a distance.

The idea that the mind can affect matter at a distance has a long history in science. However, because the possibility that the mind can affect the random numbers produced by the GCP seems to challenge the current understanding of physics, most scientists demand a high standard of evidence. Although some research findings have produced results that support the existence of the phenomenon, supportive results

are in general regarded as speculative at best. As such, most scientists tend to reject both the possibility and existence of such a phenomenon, sometimes even without closely examining the data supporting the conclusion. The rejection of a hypothesis without examining its supportive data is, however, not best practice in scientific inquiry. Instead, one should seek to be open-minded and draw conclusions based on the evidence at hand. In this spirit, the hypothesis underlying the GCP is tested by taking the project's data as given and exploring the implications of doing so.

The hypothesis can be tested using the fact that other variables produced out of seemingly unrelated data sources should react to events that are claimed to affect the GCP data. One such variable identified is internet search trends because humans tend to search for information when engaging events occur. As such, global search behavior is likely to change during and after the time the engaging event takes place, an implied regularity that can be used to test the validity of the hypothesis underlying the GCP data.

Whether the monthly Google trends data correlates with a monthly aggregate constructed out of the second-by-second composite random

* Independent researcher.

E-mail address: ulf.e.holmberg@me.com.

⁺ The author appreciates the comments received from Roger D. Nelson and George Williams on an early draft of the paper.

¹ The exact number of generators varies from under 20 to over 70.

<https://doi.org/10.1016/j.explore.2022.07.007>

Received 25 February 2022; Received in revised form 6 July 2022; Accepted 23 July 2022

Available online 6 August 2022

1550-8307/© 2022 Elsevier Inc. All rights reserved.

numbers produced by the GCP is tested. In practice, several search indexes are constructed, univariate statistical time series models are fitted, and the GCP data aggregates are added. By examining if the GCP data significantly correlates with the indexes and improves the model's statistical fit, the validity of the hypothesis can be tested.

This paper is organized as follows. The next section briefly discusses consciousness, followed by a section discussing Google Trends data and the search indexes used. Next, the GCP and its data are presented and discussed, followed by an empirical section in which the relationship between the various search indexes and the GCP data aggregates are studied. The final section concludes, discusses the results, and suggests future research.

A discussion on consciousness

Consciousness is one of humanity's greatest mysteries as no one knows what it is, what it does, or even how it emerged. The prevailing working hypothesis, in most sciences, is that consciousness is an epiphenomenon of the brain and a result of physical arrangements and information processing patterns (see, e.g.¹). This viewpoint rests on the existence of neural correlates (see, e.g.^{6,19,17}, among others), but how the brain alone can produce subjective experiences (such as the feeling of warmth, cold, or pain) is not yet understood. It is even a philosophical mystery how unconscious matter can give rise to sentient beings, an unsolved philosophical conundrum often referred to as the "hard problem of consciousness."^{7,8}

The quest to understand the nature and origins of consciousness is also difficult due to the fact that it is subjective in its very nature. Accessing the subjective character of experience for other conscious organisms is, as far as people know, impossible because humans' only knowledge of subjective experience comes from accessing it directly from the inside.³⁷ As such, the current understanding of consciousness, and the current methods for studying it, suggests that any attempt to study the phenomenal nature of this world is limited by the lack of objective data. The complexity of the question has thus given rise to a palette of conflicting and often contradicting hypotheses regarding the nature of consciousness.

Some claim that consciousness in essence is illusionary (see, e.g.¹⁰) whereas others explore alternative frameworks and claim it to be fundamental.² Some also hypothesize that consciousness is a quantum phenomenon (see, e.g.^{27,13,14}) and often these hypotheses present mutually exclusive explanations. Most hypothesized explanations on the nature of consciousness, however, tend to ignore the psi results produced within the field of parapsychology, perhaps because some of the results suggest the existence of some form of controversial nonlocal mind-matter interaction process (see, e.g.³¹). Notably, however, some interpretations with regard to results produced within the field of quantum mechanics also suggest the existence of some form of nonlocal mind-matter interaction process. For example, the von Neumann-Wigner interpretation of the observer effect suggests that consciousness itself collapses the wave function (see, e.g.^{36,18,38}). As the mechanism underlying the results in quantum mechanics remains a mystery, it could be reasonable to take some of the results produced with the field of parapsychology more seriously. A study by Cardena⁵ also points in that direction as it was found that many parapsychological studies have produced highly significant results. As such, it could be beneficial to look at alternative theories of the mind for an explanation, one that also allows for the obtained psi research results.

Several studies conducted at the Princeton Engineering Anomalies Research (PEAR) lab produced results suggestive that consciousness could have the ability to interact with physical RNGs at a distance (see,

² Panpsychists (e.g.³² and ³⁴, for instance, suggest that consciousness is fundamental and intrinsic to the natural order of the world, an idea shared with those supporting the idea of neutral monism or other related philosophies.

e.g.^{22,29}). Strengthened by such results, an international and multidisciplinary collaboration project, the GCP, was created. The GCP aimed to study the hypothesis that the output of true RNGs could be affected by events that elicit widespread emotion or draw the simultaneous attention of a large number of people, or, as more precisely stated in Bancel³⁹,

Periods of collective attention or emotion in widely distributed populations will correlate with deviations from expectation in a global network of physical RNGs.

Studies conducted under the GCP have produced highly significant results that seem to validate the project's hypothesis (see, e.g.^{30,21,23}), but even so, the results tend to be put to question.³

May and Spottiswoode²⁰ examined the results produced by the GCP and suggested that the source of the statistical deviations reported could be attributed to a "psi-mediated experimenter effect," whereas others have suggested that the GCP results are due to the experimenter selecting events supportive of the project's hypothesis.⁴ Bancel², however, analyzed the data thoroughly and rejected the simple selection hypothesis with a reasonably high level of confidence. The door was, however, left open on the possibility that a psi-mediated experimenter effect was the cause of the observed GCP data effect.

The hypothesis needs to be studied using methods that allow for a post-hoc analysis while also accounting for the possibility that the data has been exposed to a psi-mediated experimenter effect. As such, an analysis of the already selected events will not do, but thankfully, the GCP has collected data continuously for many years such that the potential data effect due to many other historical events should be contained within the data.

By simply acknowledging that the events claimed to be picked up by the data produced by the GCP should also affect other seemingly unrelated data sources, the GCP data hypothesis can be studied objectively. This is because the GCP data hypothesis implicitly suggests that some unknown statistical quantity affects the data-generation process of both the random numbers produced by the GCP and other seemingly unrelated data sources. As such, several testable "intersections" should exist, as illustrated in Fig. 1.

Because it is well known that daily engaging events will affect market sentiment,³⁵ one such seemingly unrelated data source is daily stock market returns. Notably, empirical results supporting an intersection of the type illustrated in Fig. 1 with regard to stock market returns already exist.^{15,16} Another seemingly unrelated variable that could be affected by the events claimed to affect the GCP data is global internet search trends.

Global internet search trends and global attention

Searches made on the internet can be said to derive from the human need for information. Internet searches also require the searcher be attentive to the topic searched for, and the number of searches made relates to how engaging the topic searched for is. Thus, a rise in the number of searches made globally on specific topics should relate to how engaging the topic is perceived to be and thus to global attention. This is approximately what is claimed to affect the random numbers produced by the GCP such that an intersection of the type illustrated in Fig. 1 can be tested for.

Based on this insight, publicly available data on searches made using the search engine Google is used. In particular, monthly data on popular

³ Some claim that such results could violate some laws of physics (see, e.g.²⁶, at least how the laws are understood to date.

⁴ May and Spottiswoode²⁰ claimed that the decision augmentation theory (DAT) can adequately model the GCP results. Bancel² analyzed their findings and rejected this possibility in a commentary.

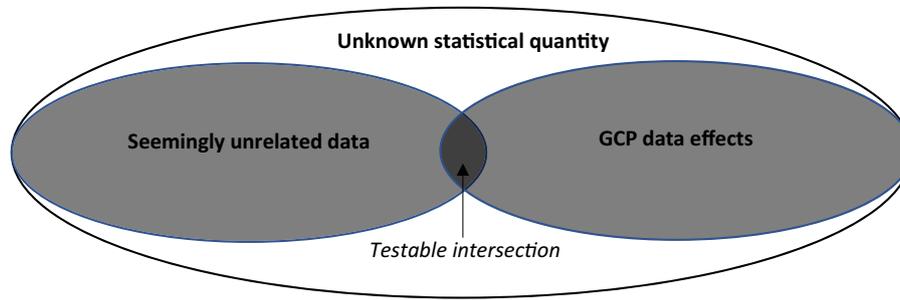


Fig. 1. The GCP data hypothesis implicitly suggests the existence of testable intersections.

news searches made annually since 2014 using Google’s “Year in Search” product is used, with data on each individual word stretching back to January 2004.^{5,6,7} Table 1 presents the 10 most searched for news-related words globally each year, and as expected, the exact words searched for vary between years, although some similar themes can be identified. In particular, searches related to shootings and hurricanes tend to recur, suggesting that natural disasters and random acts of violence could be particularly engaging events of interest.

The quantitative time series study begins with the definition of a simple index as the sum of all Google Trends index values of individual search words:

$$Simple_t = \sum_i word_{i,t} \quad (1)$$

where $word_{i,t}$ represents the Google Trends index value on the word (i) at time t .⁸ The *Simple* index thus treats all searches as equally engaging, but arguably, not all words in Table 1 are perceived as equally engaging by the global public.

Because the GCP data is hypothesized to be affected by global engagement and coherent attention, ideally only particularly emotionally engaging global topics would be considered in the index construction. Doing so, however, risks the index being subject to the modeler’s subjective judgment, an unwanted outcome in any objective study. Thankfully, this can be circumvented by simply acknowledging that the searches’ annual popularity order relates to perceived global engagement. As such, this study proceeds with the construction of a weighted summation index:

$$Weighted_t = \sum_i (W_{i,t} \times word_{i,t}) \quad (2)$$

where $W_{i,t}$ is equal to 1 if the word is the year’s most searched for word, $\frac{1}{2}$ if it is the second most searched for word, $\frac{1}{3}$ if it is the third most searched for word, and so forth.

The *Weighted* index, however, does not discriminate between searches’ perceived importance over time. As such, searches made on, for example, the city of Nice or Brussels or the country of North Korea

⁵ Google currently has between 80 and 90 percent of the global search engine market, so Google search trends should be able to “pick up” how engaging the global public perceives topics to be. It is, however, noted that inhabitants in China, North Korea, South Korea, and Russia do not use Google as their primary search engine, which could skew the results.

⁶ <https://about.google/stories/year-in-search/>

⁷ Google started reporting lists on global annual news searches in 2014, whereas the Google Trends data on the words searched for stretches back to 2004.

⁸ The Google Trends index value represents the sum of all Google searches’ relative importance over time, where the searches are related to the highest number of searches made for a given region (globally) and time (monthly between January 2004 and December 2021). A value of 100 thus indicates the highest interest in the search query since 2004, 50 indicates that it is half as popular, and 0 means that there is not enough data for the search term.

are given the same weight over the years, regardless of whether an event of importance related to the word occurred. These issues suggest that both the *Simple* index and the *Weighted* index will be subject to an underlying trend growth as more words with noticeable Google Trends index values are added to the index over time. To remedy these issues, a *Focused* version of the *Weighted* index is constructed in which only the most searched for words each year are included:

$$Focused_t = \sum_i (I_{i,t} \times W_{i,t} \times word_{i,t}), \quad (3)$$

where $I_{i,t}$ is a binary indicator variable equal to one if the word is one of the most searched for words during the year at time t and zero otherwise. As such, the *Focused* index can only be calculated for years with global news search words listed in Table 1 (i.e., for the years between 2014 and 2021).

Fig. 2 depicts the values of the three internet search indexes derived from all 10 search words. As expected, the *Simple* and *Weighted* indexes are heteroskedastic (become more volatile with time) and exhibit a small trend growth. The index values also seem to covary with each other, even though the magnitude of the intra-year variations obviously differs between the indexes. The *Simple* index exhibits the largest monthly variations and the *Focused* index the smallest, and all three indexes exhibit large intra-year volatility that can be used to study the validity of the GCP data hypothesis. For this, however, monthly aggregates constructed out of the GCP data that can be correlated with the data in Fig. 2 are needed.

The GCP data

The GCP generates and collects random numbers continuously from a network of physical RNGs. The physical random numbers are generated using quantum tunneling techniques and are hypothesized to be affected by events that elicit widespread emotion or draw the simultaneous attention of a large number of people.⁹ If so, the GCP data should covary with internet searches and with the indexes defined in the previous section. Monthly GCP data aggregates that can be correlated with the indexes are thus derived.

Denote the data produced by an individual RNG as $RNG_{i,\tau}$ for $i = 1, 2, \dots, n_\tau$, where n_τ is the total number of operating RNGs during second $\tau \in t$. The RNGs produce a series of 200 bits per second with an expected value of $\mu = 100$ and a variance of $\sigma^2 = 50$ from which n_τ standardized random numbers ($z_{i,\tau}$) can be calculated. To obtain measurable intraday data effects that can be aggregated into longer time frames, aggregate data on a standardized time frame from which the data aggregations is needed. To this end, the data is bundled into 15-minute (900 seconds) nonnegative data chunks using Stouffer’s Z-score method ($Z_{\tau,p}$).¹⁰ How this is

⁹ More information about how the data is generated can be found here: <https://noosphere.princeton.edu/gcpdata.html>

¹⁰ The intraday data found on the daily tables page on the GCP website is also bundled into 15-minute data chunks.

Table 1
The 10 news words most searched for since 2014.

Order	2014*	2015	2016	2017	2018	2019	2020	2021
1	Ebola	Charlie Hebdo	US Election	Hurricane Irma	World Cup	Copa America	Coronavirus	Afghanistan
2	ISIS	Paris	Olympics	Bitcoin	Hurricane Florence	Notre Dame	Election results	AMC Stock
3	Malaysia Airlines	Hurricane Patricia	Brexit	Las Vegas Shooting	Mega Millions Result	ICC Cricket World Cup	Iran	COVID Vaccine
4	Crimea Ukraine	Isis	Orlando Shooting	North Korea	Royal Wedding	Hurricane Dorian	Beirut	Dogecoin
5	Ferguson	Nepal	Zika Virus	Solar Eclipse	Election Results	Rugby World Cup	Hantavirus	GME Stock
6	Gaza and Israel	El Chapo	Panama Papers	Hurricane Harvey	Hurricane Michael	Sri Lanka	Stimulus checks	Stimulus Check
7	Scottish Referendum	Greece	Nice	Manchester	Kavanaugh Confirmation	Area 51	Unemployment	Georgia Senate Race
8	Oscar Pistorius trial	Baltimore Riots	Brussels	Hurricane Jose	Florida Shooting	India election results	Tesla stock	Hurricane Ida
9	-	San Bernardino	Dallas Shooting	Hurricane Maria	Greve dos caminhoneiros	台風 19 号	Bihar election result	COVID
10	-	Hurricane Joaquin	熊本 地震 Kumamoto Earthquake	April the Giraffe	Government Shutdown	Fall of Berlin Wall	Black Lives Matter	Ethereum Price

Source: Google Trends.

*Only eight news search words are reported for the year 2014.

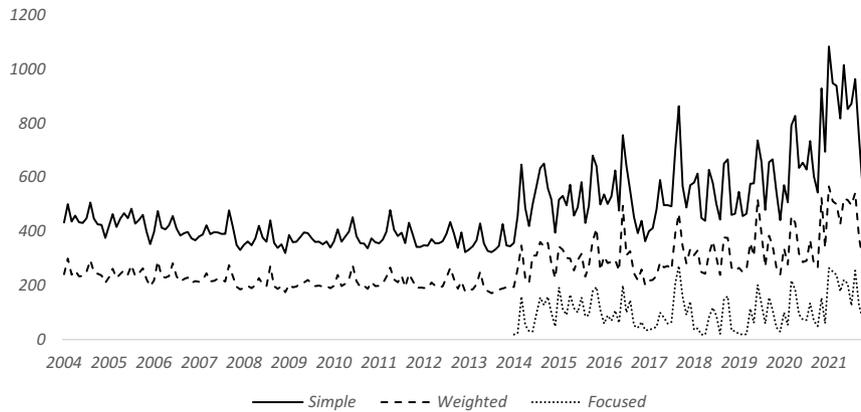


Fig. 2. Index values constructed out of 10 search words. Note: When the data is extended back before 2014, the events triggering most searches had not yet occurred, so volatility is reduced. Source: Google Trends and own calculations.

done can easiest be seen by applying the following formulas to the extracted data:¹¹

$$Z_\tau = \left| \sum_{\tau-900}^{\tau} z_\tau / \sqrt{900} \right|, \quad (4)$$

where

$$z_\tau = \left(\sum_i^{N_\tau} RNG_{i,\tau} - N_\tau \mu \right) / \sqrt{N_\tau \sigma^2} \quad (5)$$

and where N_τ is the number of active RNGs during τ . Measuring Z_τ at the end of each 15-minute interval, as is done in the daily tables section on the GCP website, 96 daily intraday measurements are obtained, and from these measurements, a daily aggregate is calculated.

The daily aggregates aim to “pick up” large and unexpected values

¹¹ Note that the formulas are the exact versions of the approximations used in Holmberg¹⁶.

that are hypothesized to occur together with engaging word events. As Holmberg¹⁵ argued, the daily maximum value of Z_τ should conceptually capture such changes, but the daily average should also be affected.¹² As such, let the daily intraday maximum value be denoted as $Max[Z_t] = max(Z_\tau)$, for $\forall \tau \in t$ and define the intraday average as $Average[Z_t] = \sum_\tau Z_\tau / 96$.

Most observations underlying the daily GCP data aggregates utilize the intraday data found in the “All Egg Composite” column on the daily tables page of the GCP website.¹³ However, because the reporting RNGs at times are affected by technical malfunctions resulting in unusually large Z_τ values, the intraday data used on all dates on which $Max[Z_t]$ is found to be larger than 5 is recalculated.¹⁴ Furthermore, data on dates with no reported daily tables values is calculated if the data can be

¹² Arguably, other intraday time frames could have also been chosen. However, large measurable intraday movements caused by engaging global events should “show up” in the aggregation procedure regardless of the exact time frame chosen.

¹³ The daily tables data underlying these aggregates can be found here: <https://global-mind.org/data/eggsummary>

¹⁴ The data can be extracted here: <https://noosphere.princeton.edu/extract.html>

accessed using the GCP's data extraction tool. Note also that all temporarily malfunctioning RNGs are removed when calculations are remade.¹⁵

Fig. 3 depicts the daily $Average[Z]$ and daily $Max[Z]$ values and their 30-day monthly average counterparts, and Table 2 presents some descriptive data. As can be seen, both the daily and monthly aggregates fluctuate around their empirical mean, suggesting that the data is stationary. This is also confirmed using the Dickey–Fuller statistic.¹¹

Empirical results

This section investigates the hypothesis underlying the GCP data empirically. To this end, let $SI_{i,t}$ denote the search word index i at time t where $i = \{Simple, Weighted, Focused\}$, and let GCP_j be the monthly average GCP data aggregate studied with $j = \{Max[Z], Average[Z]\}$.

As discussed in the previous sections, if the GCP data reacts to engaging events of perceived global importance, the daily aggregates should react. This should also be picked up by the monthly GCP data aggregates because it would cause the aggregates to be somewhat larger than ordinary during months in which search activity is high. Thus, using a simple linear regression model, whether the search indexes correlate directly with the GCP data is investigated:

$$SI_{i,t} = \alpha_{i,j} + \beta_{i,j} \times GCP_{j,t} + \varepsilon_{i,t}, \quad (6)$$

where $\varepsilon_{i,t}$ is assumed to be a white noise time series. Initially, the focus is on the two longer time series indexes (*Simple* and *Focused*), and because the search word indexes are likely to be heteroskedastic and serially correlated (Fig. 2), heteroskedasticity and autocorrelation consistent (HAC) standard errors are used when assessing the parameters' significance.²⁴ Table 3 presents estimates on Eq. (6), obtained using ordinary least squares.

The results in Table 3 indicate that the GCP data indeed covaries with global internet searches. In fact, β is positive and significant using both the monthly $Max[Z]$ and $Average[Z]$ in Eq. (6), suggesting that search activity is indeed elevated during months in which GCP data aggregates are large. A closer examination of the models' residuals using the augmented Dickey–Fuller test statistic⁹, however, suggests that the models suffer from a unit root and the results presented in Table 1 are only indicative.

The unit root problem is addressed by modeling the stationary first difference. However, because the index values themselves partly include the sought-after GCP data effect and because a coherent emotional response of an event is likely to occur instantaneously, whereas global internet search trends could react more slowly, one period's lagged changes in the aggregate GCP data is studied. For tractability, a linear dependence is assumed:

$$\Delta SI_{i,t} = \alpha_i + \beta_i \times \Delta GCP_{t-1} + u_{i,t}, \quad (7)$$

where the error term $u_{i,t}$ contains the models' autoregressive components:

$$u_{i,t} = \sum_j^J \rho_{i,j} u_{i,t} + \varepsilon_{i,t}. \quad (8)$$

In Eq. (8), J is the number of autoregressive parameters, and $\varepsilon_{i,t}$ is a pure white noise process subject to the usual assumptions.¹⁶ It is thus implicitly assumed that an autoregressive process is an adequate description of the rise and fall in popularity of internet searches. The autocorrelation order is determined using the Box–Jenkins method³.

From Eq. (7), it can be understood that a test for the hypothesis underlying the GCP boils down to a test for whether $\beta_i = 0$ such that the

GCP data hypothesis can be tested using a standard t-statistic. Furthermore, the null hypothesis ($\beta_i = 0$) suggests that univariate models better fit the data such that the hypothesis can also be investigated by simply comparing the GCP data-dependent models' statistical fit with their univariate counterparts.

Table 4 presents estimates on the models in Eqs. (7) and (8), and in general, evidence in favor of the GCP data hypothesis is found ($\beta \neq 0$). The estimates also reveal that the GCP data-dependent models have a better statistical fit than their univariate counterparts because the GCP data-dependent models reduce the models' Akaike information criterion (AIC) in 15 out of the 18 the models studied¹. Note also that a rise in the GCP data aggregates results in an increase in internet searches, confirming the simple linear regression findings in Table 3.

It is also found that $Max[Z]$ captures the GCP data effect to a greater degree than $Average[Z]$ and that the GCP data covaries more strongly with the indexes constructed out of only the most searched for words each year (the GCP data reduces the AIC more when the 3- and 5-word indexes are studied). The coefficient of determination (R^2) also tends to increase if the words are weighted in accordance with their popularity (*Weighted* and *Focused*), a result that fits well with the hypothesis underlying the GCP data. The best fitted models in the table are also the ones fitted on the *Focused* indexes, even though those models are estimated using a lesser amount of data.

Examining the 5- and 3-word indexes in more detail

The results in Table 4 point toward the validity of the hypothesis underlying the GCP data. Because the estimates also suggest that the GCP data covaries more strongly with the indexes constructed out of 5 and 3 search words, the robustness of the results is investigated by examining the 5- and 3-word indexes in more detail.

Recalling that the words listed in Table 1 seem to suggest that searches on natural disasters and random acts of violence are somewhat more engaging (searches related to hurricanes or shootings recur over the years), whether the results are affected by the inclusion of the words "Hurricane" and "Shooting" is investigated. Additionally, searches made on the word "Earthquake" are included, arguably a natural disaster that can be highly engaging. These searches are added to the 5- and 3-word indexes, and the modified indexes are denoted as boosted indexes. Table 5 presents the results.

Table 5 indicates that the significance of the GCP data aggregates increases, and the models' statistical fit is improved (*Impact on AIC*). As such, it seems like the search words included in the boosted indexes indeed covary strongly with the GCP data aggregates, an interesting finding that suggests avenues for further exploration. The results also strengthen the results in Table 4, as again, $Max[Z]$ is the GCP data aggregate that captures the GCP data effect to a greater degree. Weighting the search words with regard to their annual popularity also seems to increase the models' fit as the *Focused* index model reduces the models' AIC the most.

The robustness of the results is investigated by splitting the sample in two and performing a subsample analysis on the split sample. Table 6 presents the results, and as can be seen, $Average[Z]$ loses its significance, whereas the $Max[Z]$ aggregate remains highly significant ($P < 0.01$). Also, by conditioning the index changes on the GCP data aggregates, the model fit improves (*Impact on AIC*). Furthermore, using the GCP data-dependent parameter estimates from the first period to forecast the second period's search trends, the forecasted root mean square error (RMSE) is reduced by 7 to 8 percent.

Finally, the validity and robustness of the results is further investigated by splitting the data into several three-year time periods from which one-year out-of-sample forecasts are made. In doing so, the same model-dependence structure is used as in Table 6 but estimated on a rolling three-year sample. Fig. 4 depicts the validation procedure.

Table 7 presents the out-of-sample validation results and shows how much the one-year out-of-sample forecasts' RMSE is reduced if the

¹⁵ Only about 2 percent of all 6510 daily observations are recalculated.

¹⁶ The residuals are independent, normally distributed, and homoscedastic.

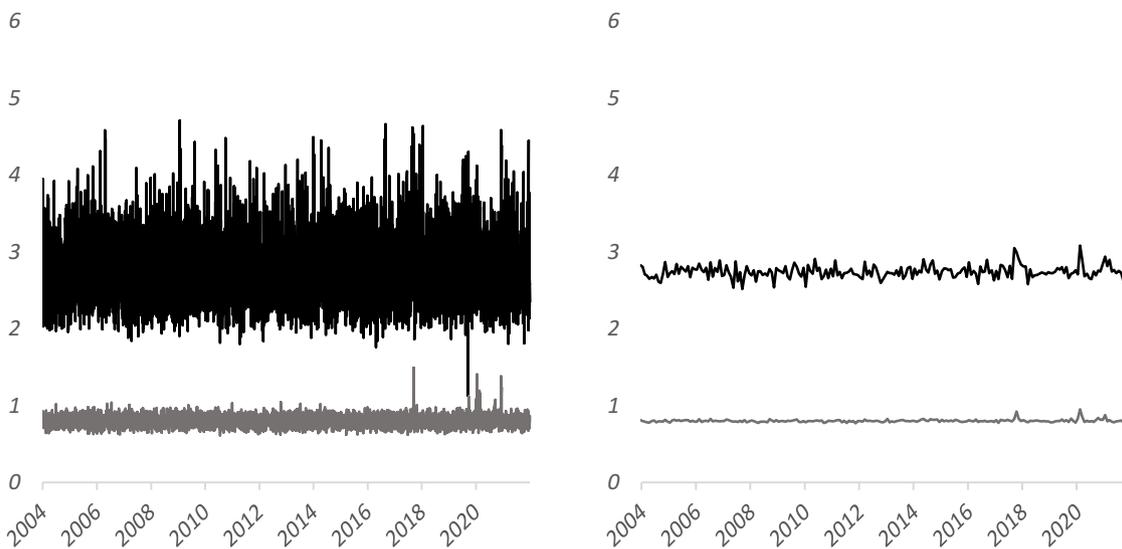


Fig. 3. The GCP data aggregates. Note: From 6510 daily observations between 2004-01-01 and 2021-12-28 from which monthly averages have been calculated.

Table 2
Descriptive data.

	Daily Average[Z]	Max[Z]	Monthly Average[Z]	Max[Z]
Minimum	0.61	1.13	0.77	2.51
Average	0.80	2.74	0.80	2.74
Median	0.80	2.70	0.80	2.74
Maximum	1.49	4.71	0.95	3.07
Standard deviation	0.07	0.41	0.02	0.08
Skewness	1.13	0.78	3.76	0.62
Kurtosis	6.86	1.19	23.75	2.22

Note: From 6510 daily observations between 2004-01-01 and 2021-12-28 from which monthly averages have been calculated.

Table 3

Simple linear regression estimates on the *Simple* and *Weighted* indexes, Results supporting the GCP data hypothesis are boldface.

	<i>Simple</i> 10 words		5 words		3 words	
	Max[Z]	Average[Z]	Max[Z]	Average[Z]	Max[Z]	Average[Z]
α	-359.72	-671.13	-257.65	-309.59	-139.68	-147.18
$\beta_{Max[Z_t]}$	306.13*	-	171.88**	-	101.90**	-
$\beta_{Average[Z_t]}$	-	1436.47**	-	653.15**	-	358.16***
R ²	0.03	0.04	0.03	0.02	0.02	0.02
	<i>Weighted</i> 10 words		5 words		3 words	
	Max	Average	Max	Average	Max	Average
α	-205.32	-347.64	-650.48	-1075.14**	-53.99	-64.02
$\beta_{Max[Z_t]}$	171.67**	-	299.58**	-	51.84	-
$\beta_{Average[Z_t]}$	-	765.22**	-	1547.88**	-	189.98
R ²	0.03	0.03	0.04	0.06	0.01	0.01

Significance levels: * 10%, ** 5%, and *** 1%.

Note: Estimated using ordinary least squares with HAC standard errors using EViews 12. The Simple and Weighted models utilize the full sample of data available from January 2004 to December 2021.

forecasts are conditioned on $Max[Z_{t-1}]$. As can be seen, the GCP data adds useful information to the forecasts because models dependent on the GCP data reduce the out-of-sample forecasts' RMSE by as much as 8.26 percent. However, during 2020, the GCP data-dependent forecasts' RMSE *increased* compared with their univariate counterpart. The cause of the anomalous 2020 results is thus analyzed in more detail.

Fig. 5 depicts the 5-word *Focused* index's subcomponents during 2020 and shows that the monthly average $Max[Z_t]$ increased sharply in January at the same time as internet searches on the word "Iran" increased. The GCP data aggregate then dropped in February and

reached a local low in March when searches on the word "Coronavirus" skyrocketed.¹⁷ At first, this finding does not align well with the other results presented in this paper. However, the delayed *Focused* index response observed could possibly be motivated by the fact that the virus originated from China and that China's search engine market is dominated by the search engine Baidu. Because the search trends studied are measured on searches made on Google, the initial spread of the virus could have affected the GCP data without affecting the search indexes. Noting also that searches on the word "Coronavirus" are given a weight of 1, it can be suspected that the onset of the pandemic was both the most engaging world event and the cause of the anomalous large January rise in $Max[Z]$, an interesting avenue for future research to explore.

Concluding remarks

This study has examined whether global internet searches made using the search engine Google covary with an aggregate constructed out of the data produced by the GCP. The results suggests that they do, which partially confirms the GCP findings. Additionally, because the

¹⁷ Before the COVID-19 virus causing the pandemic was given an official name, it was commonly referred to as the "coronavirus."

Table 4
Estimates on $\Delta SI_{i,t}$ Results supporting the GCP data hypothesis are boldface.

	<u>10- word index</u>		Weighted	Average[Z]	Focused	Average[Z]
	Simple	Average[Z]				
	Max[Z]		Max[Z]		Max[Z]	
α	0.28	0.28	-0.01	-0.01	0.32	0.32
$\beta_{Max[Z_{t-1}]}$	103.24*	-	83.57**	-	120.87*	-
$\beta_{Average[Z_{t-1}]}$	-	209.48	-	276.15	-	375.54
ρ_1	-0.38***	-0.37***	-0.45***	-0.45***	-0.48***	-0.46***
ρ_2	-0.27***	-0.27***	-0.28***	-0.29***	-0.32***	-0.33***
R^2	0.16	0.15	0.20	0.19	0.20	0.20
Impact on AIC	-0.01%	0.09%	-0.07%	0.03%	-0.07%	0.00%
Durbin-Watson	2.08	2.07	2.06	2.05	2.05	2.04
	<u>5-word index</u>		Weighted	Average[Z]	Focused	Average[Z]
	Simple	Average[Z]				
	Max[Z]		Max[Z]		Max[Z]	
α	-0.04	-0.04	0.20	0.20	0.16	0.17
$\beta_{Max[Z_{t-1}]}$	111.33**	-	207.91***	-	111.56**	-
$\beta_{Average[Z_{t-1}]}$	-	396.22*	-	925.90***	-	328.00*
ρ_1	-0.57***	-0.56***	-0.58***	-0.59***	-0.62***	-0.60***
ρ_2	-0.32***	-0.34***	-0.26***	-0.30***	-0.34***	-0.34***
R^2	0.27	0.26	0.26	0.26	0.29	0.28
Impact on AIC	-0.14%	-0.03%	-0.16%	-0.14%	-0.29%	-0.14%
Durbin-Watson	2.06	2.04	2.04	2.04	2.00	1.99
	<u>3-word index</u>		Weighted	Average[Z]	Focused	Average[Z]
	Simple	Average[Z]				
	Max[Z]		Max[Z]		Max[Z]	
α	-0.09	-0.09	-0.01	-0.01	0.21	0.22
$\beta_{Max[Z_{t-1}]}$	88.22**	-	52.14**	-	85.02**	-
$\beta_{Average[Z_{t-1}]}$	-	403.18*	-	247.88*	-	288.66**
ρ_1	-0.52***	-0.53***	-0.55***	-0.56***	-0.63***	-0.63***
ρ_2	-0.28***	-0.30***	-0.29***	-0.31***	-0.36***	-0.37***
R^2	0.24	0.24	0.25	0.26	0.31	0.31
Impact on AIC	-0.16%	-0.16%	-0.04%	-0.06%	-0.26%	-0.23%
Durbin-Watson	2.06	2.05	2.09	2.08	1.99	2.00

Significance levels: * 10%, ** 5%, and *** 1%.

Note: Estimated using ARMA Maximum Likelihood (OPG – BHHH) in EViews 12. The Simple and Weighted models utilize the full sample of data available from January 2004 to December 2021, and the Focused index estimates data between January 2014 and December 2021.

Table 5
Estimates on $\Delta SI_{i,t}$ boosted indexes, Results supporting the GCP data hypothesis are boldface.

	<u>5-word index</u>		Weighted	Average[Z]	Focused	Average[Z]
	Simple	Average[Z]				
	Max[Z]		Max[Z]		Max[Z]	
α	-0.01	-0.01	0.07	0.07	0.19	0.20
$\beta_{Max[Z_{t-1}]}$	137.78***	-	224.86***	-	161.52***	-
$\beta_{Average[Z_{t-1}]}$	-	404.79*	-	874.16**	-	365.28*
ρ_1	-0.50***	-0.50***	-0.53***	-0.54***	-0.54***	-0.51***
ρ_2	-0.30***	-0.31***	-0.25***	-0.28***	-0.29**	-0.28**
R^2	0.23	0.22	0.23	0.23	0.25	0.22
Impact on AIC	-0.22%	-0.02%	-0.20%	-0.10%	-0.53%	-0.09%
Durbin-Watson	2.07	2.05	2.05	2.05	2.02	2.01
	<u>3-word index</u>		Weighted	Average[Z]	Focused	Average[Z]
	Simple	Average[Z]				
	Max[Z]		Max[Z]		Max[Z]	
α	-0.05	-0.05	0.03	0.03	0.22	0.24
$\beta_{Max[Z_{t-1}]}$	119.45***	-	85.13***	-	140.11***	-
$\beta_{Average[Z_{t-1}]}$	-	428.68***	-	279.11*	-	347.26**
ρ_1	-0.47***	-0.48***	-0.49***	-0.51***	-0.57***	-0.56***
ρ_2	-0.27***	-0.28***	-0.27***	-0.28***	-0.32***	-0.31***
R^2	0.22	0.21	0.22	0.22	0.29	0.25
Impact on AIC	-0.30%	-0.13%	-0.21%	-0.04%	-0.66%	-0.21%
Durbin-Watson	2.08	2.07	2.09	2.09	2.02	2.02

Significance levels: * 10%, ** 5%, and *** 1%.

Note: Estimated using ARMA Maximum Likelihood (OPG – BHHH) in EViews 12. The Simple and Weighted models utilize the full sample of data available from January 2004 to December 2021, and the Focused index estimates data between January 2014 and December 2021.

analysis makes use of historical search data on the most popular searches globally, the possibility that the correlations found are due to a psi-mediated experimenter effect seems unlikely. Instead, the results point toward the possibility that some unknown statistical quantity causes internet search trends to react in conjunction with variations in the GCP

data. One such quantity is focused attention, such that the results point toward the validity of the hypothesis underlying the GCP (i.e., that events that elicit widespread emotion or draw the simultaneous attention of large numbers of people affect the output of hardware-generated random numbers).

Table 6

Subsample estimates for selected disaster-added *Focused* indexes, Results supporting the GCP data hypothesis are boldface.

	<i>Boosted 5-word index</i>		<i>2017-2021</i>	
	<i>2014-2017</i>			
	<i>Max[Z]</i>	<i>Average[Z]</i>	<i>Max[Z]</i>	<i>Average[Z]</i>
α	1.10	1.21	0.35	0.61
$\beta_{Max[Z_{t-1}]}$	193.20***	-	207.13***	-
$\beta_{Average[Z_{t-1}]}$	-	646.21	-	349.76
ρ_1	-0.45**	-0.43**	-0.55***	-0.51***
ρ_2	-0.28	-0.30	-0.32*	-0.29*
R^2	0.26	0.21	0.27	0.20
<i>Impact on AIC</i>	-0.77%	-0.22%	-0.94%	-0.01%
<i>Durbin-Watson</i>	2.11	2.10	1.99	1.99
<i>RMSE reduction*</i>	-	-	-6.40%	0.27%
	<i>Boosted 3-word index</i>		<i>2017-2021</i>	
	<i>2014-2017</i>			
	<i>Max[Z]</i>	<i>Average[Z]</i>	<i>Max[Z]</i>	<i>Average[Z]</i>
α	0.80	0.90	0.22	0.47
$\beta_{Max[Z_{t-1}]}$	182.31***	-	185.49***	-
$\beta_{Average[Z_{t-1}]}$	-	633.69	-	316.49
ρ_1	-0.55***	-0.54***	-0.55***	-0.55***
ρ_2	-0.36**	-0.34*	-0.34**	-0.31**
R^2	0.32	0.28	0.32	0.23
<i>Impact on AIC</i>	-1.18%	-0.45%	-1.18%	-0.07%
<i>Durbin-Watson</i>	2.09	2.07	2.01	2.03
<i>RMSE reduction*</i>	-	-	-8.02%	-0.07%

Significance levels: * 10%, ** 5%, and *** 1%.

Note: Estimated using ARMA Maximum Likelihood (OPG – BHHH) in EViews 12.

The hypothesis is tested by calculating the monthly average of the daily aggregate GCP data variables *Max[Z]* and *Average[Z]*. These variables should be able to capture unexpectedly high intraday Z-score values during days on which engaging events occur, one of the many expected outcomes if the GCP data hypothesis holds true. Changes in the monthly average *Max[Z]* and *Average[Z]* are then correlated with indexes constructed out of Google Trends search data, and because changes in these indexes should, to some degree, reflect changes in how intensely the public feels affected by an event, the validity of the GCP data hypothesis can be tested.

A statistical time series analysis reveals that both the monthly *Average[Z]* and *Max[Z]* variables significantly correlate with all studied search word indexes, and the most significant correlation is found on *Max[Z]*. The finding that the GCP data correlates with global search trends is also further strengthened if searches on the words “Earthquake,” “Hurricane,” and “Shooting” are included in the index construction ($P < 0.01$). It is also found that the indexes on which the words are weighted with regard to their annual popularity are better fitted to

the GCP data. Furthermore, the results show that out-of-sample forecasts on internet search trends can be improved, sometimes by as much as 8 percent, when the forecasts are conditioned on the information contained within the GCP data aggregates. The results presented herein thus both provide empirical support in favor of the hypothesis underlying the GCP and point toward how the data can be put to practical use.

What mechanism could cause the GCP data to react together with changes in global internet search trends? The prevailing working hypothesis with regard to consciousness suggests that consciousness is an epiphenomenon of the brain and the result of physical arrangements and complex information processing patterns. Unless the assumed information processing mechanism gives rise to the observed GCP data effect, this hypothesis does not seem to be well equipped to explain the results presented herein.

Alternative hypotheses with regard to consciousness exist, and among them, the electromagnetic field theories of consciousness (see, e.g. ^{28,25}) or quantum consciousness ideas such as the orchestrated objective reduction hypothesis (see, e.g. ^{27,13,14}) can be mentioned. However, because the electromagnetic field of the human brain has been found to only stretch out about 63 cm from the human skull,⁴ and because the orchestrated objective reduction hypothesis postulates that consciousness originates locally at the quantum level inside neurons, they alone cannot explain the results presented in this study. Perhaps, instead, other less accepted ideas of the mind could also be considered.

Some of the more parsimonious alternative hypotheses postulate a field mechanism of the mind, possibly also with the ability to affect matter at a distance (see, e.g. ³³). Such alternative explanations more closely relate to contemporary philosophical ideas on the nature of consciousness, ideas that often require the existence of a unified consciousness field of sorts (see, e.g. ³⁷). Whatever the mechanism behind the results might be, the results suggest that the prevailing paradigm with regard to consciousness needs to be discussed, as the results cannot be understood using the current understanding of consciousness alone.

One way to advance the understanding of consciousness in light of these results could be to discuss how the results fit within the current

Table 7

RMSE reduction, one-year out-of-sample forecasts, Results supporting the GCP data hypothesis are boldface.

	2017	2018	2019	2020	2021
<i>5-word index</i>	-4.85%	-8.26%	-3.40%	46.14%	0.27%
<i>3-word index</i>	-7.37%	-5.48%	-3.33%	20.76%	-0.61%

Note: RMSE reduction obtained from models using ARMA Maximum Likelihood (OPG – BHHH) in EViews 12.

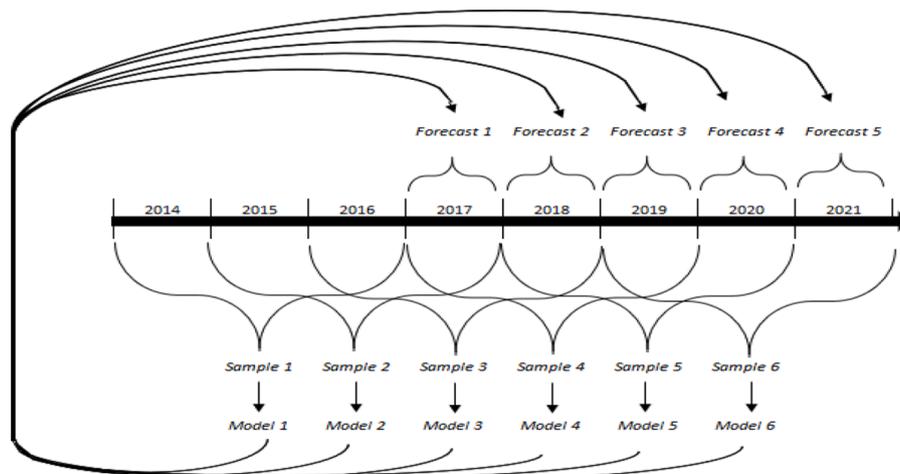


Fig. 4. Out-of-sample forecast validation procedure.

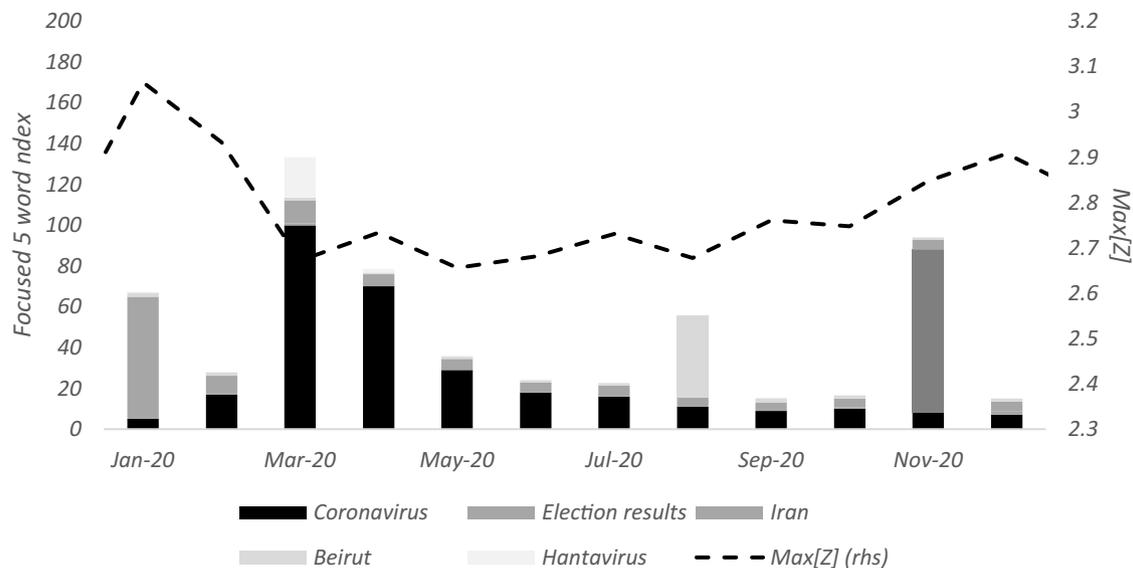


Fig. 5. The five most popular searches made during 2020 and the evolution of $Max[Z]$. Source: Google Trends and own calculations.

understanding of physics. It has been claimed that the results produced by the GCP cannot hold true because if the mind can stretch beyond the head and affect RNGs at a distance, the distance between the RNG and the mind would be a major factor describing the strength of such an effect. Such criticism stems from the fact that many physical quantities have been found to be inversely proportional to the square of the distance from the source of that physical quantity (the inverse-square law), but because some physical processes are unaffected by distance (e.g., quantum entanglement), the idea that such a law is at play here cannot be ruled out.

It could also be speculated that a consciousness-related quantum mechanical effect could be affecting the RNGs at a distance such that the results could be said to shed some light on how the empirically verified observer effect in quantum mechanics should be interpreted. This, as a mind-matter interaction of the type suggested by the results, seems to favor the von Neumann-Wigner interpretation (i.e., that consciousness itself collapses the wave function) (see, e.g. ^{36,18,38}). It should, however, also be noted that other mechanisms probably play a big part in explaining the results because it can be assumed that the individuals emotionally affected by the events in general are unaware of the physical RNGs' existence.

The results presented herein open up many avenues for future research to explore as they suggest that the current paradigm with regard to consciousness needs to be revised. Additionally, as global events of interest have been studied, the question of whether distance affects the observed GCP data effect remains open and an interesting avenue for future research to explore. Research could be directed toward understanding how the GCP data interacts with global internet searches more thoroughly and possibly also how it reacts together with local search trends. It also appears that some specific search words covary more strongly with the GCP data aggregates, an interesting finding that hints toward further exploration. Finally, it is probable that the GCP data can be linked to many other variables related to attention and engagement, which could be explored in future research.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.explore.2022.07.007.

References

- Akaike H. A new look at the statistical model identification. *IEEE Trans Autom Control*. 1974;19:716–723.
- Bancel P. Reply to May and Spottiswoode's "The Global Consciousness Project: Identifying the Source of Psi." *J Scient Explor*. 2011;25:690–694.
- Box G, Jenkins G. *Time Series Analysis: Forecasting and Control*. San Francisco, CA: Holden-Day; 1970.
- Brazdzionis J, Wiginton J, Patchana T, et al. Measuring the electromagnetic field of the human brain at a distance using a shielded electromagnetic field channel. *Cureus*. 2022;14:e23626. <https://doi.org/10.7759/cureus.23626>.
- Cardena E. The experimental evidence for parapsychological phenomena: a review. *Am Psychol*. 2018;73:663–677.
- Cotterill R. Cooperation of the basal ganglia, cerebellum, sensory cerebrum and hippocampus: Possible implications for cognition, consciousness, intelligence and creativity. *Prog Neurobiol*. 2001;64:1–33.
- Chalmers D. Facing up to the problem of consciousness. *J Consciousness Stud*. 1995;2: 200–219.
- Chalmers D. Consciousness and its place in nature. In: Stich S, Warfield F, eds. *Blackwell Guide to the Philosophy of Mind*. Malden, MA: Blackwell; 2003.
- Cheung Y-W, Lai KS. Lag order and critical values of the augmented Dickey-Fuller test. *J Bus Econ Stat*. 1995;13:277–280.
- Dennett, D. (1991). A. Lane (ed.). *Consciousness Explained*. The Penguin Press.
- Dickey DA, Fuller WA. Distribution of the estimators for autoregressive time series with a unit root. *J Am Statist Assoc*. 1979;74:427–431.
- Güzeldere G. *The Nature of Consciousness: Philosophical debates*. Cambridge, MA: MIT Press; 1997.
- Hameroff S. Quantum computation in brain microtubules? The Penrose-Hameroff "Orch OR" model of consciousness. *Philos Trans R Soc A*. 1998;356:1869–1896.
- Hameroff S. How quantum brain biology can rescue conscious free will. *Front Integr Neurosci*. 2012;6. <https://doi.org/10.3389/fnint.2012.00093>.
- Holmberg U. Stock returns and the mind: An unlikely result that could change our understanding of consciousness. *J Consciousness Stud*. 2020;27:31–49.
- Holmberg U. Revisiting stock returns and the mind: Digging deeper into the data. *J Consciousness Explor Res*. 2021;2:207–223.
- Koch C. *The Quest for Consciousness*. Englewood, CO: Roberts & Company; 2004.
- English translation in London F, Bauer E. La théorie de l'observation en mécanique quantique. In: Wheeler JA, Zurek WH, eds. *Quantum Theory and Measurement*. Princeton, New Jersey: Princeton University Press; 1939:217–259. Location.
- Llinás R. *I of the Vortex. From Neurons to Self*. Cambridge, MA: MIT Press; 2002.
- May EC, Spottiswoode JP. The global consciousness project: Identifying the source of psi. *J Scient Explor*. 2011;25:663–682.
- Nelson RD, Radin D, Shoup R, Bancel P. Correlations of continuous random data with major world events. *Found Phys Lett*. 2002;15:537–550.
- Nelson RD, Jahn R, Dunne B. Operator-related anomalies in physical systems and information processes. *J Soc Psychological Res*. 1986;53:261–285.
- Nelson RD, Bancel P. Effects of mass consciousness in random data during global events. *Explore*. 2011;7:373–383.
- Newey WK, West KD. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*. 1987;55:703–708.
- McFadden J. The conscious electromagnetic information (CEMI) field theory: The hard problem made easy? *J Consciousness Stud*. 2002;9:45–60.
- Park RL. *Voodoo Science: The Road from Foolishness to Fraud*. Reprint. Oxford: Oxford University Press; 2000.

- 27 Penrose R. *The Emperor's New Mind: Concerning Computers, Minds and The Laws of Physics*. Oxford: Oxford University Press; 1989.
- 28 Pocket S. The electromagnetic field theory of consciousness: a testable hypothesis about the characteristics of conscious as opposed to non-conscious fields. *J Consciousness Stud.* 2012;19:191–223.
- 29 Rowe WD. Physical measurements of episodes of focused group energy. *J Scient Explor.* 1998;12:569–583.
- 30 Radin D. Exploring relationships between random physical events and mass human attention: asking for whom the bell tolls. *J Scient Explor.* 2002;16:533–547.
- 31 Radin D, Nelson RD, Dobyns Y, Houtkooper J. Assessing the evidence for mind-matter interaction effects. *J Scient Explor.* 2006;20:361–374.
- 32 Seager W. Consciousness, information, and panpsychism. *J Consciousness Stud.* 1995; 2:272–288.
- 33 Sheldrake R. *A New Science of Life: The Hypothesis of Formative Causation*. third edition 2009. Los Angeles, CA: J.P. Tarcher; 1981.
- 34 Strawson G. Realistic monism - why physicalism entails panpsychism. *J Consciousness Stud.* 2006;13:3–31.
- 35 Shiller RJ. Narrative economics. *Am Econ Rev.* 2017;107:967–1004.
- 36 Von Neuman, J. (1932). *Mathematical Foundations of Quantum Mechanics*. R. T. Beyer (trans.). Princeton University Press (1996 edition).
- 37 Williams G. Can the psi data help us make progress on the problem of consciousness? *J Consciousness Stud.* 2021;28:145–172.
- 38 Wigner E, Margenau H. Remarks on the mind body question. Symmetries and reflections, scientific essays. *Am J Phys.* 1967;35:1169–1170.
- 39 Bancel Peter. An analysis of the global consciousness project. In: Broderick Damien, Goertzel Ben, eds. *Evidence for Psi: Thirteen Empirical Research Reports*. Jefferson, North Carolina: McFarland & Company; 2014:255–277 Chapter 11. In press.