Algorithms without frontiers? How language-based algorithmic information disparities for suicide crisis information sustain digital divides over time in 17 countries

Sebastian Scherr [a], Florian Arendt [b] and Mario Haim [c]

[a]Texas A&M University, College Station, TX, USA; [b]University of Vienna, Austria, Austria; [c]LMU Munich, Munich, Germany

ABSTRACT

This study focuses on the changes in the global digital divide produced by language-based, algorithmic information disparities in relation to crisis-prevention resources for suicide available through the Google search engine. We used agent-based testing to emulate Google searches performed in 17 countries and in 16 different languages as a direct replication and extension of previous work. We compare data collected in 2017 with data collected in 2021. Our analyses revealed that Google searches in English from within the United States still have the highest likelihood of triggering the display of additional crisis-prevention information prominently shown in addition to the regular search results (i.e., Google's suicide-prevention result). Searches in Spanish from within the United States are informationally disadvantaged. Display rates are only slightly lower in other English-speaking countries and when searches are performed in English. While information disparities and digital divides narrowed between 2017 and 2021, substantial differences in the display of crisis-prevention resources remain observable within multilingual countries, especially when other languages compete with English. In Bahrain, South Africa, and Sweden, the crisis-prevention information functionality seems unimplemented. Our findings suggest that the use of automated computational methods is both useful to continuously observe the implementation of new algorithmic functionalities and necessary to hold global media institutions accountable for their actions.

Google's corporate mission is to 'organize the world's information and make it universally accessible and useful' (Google, 2021b). That not only means providing straight facts when requested, but also adequate information that can be used to explore a topic from different angles. In that, Google not only dominates the global search engine market (Chaffey, 2022), but also customizes its search results, which, in turn, makes them appear slightly different for different searchers. Google has also implemented functionalities to detect 'emergency situations, such as suicide, poison concerns, sexual assault, or drug addiction' (Google, 2021a). For example, when detecting search behavior

CONTACT Sebastian Scherr scherr@tamu.edu Texas A&M University, College Station, TX, USA

potentially indicative of suicidal tendencies, Google displays a country-specific additional info box ('suicide-prevention result,' SPR) on top of the main search results containing crisis-prevention resources, including local telephone hotlines and chat offers.

As much as these efforts can be appreciated, Arendt et al. (2020) showed that even a slight modification, such as adding the name of a celebrity to a suicide-related search (e.g., '*avicii* suicide' relative to 'suicide'), totally inhibits Google's algorithms from displaying crisis-prevention resources. Similarly, Haim et al. (2021) showed that crisis-prevention resources were displayed much less frequently in response to potentially harmful, suicide-related searches when method-specific opioids or drugs were added (e.g., 'easy way to commit suicide *fentanyl*'). The display frequency of the Suicide Prevention Result (SPR; Scherr et al., 2019) has also been at the center of a set of three computational studies that provided a globally comparative estimate of the extent of this phenomenon and thus highlighted algorithmic territory disparities in global health information, driven mainly by region and language.

While acknowledging the research on global digital divides that might amplify gaps in access to health information over time (Norris, 2001), much less is known about whether and how digital divides change over time. Big Tech companies, including Google/Alphabet, do not only not disclose their algorithmic decision-making processes, but have also been called out for 'black box gaslighting' (Cotter, 2021) – a concept according to which opaque decision-making is leveraged to hinder empirical insights and thus to hamper investigations into algorithmic accountability while claims about the potential benefits of artificial intelligence and machine learning being able to close existing digital divides also largely remain empirically untested. The largest comparative study to date that compared the display rate of Google's SPR in 11 countries using data collected in 2017 (Scherr et al., 2019) revealed a display rate of slightly over 90% after potentially harmful search terms and about 55% after potentially helpful search terms – when the searches were performed in English, and at best less than half of that if another language was used. Given the scarcity of research on *changes in Google's information disparities as well as in informational digital divides* at a global level, combined with no feedback from Big Tech on such empirical observations, we refrained from hypothesizing change, but instead employed a longitudinal monitoring approach and asked: Did the SPR's display frequency as an indicator for access to ad hoc health crisis information change between 2017 and 2021 worldwide?

In order to answer our research question, we used agent-based testing as a digital research method to perform a total of 1,607,145 Google searches in 17 countries (Australia, Bahrain, Brazil, Canada, China, France, Germany, India, Ireland, Italy, Japan, Singapore, South Africa, South Korea, Sweden, the United Kingdom, and the United States) and in 16 languages (Afrikaans, Arabic, Cantonese, English, French, German, Hindi, Italian, Japanese, Korean, Malayan, Mandarin, Portuguese, Spanish, Swedish, and Telugu) in the domain of suicide prevention as a direct replication and monitoring extension of Scherr et al. (2019). We will show that Google's SPR is now more consistently shown at higher rates both after potentially harmful and helpful searches, but still suffers from a global digital divide that contributes to disparities in access to ad hoc health crisis information in 'other-than-English-speaking' countries and within multilingual countries.

Google's suicide-prevention efforts: laudable, but in need of improvement

For over a decade, Google Search has displayed an additional SPR – an info box with country-specific helpline telephone numbers, websites, and/or chats (see Figure 1 for a screenshot), positioned at the very top of the results page if triggered by 'certain search queries' (Zeiger, 2010). Search terms that trigger the display of the SPR can be either indicative of potentially harmful (e.g., 'most effective way to commit suicide') or helpful intentions (e.g., 'how to overcome suicide thoughts'), and in both cases, the SPR is displayed right at the moment when such thoughts are translated into actual search behavior.

Originally, a Google employee raised the company's awareness of suicide as an issue (Cohen, 2010) that could be addressed by implementing crisis information in response to potentially harmful searches related to suicide (e.g., method-related information on how to kill oneself). Google consequently implemented an info box at the top of all search

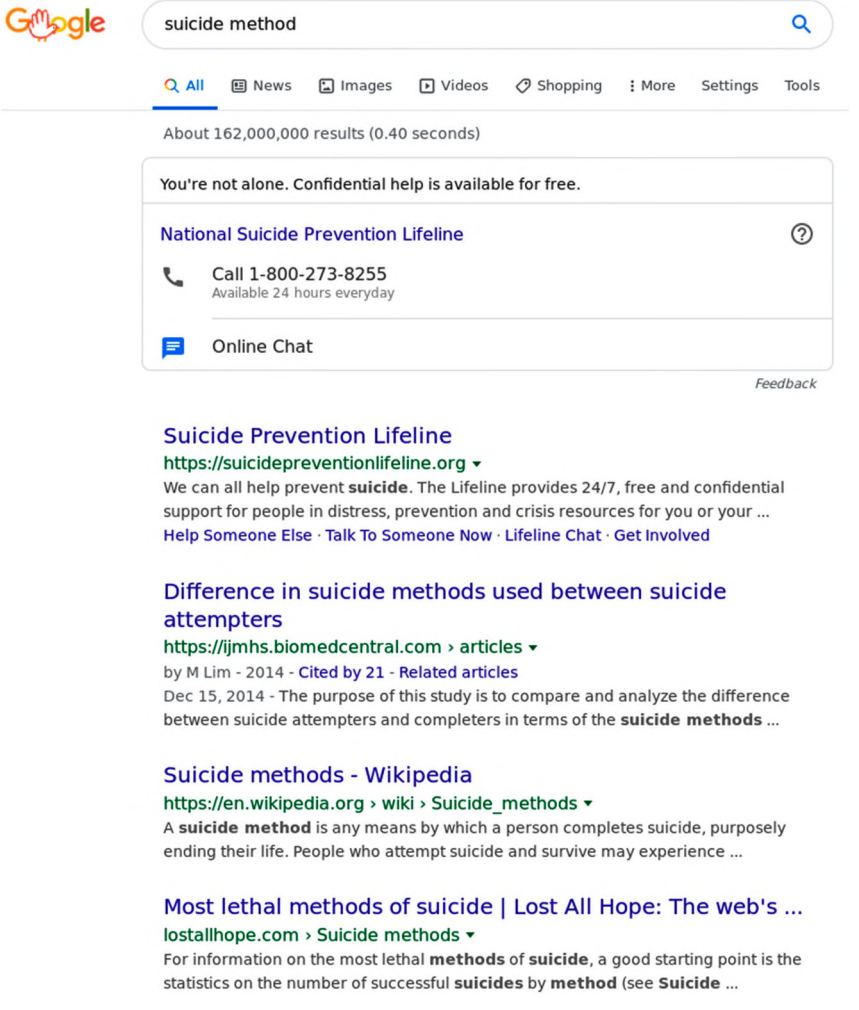

Figure 1. Screenshot of Google's Suicide-Prevention Result (SPR) on Top of the Regular Search Results

results when a suicide-related, potentially harmful search term was entered. The SPR originally contained the National Suicide Prevention Lifeline and was available in 14 countries (Australia, Belgium, France, Germany, Hungary, Italy, the Netherlands, New Zealand, Norway, Spain, Sweden, Switzerland, the United Kingdom, and the United States). In November 2010, Google estimated the increase in hotline calls in the United States to be 9% in the months after having implemented the SPR (Google Official Blog, 2010, November) and planned to implement the technology into its software in more countries.

Importantly, suicide-related searches can be indicative of acute suicidality (Arendt & Scherr, 2017). However, there is incredibly little evidence for the effects of the SPR implementation, and we are only aware of two pieces of anecdotal evidence in this regard. First, the previously cited Google blog entry quotes a 9% increase in 'legitimate calls' as reported by the National Suicide Prevention Lifeline in the United States within an unspecified timeframe after the implementation. Second, increased phone calls are a desirable outcome given that the use of telephone counseling services can decrease suicidality (Gould et al., 2018). Third, there is reported evidence that by linking Google advertisements to suicide-related search terms, one-in-three users who saw such a Google ad also clicked on it (Berlinquette, 2019). We see only small negative repercussions of 'false positives' in this case, that is, when people see suicide-related help information when googling for something suicide-related even though they are not in an acute suicidal crisis. Taken together, it is fair to say, from a suicide-prevention perspective, that high Google SPR display rates are desirable and likely to help in saving lives.

However, there is room for improvement: The SPR display rates not only vary across search queries due to baked-in yet largely unknown algorithmic mechanisms, but the SPR display algorithm is also sensitive toward specific keywords and is not so much determined by the users' past browsing behavior (Haim et al., 2017). Moreover, the display frequency of the SPR is much higher after potentially harmful searches as compared to those indicative of help-seeking intentions; it varies across countries and with the language used to perform the search; and when additional words are added to the search (e.g., the name of a celebrity who died by a specific suicide method, or the name of an opioid used for overdosing), the SPR display frequency drops substantially (Arendt et al., 2020; Haim et al., 2017; Scherr et al., 2019). The current algorithmic decision-making around Google's SPR is therefore best described as subpar. This is problematic because the SPR has been developed as an automated nudge (Thaler & Sunstein, 2008) toward emergency information for people in need of help and independent of whether their intentions are helpful or harmful at the moment of their search; as of now, multiple factors prevent the SPR from unlocking its full suicide-preventive potential, which was originally set up to support Google's corporate mission.

Digital divide dynamics: how language prevents them from narrowing

Nudges, prompts, or primes delivered through various channels, including fictional programs or public health campaigns, have long been known to be an effective tool to promote health, including suicide prevention (Kennedy et al., 2004; Knox & Bossarte, 2009;

Thaler & Sunstein, 2008). Hence, Google's own data and previous health communication research are consistent with the assumption that the display of the SPR is likely to elicit net beneficial effects. This is also in line with conceptualizations of suicidality as a disorder that can be simultaneously characterized by a restricted suicidal tunnel vision with suicidal plans at the top of the minds of suicidal people (Wasserman, 2016a, 2016b), but also as a state of high ambivalence toward life and death (Wasserman, 2016a). Thus, even small nudges, such as Google's SPR, should be able to prevent suicide, especially if they reach suicidal people at the right time, that is, in a moment of ambivalence where even small nudges might be enough to change their direction. As such, the SPR can be understood as just-in-time, health crisis information that is presented to individuals when their actual search behavior triggers Google's algorithm to display an ad hoc crisis intervention prominently on top of the regular search results.

However, previous research in this field (Arendt et al., 2020; Haim et al., 2017; Scherr et al., 2019) shed light on the cross-country disparities in the display frequency of Google's SPR. Conceptually, only showing health crisis information to some users under some conditions contributes to digital inequalities (DiMaggio & Hargittai, 2001) and ultimately to a 'digital divide' (Norris, 2001) between users regarding their access to health crisis information. The two concepts were borrowed from descriptions of socioeconomic disparities that were responsible for different degrees of physical access to computers and the Internet, and have been refined as second-level digital divides (i.e., differences in Internet use) since Internet-access gaps largely disappeared (Pew Research Center, 2009). Second-level digital divides include disparities due to social or racial/ethnic factors (Hargittai & Hinnant, 2008; Zillien & Hargittai, 2009) and therefore implicitly touch upon language as another potential contributing factor for online information disparities. Second-level digital divides have been conceptualized as complex, dynamic, and with a tendency to further deepen rather than narrow automatically (van Dijk, 2005; van Dijk & Hacker, 2003), and therefore they more strongly represent problems relating to inclusivity, institutions, and individual proficiency (Chakravorti, 2021).

Although research on Internet-use inequalities eventually tapped into the health effects of search engines and social media (Kontos et al., 2010) and also discussed digital divides against the backdrop of content production and geography (Robinson, 2009; Robinson et al., 2015; Schradie, 2011), longitudinal research focusing on the *changes in digital divides over time* (i.e., the narrowing or widening of digital divides) is scarce. As in arguments by Warschauer (2002), even if equal technological access was granted, language, media literacy, as well as structural influences would still maintain digital divides (see also Graham, 2011). To date, we have also learned that we must include algorithms on the list and consider their main coding language (= English) as an additional contingency factor for their functionality – content-related flags, filters, and functions work better in English, which makes today's digital disparities appear to be a code-language issue (Scherr et al., 2019; Valencia, 2021). It remains largely unclear how the Big Tech companies will improve their code functionality in languages other than English over time, and to what extent this can be related to digital divides as part of health information disparities.

Information disparities at google, implementation gaps, and a need for continuous monitoring

Between 2017 and 2019, multiple studies showed that the crisis-prevention resources for suicide (SPRs) were not equally displayed for searches that a) implied help-seeking intentions (e.g., 'suicide help') as opposed to harmful intentions (e.g., 'easy way to commit suicide'); b) were performed from English-speaking countries; c) were performed in English as compared to other languages, even within multilingual countries; and d) contained additional search terms that further specified a suicide-related search either by adding the name of a celebrity suicide or drugs and opioids used for overdosing (Arendt et al., 2020; Haim et al., 2017; Haim et al., 2021; Scherr et al., 2019). Implementers at Google either did not understand these facets in their full complexity (from a suicide-prevention perspective, a high, across-the-board display frequency is desirable), or they underestimated the impact of their work (it remains unclear to what extent displayed help resources boost help-seeking activities) and therefore left gaps in the implementation. *Implementation gap theory* (Hudson et al., 2019) helps in understanding how internal processes can contribute to the subpar execution of implementing changes in complex systems, and strongly recommends systematically observing the implementation steps taken in order to hold Big Tech accountable. If the algorithms cannot be satisfactorily observed, there will be no institutional accountability for Big Tech and no exercise of power by governments, NGOs, and essentially all those outside of Big Tech.

In the following empirical part of this manuscript, we present a direct replication and monitoring extension of the most recent observations of digital health information disparities as produced by Google's algorithm for its so-called SPR (Scherr et al., 2019). This study is an empirical example that demonstrates how continuous, external evaluations of the functionality of algorithms not only shed light on the changes (i.e., the narrowing or widening) of existing digital divides, but also define areas of accountability, and might therefore contribute to pushing Big Tech to actually help making the world a better place. As of the writing of this manuscript, we are not aware that a system of continuous evaluation of the implementation of Google's SPR has been put in place, and by arguing for the notion that digital divides are unlikely to narrow themselves without further input (van Dijk, 2005; van Dijk & Hacker, 2003), we reused and cast the most recent observations (Scherr et al., 2019) into three hypotheses:

H1: There are significant differences in the display rates of Google's SPR between a set of 17 countries.

H2: There are significant differences in the display rates of Google's SPR within a subset of countries with more than one official language.

H3: There are significant differences in the display rates of Google's SPR between a subset of English-speaking countries.

Method

Digital method setup

We employed an agent-based testing (Haim, 2019) approach as a computational method that allowed us to continuously observe the implementation of Google's SPR over time

(Haim, 2020). Building on the ScrapeBot implementation, we emulated human-like search behavior on a large scale. In total, virtual agents performed $N = 1{,}607{,}145$ Google web searches over the course of several weeks in 2021 as part of three separate studies. We used all available locations from Amazon Web Services (AWS) at the time the studies were performed to locate our study. Specifically, the servers were located in Australia, Bahrain, Brazil, Canada, China (Hong Kong), France, Germany, India, Ireland, Italy, Japan, Singapore, South Africa, South Korea, Sweden, the United Kingdom, and the United States. Installing virtual agents solely on AWS servers not only allowed us to emulate search behavior that Google cannot distinguish from actual searches performed by locals in the country (mostly through IP traceback/backtracking and more reliably than by using VPN clients), but it also creates a globally comparable study setup. To further increase the external validity of our approach, we increased within-country variance by using two servers per server location and made use of multiple server locations in countries where multiple locations were available (i.e., Japan, the United States). Besides provider-dependent assignments, we used constant IP addresses for our servers to equalize possible mediating effects. We used a central database for the maintenance and orchestration of all servers and virtual agents.

General digital procedure

We directly replicated and extended Scherr et al.'s (2019) study and used the same three lists of suicide-related helpful, suicide-related harmful, or suicide-unrelated search terms (see this OSF https://osf.io/hnmte/?view_only=927cdd5fd7fb466dafbb1c485c2a46a6). Suicide-related search terms were originally suggested by Biddle et al. (2008), unrelated terms were nouns, verbs, or adjectives from the Berlin Affective Word List (Võ et al., 2009) with a medium size arousal value. The original list of search terms was initially extended by suggestions from local suicide experts and translated by them or native speakers from each country. All translators were invited to further extend the list of search terms by adding additional, country-specific expressions. Translated into classic experimental research terminology, we emulated searches using three different lists of queries (i.e., independent variables) that built on (a) suicide-related potentially helpful, (b) suicide-related potentially harmful, and (c) suicide-unrelated control terms in line with existing studies (Arendt et al., 2020; Haim et al., 2017; Haim et al., 2021; Scherr et al., 2019).

Besides the different groups of search terms, all virtual agents had the same instructions: After each performed search, the display of the SPR was stored as the main (binary) outcome (i.e., dependent variable). Emulations were performed using 'ScrapeBot' (Haim, 2019), a group of Python scripts that use Selenium for Firefox to repeatedly run so-called recipes – a set of accurate instructions for virtual behavior on a distributed set of servers. Every run started 'fresh,' that is, ScrapeBot set up a new, realistic environment with a visual desktop display, legitimate language and location specifications, as well as real browser signatures. In order to resemble human search behavior, random slow typing was implemented, as well as waiting times and scrolling through search results.

Specifically, each recipe instructed virtual agents to first navigate to https://www.google.com/ then randomly draw one search term from the agent's experimental group's list of search terms, and type it into Google's query-input box. After submitting the

search, it stored whether the SPR was displayed (= 1) or not (= 0). Figure 1 shows an example of how the SPR looked in 2021.

Taking a screenshot of the search results for 5% of all runs per recipe allowed us to visually inspect that the system had operated correctly. We also scraped and stored all information boxes, including the SPR, as clear text for documentation. For backup reasons, we also stored all links to organic, first-page search results as well as the first-page source code.

Across all three studies presented here, data were collected in 2021 between 6 March and 5 April (end dates included). On all AWS servers, ScrapeBot started operating every two minutes. Each ScrapeBot randomly picked recipes from the central database and ran them. During runtime, ScrapeBot (Haim, 2019) meticulously monitored its own behavior. Due to the stability of the SPR display frequencies previously observed (e.g., Scherr et al., 2019), the data presented here represent the daily share (in %) of searches for which the SPR had been displayed. To include standard errors, we also employed bootstrapping for the generation of 95% confidence intervals. All data, codes, and recipes as well as per-group and per-study break-downs of performed searches are available at OSF under https://osf.io/hnmte/.

Results

Study 1: Differences between countries narrowed between 2017 and 2021

The first study is a direct replication and monitoring extension of Scherr et al.'s (2019) global study in 17 countries. The original study was based on data collected in 2017 and tested for differences in the display rates of Google's SPR between all countries available through the AWS server environment at the time of data collection. In addition to the original study (Australia, Brazil, Canada, Germany, India, Ireland, Japan, Singapore, South Korea, the United Kingdom, and the United States), six countries were added for this 2021 replication (Bahrain, China, France, Italy, South Africa, and Sweden). For each country, we used the main spoken language by the majority of people and had the lists of search terms translated by professional translators fluent in these languages (i.e., Afrikaans, Arabic, Cantonese, English, French, German, Hindi, Italian, Japanese, Korean, Malayan, Mandarin, Portuguese, Spanish, Swedish, and Telugu). Figure 2 shows the relative display frequencies of the SPR based on a total of $N = 596{,}068$ Google searches performed between 6 and 12 March, 2021 (end dates included). Random inspections of the screenshots that were taken revealed no irregularities other than the main dependent variable (i.e., the SPR) being displayed or not.

We observed some similarities and several changes from 2017 to 2021: First, the SPR display rates remain substantially higher after harmful ($_{hrm}$) as compared to helpful ($_{hlp}$) suicide-related searches, and the control terms continue to never trigger any SPR displays anywhere. Second, in all English-speaking countries (i.e., Australia, Canada, Ireland, Singapore, the United Kingdom, and the United States), the SPR display rate after harmful search terms is now consistently over 90%; these numbers increased in all countries that had been observed in 2017 (Australia: +8.0%; Canada: +100%; Ireland: +7.9%; the United Kingdom: +8.2%; the United States: +8.4%). For helpful search terms, the display frequency in English-speaking countries is now around 68%, second only to Germany,

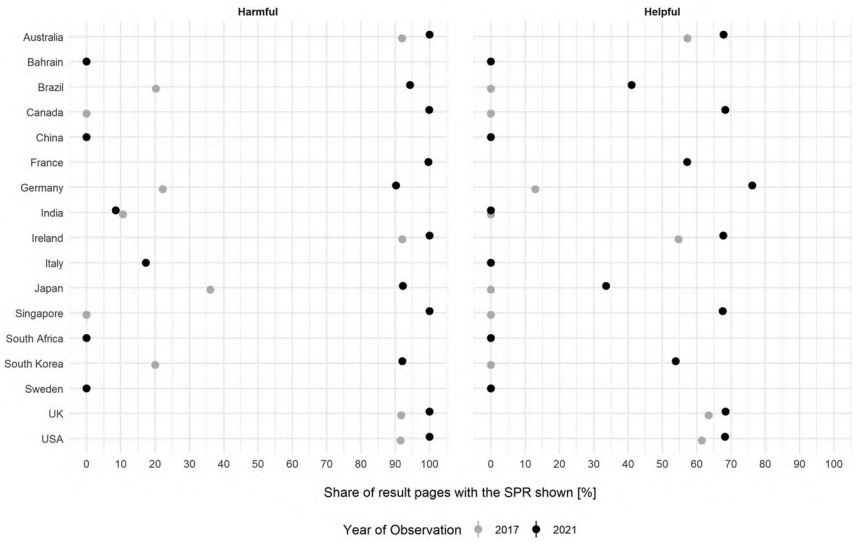

Figure 2. Between-Country Differences in the Display Frequency of Google's Suicide-Prevention Result (SPR) Note. Countries with only one black data point represent countries that were not included in the previous study (i.e., Bahrain, China, France, Italy, South Africa, and Sweden).

where it was higher, at 76.2%. Third, between 2017 and 2021, the SPR display frequency increased substantially in most of the other countries after both harmful and helpful search terms (Brazil: +7.4%$_{hrm}$ / +41.0%$_{hlp}$, Germany: +68.0%$_{hrm}$ / + 63.2%$_{hlp}$, Japan: +56.2%$_{hrm}$ / +33.6%$_{hlp}$, and South Korea: +72.0%$_{hrm}$ / +53.9%$_{hlp}$), and in Canada, the SPR seems to only just have been implemented in English (Canada: +100%$_{hrm}$ / +68.3%$_{hlp}$). Thus, the digital divide regarding algorithmic health crisis information disparities narrowed between 2017 and 2021 in all countries observed in both years, except for in India, where the gap widened (−2.1%$_{hrm}$ / ±0%$_{hlp}$), and in China, where the SPR is still never displayed. New first-time data is reported for Bahrain (0%), South Africa (0%), and Sweden (0%), as well as Italy (17.3%$_{hrm}$ / 0%$_{hlp}$), and France (100%$_{hrm}$ / 57.2%$_{hlp}$). Taken together, there are still significant differences between countries in regard to the display frequency of Google's SPR, which makes us accept H1, but we also observed a substantial narrowing of the digital divide in many countries, which answers the question that guided this research in the first place. However, while the divide is narrowing in many countries, some countries have been 'left behind' thus discrediting Google's mission statement.

Study 2: differences within multilingual countries remained between 2017 and 2021

For an investigation of the SPR display frequency in countries with more than one official language that is widely spoken, we used the same study setup as in Study 1, except for the observation time (14 March through 20 March, 2021, end dates included) and the number of searches performed (N = 645,164). Since individuals are likely to search in their language, that is, the one that they feel most comfortable with, we tested Spanish and

English for the United States, Afrikaans and English in South Africa, English and French in Canada, Hindi, Telugu, and English in India, Arabic and English in Bahrain, Mandarin, Malayan, and English in Singapore, and Mandarin, Cantonese, and English in China (see Figure 3).

Our analysis shows that substantial language-based within-country discrepancies persist. Among those countries investigated for the first time as part of a multilingual analysis in 2021, both in South Africa (English, Afrikaans) and Bahrain (Arabic), the SPR is never displayed. In China, the SPR was never displayed for searches in Cantonese or Mandarin, but it was for English searches ($92.3\%_{hrm}$ / $31.6\%_{hlp}$). In the United States, the SPR was virtually always displayed after searches in English ($100\%_{hrm}$ / $68.3\%_{hlp}$) as in Study 1, but substantially less so when people searched in Spanish ($94.0\%_{hrm}$ / $25.4\%_{hlp}$) – particularly in regard to helpful information. Thus, searches for suicide-prevention resources within the United States performed in English are 2.7 times as likely to immediately receive this information as part of the SPR as compared to US-based searches in Spanish. In Canada, however, both searches in English ($100\%_{hrm}$ / $68.9\%_{hlp}$) and French ($99.3\%_{hrm}$ / $62.0\%_{hlp}$) yielded similar SPR display frequencies. In India, searches in English yielded comparably substantial display frequencies, as elsewhere ($100\%_{hrm}$ / $68.4\%_{hlp}$), while only some harmful searches in Hindi ($12.9\%_{hrm}$ / $0\%_{hlp}$), but no searches at all in Telugu triggered any SPR displays – thus, having immediate access to suicide-crisis information in India is 7.8 times as likely for potentially

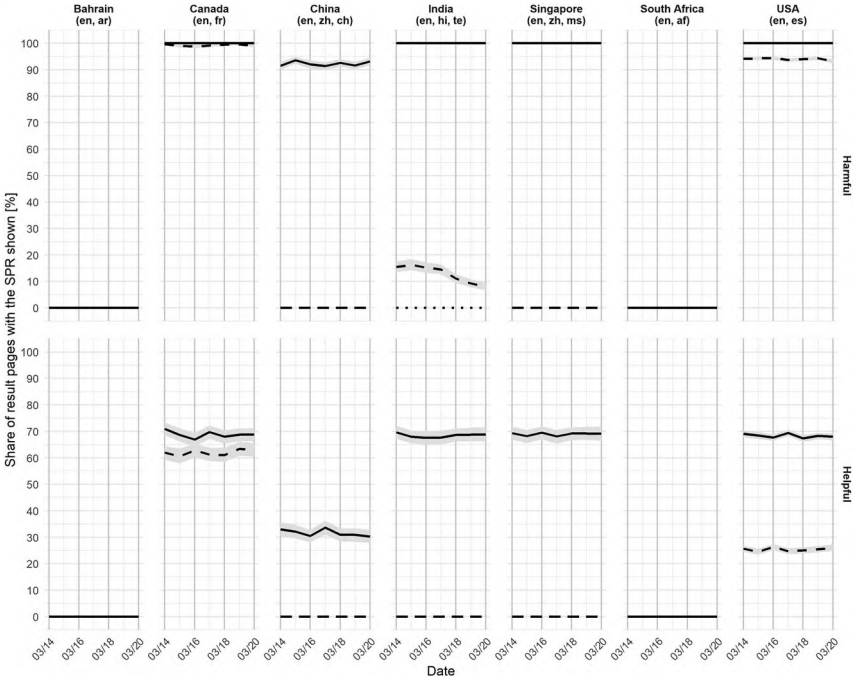

Figure 3. Within-Country Differences in the Display Frequency of Google's Suicide-Prevention Result (SPR) Note. Solid lines correspond to the first language mentioned in parentheses (i.e., English), dashed lines correspond to the second language mentioned, and dotted lines to the third language. 95% confidence intervals in gray are based on $n = 1000$ bootstrapping samples.

harmful Google searches in English than for searches in Hindi, and is impossible to achieve in Telugu. In Singapore, we also found the highest display frequencies after English searches ($100\%_{hrm}$ / $68.9\%_{hlp}$) and no displays for searches in Mandarin or Malayan. These observations strongly support hypothesis H2.

Study 3: Differences between English-speaking countries

Finally, we looked at the SPR display frequency in English-speaking countries using the same study setup as in Studies 1 and 2, except for the observation time (30 March through 5 April, 2021, end dates included) and the number of searches performed (N = 365,913). We only compared the SPR display frequency after Google searches in English (see Figure 4).

The United States yields significantly higher display frequencies both for harmful and helpful suicide-related searches ($100\%_{hrm}$, $68.3\%_{hlp}$) than any other country. Significantly lower, yet similar display frequency rates were observed for the United Kingdom ($93.3\%_{hrm}$, $37.6\%_{hlp}$), Singapore ($93.3\%_{hrm}$, $37.4\%_{hlp}$), Australia ($93.8\%_{hrm}$, $36.6\%_{hlp}$), Canada ($93.7\%_{hrm}$, $36.7\%_{hlp}$), Ireland ($94.5\%_{hrm}$, $36.7\%_{hlp}$), and India ($93.1\%_{hrm}$, $36.4\%_{hlp}$). In China, in turn, the display frequency is slightly but significantly lower ($92.2\%_{hrm}$, $31.4\%_{hlp}$), and both in South Africa and Bahrain, the SPR never appeared.[1] We therefore accept H3, confirming significant differences in the display rates of Google's SPR between countries in which English is widely spoken.

Discussion

We employed agent-based testing to emulate 1.6 million human-like Google searches in order to evaluate the implementation of an algorithmic functionality at Google Search

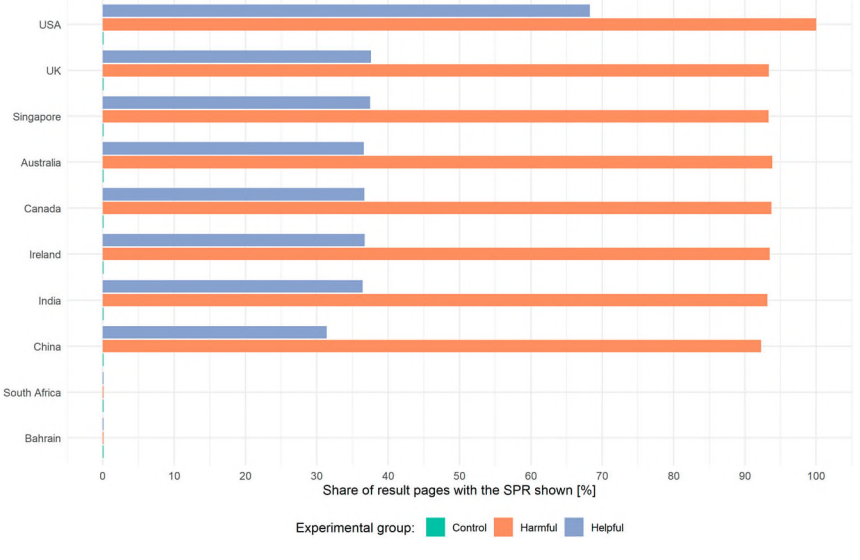

Figure 4. Differences in the Display Frequency of Google's Suicide-Prevention Result (SPR) in English-Speaking Countries

that aims at providing crisis-prevention resources. Google's SPR has been in place since 2010 and had previously been evaluated on a global scale (Scherr et al., 2019). We replicated and extended this work and especially focused on monitoring the *changes in digital divides* previously observed as part of algorithmic, language-based health information disparities now in 17 countries (Australia, Bahrain, Brazil, Canada, China, France, Germany, India, Ireland, Italy, Japan, Singapore, South Africa, South Korea, Sweden, the United Kingdom, and the United States) and in 16 languages (Afrikaans, Arabic, Cantonese, English, French, German, Hindi, Italian, Japanese, Korean, Malayan, Mandarin, Portuguese, Spanish, Swedish, and Telugu).

Our analysis revealed that digital divides both have and have not narrowed since the last evaluation based on data collected in 2017, that is, 4 years ago (Scherr et al., 2019). There are four main findings for the present 2021 assessment: First, in most countries where Google's SPR is implemented as a functionality, users see suicide-crisis information over 90% of the time when they search for something suicide-related and potentially harmful. Users also see such information about 60–70% of the time when they are actively looking for help in a suicide-related, potential crisis. Most display frequencies increased between 2017 and 2021, indicating the narrowing of the language-based and algorithmically sustained digital divide. This is a considerable improvement. Yet, Google still displays the SPR more frequently after suicide-related, potentially harmful searches as compared to searches that are likely to involve seeking help, despite the low cost of 'false positives.' Second, the SPR display frequency is still higher after searches are performed in English and is now almost always displayed in English-speaking countries. Informational discrepancies for the benefit of English-language searches therefore continue to exist between English- and non-English-speaking countries and within multilingual countries in which English is also spoken. Third, miscellaneous observations included slightly higher SPR display frequencies in English-speaking countries since 2017, the initiation of the implementation of the SPR in Canada, which must have occurred since 2017, the initiation of entirely new observations beginning in 2021 (France, Italy, Bahrain, China, South Africa, and Sweden), and interestingly, India, which was the only country where we observed a slight decline in the SPR display frequency.

Our findings underline the importance of a continuous, external assessment of the implementation of new functionalities within algorithmically curated information spaces on a global scale. Without such external assessment, accountability and control over how algorithmic functionalities are implemented and how they effectively operate over time remains unclear, and only speculations can be made about how global informational disparities prevail over extended periods of time. Implementation gap theory (Hudson et al., 2019) suggests an entangled process of algorithmic development and its ongoing evaluation to avoid implementation gaps in the first place, and to monitor and be able to close implementation gaps along the way. In the case of Google's SPR, English-speaking individuals in the United States had a 100% chance of receiving crisis-prevention information from Google directly, on top and apart from the regular search results right at the moment when they googled for a suicide-related, potentially harmful term, such as 'most effective way to commit suicide'; the chances were about 68% when their search implied help-seeking intentions (e.g., 'how to overcome suicide thoughts'). However, the same search terms typed into Google in Spanish, but also from within the United

States, made the same suicide-prevention information available only in 95% of potentially harmful searches and in only 25% of help-seeking searches, which highlights the divide within the United States.

Without a systematic, continuous evaluation of the display rates, we have to rely on Google's word that such information disparities will be addressed and eliminated. The SPR is just one example of an area in which the functionality of the Google algorithm became at least somewhat visible. There is a need for evaluation during every step of the implementation process, both for successful implementation and for productive and uninterrupted use by diverse groups of users across the globe. Extended global access to algorithmically curated media environments will increase the need for new, systematic evaluation methods, including agent-based testing, in order to simulate different scenarios and communities using different languages, terms, and expressions in which an algorithm is operating. Our 2021 analyses suggest that Google has closed some of the implementation gaps since 2017, but our observations in India, but also within the United States, suggest that these changes are not necessarily reflective of the entire environment that Google, as a global company, taps into. A combination of continuous evaluations and strengthening local algorithm development that also takes cultural differences in how people express themselves vis-à-vis search engines could be a sustainable solution (see also Valencia, 2021).

Implementation gaps are not necessarily visible to experts, politicians, or customers, and therefore they elude themselves from public scrutiny, which makes them hard to detect. We can only speculate about the origins of the observed information disparities for Google's SPR, and there is no way to see from the outside whether Google omitted some of the English search terms in other languages or incorrectly translated them, thus resulting in lower display frequencies in other countries. Surprisingly, our questions to the Tech giant have remained unanswered to this day. Observed challenges are admittedly complex for Google given that users cannot only influence their localization virtually (e.g., through the use of virtual private networks; VPNs), but also due to increasing global mobility and the changing map of global language diversity (Hua et al., 2019). Substantial within-country gaps in the display frequency of the SPR are not limited to Spanish in the United States—the same applies to other countries and to other global languages, such as Mandarin Chinese and Hindi.

Finally, Google's SPR might affect the user search experience regarding user retention and engagement when the SPR is shown unintentionally (from a user's perspective). Furthermore, Google's SPR might as well produce an '*Info Box Effect*' on users. There is limited evidence that the presentation of additional info boxes (e.g., Google's Knowledge Panel) significantly reduces the overall search time and the number of actual search results people clicked on (Epstein & Mohr, 2018). Against this backdrop, one could ask under what conditions Google decides to display additional info boxes and with what effects on information seeking given that an info box display might affect the attention paid to ranked search results. Further studies should more closely look at this.

Limitations

We evaluated Google's SPR display frequency across countries, languages, and over the course of three short time periods. Although previous research revealed that the display

rates remain very stable over time (Haim et al., 2017; Scherr et al., 2019), we cannot rule out or generalize from these studies to make similar observations in all countries, especially those that have not been assessed before. There might also be stronger short-term variability in the display frequency after special circumstances, such as prominent celebrity suicides (Arendt et al., 2020). Although we carefully watched the news during the data-collection periods and systematically collected screenshots as control samples of the functionality of our approach, and even though we were not aware of any prominent suicides, there might have been local suicide news reports, and there is a chance that these could have triggered different local responses at Google. Longer observation periods could shed light on such possible variations. Moreover, we restricted our analysis to all server locations that were available through the Amazon AWS environment at the time of data collection. Alternative approaches would allow for a higher number of countries – however, at the cost of less comparability. For the same reason, we also used the same word lists as those used in previous studies (e.g., Scherr et al., 2019) instead of changing or extending them. Finally, local contexts as well as cultural differences on how to deal with the issue of suicide might have had an impact on the literal and social translations of the search terms, and in turn, Google search results might as well acknowledge them. This limitation could apply both to within and between country differences, e.g., a query that clearly suggests a crisis in one language but less so in another language. It also remains unclear in how far Google's algorithm acknowledges or discounts for local suicide rates. Future research could further investigate if Google's SPR can be a more effective nudge, prompt, or prime if it varied more according to culture and geography.

Conclusion

Holding today's 'Big Tech' accountable is a future challenge for which useful tools and procedures have yet to be developed. We targeted Google's effort of additionally displaying crisis-prevention resources that are displayed in addition to the regular search results as an informational shortcut for those in immediate need. We simulated over 1.6 million Google searches in the domain of suicide prevention using human-like, agent-based testing and compared the display frequencies of Google's suicide-prevention result (SPR) in 17 countries and for searches performed in 16 languages. While we still found language-based, algorithmic information disparities (i.e., crisis-prevention information that was not always displayed when potentially needed) both between and within countries, we also observed higher display rates in English-speaking countries as compared to 2017. Only the continuous, external evaluation of the algorithmic functionality can help scholars to evaluate algorithmic content curation and to help policymakers hold both media and tech institutions accountable for the huge social impact they have.

Note

1. As a control condition, we also included a list of search terms that were unrelated to suicide. The SPR never showed up and showed up nowhere in response to any of the control terms (which is not explicitly mentioned in the results).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

There has been no financial support that could have influenced the outcome.

Data availability statement

All data is available open access from their OSF repository under https://osf.io/hnmte/.

Notes on contributors

Sebastian Scherr (PhD, University of Munich) is an Assistant Professor of Health Communication at the Department of Communication at Texas A&M University, USA. His research interests focus on individual and structural susceptibility factors for media effects in the domains of health and political communication, with a special emphasis on mental health, suicide prevention, and empirical methods.

Florian Arendt (PhD, University of Vienna) holds the Tenure Track Professorship in Health Communication at the Department of Communication, University of Vienna, Austria. His research focuses on health communication with a special emphasis on suicide prevention.

Mario Haim (PhD, University of Munich) is a Full Professor of Communication Science, especially Computational Communication Research, at the Department of Media and Communication at LMU Munich, Germany. His main research interests are computational journalism, news use within algorithmically curated media environments, and computational social science.

ORCID

Sebastian Scherr http://orcid.org/0000-0003-4730-1575
Florian Arendt http://orcid.org/0000-0003-1107-8682
Mario Haim http://orcid.org/0000-0002-0643-2299

References

Arendt, F., Haim, M., & Scherr, S. (2020). Investigating google's suicide-prevention efforts in celebrity suicides using agent-based testing: A cross-national study in four European countries. *Social Science & Medicine*, *262*, Article 112692. https://doi.org/10.1016/j.socscimed.2019.112692

Arendt, F., & Scherr, S. (2017). Optimizing online suicide prevention: A search engine-based tailored approach. *Health Communication*, *32*(11), 1403–1408. https://doi.org/10.1080/10410236.2016.1224451

Berlinquette, P. (2019). I used Google ads for social engeneering. It worked. *The New York Times*. https://www.nytimes.com/2019/07/07/opinion/google-ads.html

Biddle, L., Donovan, J., Hawton, K., Kapur, N., & Gunnell, D. (2008). Suicide and the internet. *BMJ*, *336*(7648), 800–802. https://doi.org/10.1136/bmj.39525.442674.AD

Chaffey, D. (2022). *Search engine marketing statistics* 2022. https://www.smartinsights.com/search-engine-marketing/search-engine-statistics/

Chakravorti, B. (2021). How to Close the Digital Divide in the U.S. *Harvard Business Review*. https://hbr.org/2021/07/how-to-close-the-digital-divide-in-the-u-s

Cohen, N. (2010). *'Suicide' query prompts Google to offer hotline*. http://www.nytimes.com/2010/04/05/technology/05google.html

Cotter, K. (2021). "Shadowbanning is not a thing": black box gaslighting and the power to independently know and credibly critique algorithms. *Information, Communication & Society*, 1–18. https://doi.org/10.1080/1369118X.2021.1994624

DiMaggio, P., & Hargittai, E. (2001). From the 'digital divide'to 'digital inequality': Studying Internet use as penetration increases. *Princeton: Center for Arts and Cultural Policy Studies, Woodrow Wilson School, Princeton University*, *4*(1), 4–2.

Epstein, R., & Mohr, R. (2018, April). *The answer bot effect (ABE): Another surprising way search engines can impact opinions* 98th Annual Meeting of the Western Psychological Association, Portland, OR. https://aibrt.org/downloads/EPSTEIN_%26_MOHR_2018-WPA-The_Answer_Bot_Effect-ABE-WP_17_04.pdf

Google. (2021a). *Find crisis prevention resources with Google search*. Retrieved 01.07.2021 from https://support.google.com/websearch/answer/9988513?*p*=crisis_prevention_info

Google. (2021b). *Our approach to search*. Retrieved 01.07.2021 from https://www.google.com/search/howsearchworks/mission/

Google Official Blog. (2010, November). *Helping you find emergency information when you need it* Retrieved 23rd June 2017 from https://googleblog.blogspot.de/2010/11/helping-you-find-emergency-information.html

Gould, M. S., Lake, A. M., Galfalvy, H., Kleinman, M., Munfakh, J. L., Wright, J., & McKeon, R. (2018). Follow-up with callers to the National Suicide Prevention Lifeline: Evaluation of callers' perceptions of care. *Suicide and Life-Threatening Behavior*, *48*(1), 75–86. https://doi.org/10.1111/sltb.12339

Graham, M. (2011). Time machines and virtual portals. *Progress in Development Studies*, *11*(3), 211–227. https://doi.org/10.1177/146499341001100303

Haim, M. (2019). *ScrapeBot: A selenium-based tool for agent-based testing*. Retrieved 01.07.2021 from https://github.com/MarHai/ScrapeBot/

Haim, M. (2020). Agent-based testing: An automated approach toward artificial reactions to human behavior. *Journalism Studies*, *21*(7), 895–911. https://doi.org/10.1080/1461670X.2019.1702892

Haim, M., Arendt, F., & Scherr, S. (2017). Abyss or shelter? On the relevance of web search engines' search results when people google for suicide. *Health Communication*, *32*(2), 253–258. https://doi.org/10.1080/10410236.2015.1113484

Haim, M., Scherr, S., & Arendt, F. (2021). How search engines may help reduce drug-related suicides. *Drug and Alcohol Dependence*, *226*, Article 108874. https://doi.org/10.1016/j.drugalcdep.2021.108874

Hargittai, E., & Hinnant, A. (2008). Digital inequality: Differences in young adults' use of the Internet. *Communication Research*, *35*(5), 602–621. https://doi.org/10.1177/0093650208321782

Hua, X., Greenhill, S. J., Cardillo, M., Schneemann, H., & Bromham, L. (2019). The ecological drivers of variation in global language diversity. *Nature Communications*, *10*(1), Article 2047. https://doi.org/10.1038/s41467-019-09842-2

Hudson, B., Hunter, D., & Peckham, S. (2019). Policy failure and the policy-implementation gap: Can policy support programs help? *Policy Design and Practice*, *2*(1), 1–14. https://doi.org/10.1080/25741292.2018.1540378

Kennedy, M. G., O'Leary, A., Beck, V., Pollard, K., & Simpson, P. (2004). Increases in calls to the CDC national STD and AIDS hotline following AIDS-related episodes in a soap opera. *Journal of Communication*, *54*(2), 287–301. https://doi.org/10.1111/j.1460-2466.2004.tb02629.x

Knox, K. L., & Bossarte, R. M. (2009). *Suicide prevention research - Enabling activities funded through VA's office of mental health* 2009 Department of Defense/Veterans Administration Suicide Prevention Conference, San Antonio, TX.

Kontos, E. Z., Emmons, K. M., Puleo, E., & Viswanath, K. (2010). Communication inequalities and public health implications of adult social networking site use in the United States. *Journal of Health Communication*, *15*(Suppl 3), 216–235. https://doi.org/10.1080/10810730.2010.522689

Norris, P. (2001). *Digital divide: Civic engagement, information poverty, and the Internet worldwide*. Cambridge University Press.

Pew Research Center (2009). *Generations online in* 2009. Retrieved 31.10.2017 from http://www.pewinternet.org/2009/01/28/generations-online-in-2009/

Robinson, L. (2009). A taste for the necessary. *Information, Communication & Society*, *12*(4), 488–507. https://doi.org/10.1080/13691180902857678

Robinson, L., Cotten, S. R., Ono, H., Quan-Haase, A., Mesch, G., Chen, W., Schulz, J., Hale, T. M., & Stern, M. J. (2015). Digital inequalities and why they matter. *Information, Communication & Society*, *18*(5), 569–582. https://doi.org/10.1080/1369118X.2015.1012532

Scherr, S., Haim, M., & Arendt, F. (2019). Equal access to online information? Google's suicide-prevention disparities may amplify a global digital divide. *New Media & Society*, *21*(3), 562–582. https://doi.org/10.1177/1461444818801010

Schradie, J. (2011). The digital production gap: The digital divide and Web 2.0 collide. *Poetics*, *39* (2), 145–168. https://doi.org/10.1016/j.poetic.2011.02.003

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.

Valencia, S. (2021). Misinformation online is bad in English. But it's far worse in Spanish:Our research found Facebook, YouTube and other platforms aren't doing enough to combat falsehoods. *The Washington Post*. https://www.washingtonpost.com/outlook/2021/10/28/misinformation-spanish-facebook-social-media/

van Dijk, J. (2005). *The deepening divide: Inequality in the information society*. Sage.

van Dijk, J., & Hacker, K. (2003). The digital divide as a complex and dynamic phenomenon. *The Information Society*, *19*(4), 315–326. https://doi.org/10.1080/01972240309487

Võ, M. L.-H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin affective word list reloaded (BAWL-R). *Behavior Research Methods*, *41*(2), 534–538. https://doi.org/10.3758/BRM.41.2.534

Warschauer, M. (2002). Reconceptualizing the digital divide. *First Monday*, *7*(7). https://doi.org/10.5210/fm.v7i7.967

Wasserman, D. (2016a). Suicidal people's experiences of trauma and negative life events. In D. Wasserman (Ed.), *Suicide: An unnecesary death* (pp. 145–228). University Press.

Wasserman, D. (2016b). *Suicide: An unnecessary death (2 ed.)*. University Press.

Zeiger, R. (2010). Helping you find emergency information when you need it. *Official google.org Blog* http://blog.google.org/2010/11/helping-you-find-emergency-information.html

Zillien, N., & Hargittai, E. (2009). Digital distinction: Status-specific types of Internet usage. *Social Science Quarterly*, *90*(2), 274–291. https://doi.org/10.1111/j.1540-6237.2009.00617.x