Creativity in the digital age: The effect of search algorithms on the ideation process

Ben Wolfson, New York University, bw916@stern.nyu.edu
Moran Lazar, Tel Aviv University, moranlazar@tauex.tau.ac.il
Hila Lifshitz-Assaf, Warwick University, hdiginnovation@gmail.com

Introduction

Today, there is a burgeoning debate about the effect of search algorithms on our society. On the one hand, these algorithms provide user experiences that are consistent with user expectations, allowing users unparalleled access to information (Storm et al., 2017). On the other hand, these algorithms polarize society by exposing individuals to like-minded opinions while obfuscating contradictory ones (Haidt, 2022; Pariser, 2014). In the scholarly literature, the effects of search algorithms have been studied in the realms of social media and news recommendation (Chitra & Musco, 2020; Holone, 2016). Still, we know little about the impact of these same algorithms on creativity and innovation. Recent literature has called for further investigation of the use of algorithms on work-related knowledge creation and innovation (Wang & Nickerson, 2017). Our paper offers a new theoretical account investigating how the idea creation process is affected by algorithmic search tools.

In this paper, we theorize why and when algorithmic search impacts the creative heart of the innovation process, idea generation. We build upon the information systems literature on digital innovation and representation theory to suggest that the algorithmic search is affecting the knowledge search process in ways that are critical for the creative idea generation process. Specifically, we utilize an experimental research design to test the relationship between algorithmic search tools and idea creation. In this experiment, we manipulate the search process people go through by providing an *instrumental search* tool, Google, versus an *exploratory* search tool, a new algorithm created for this study).

Abbreviated Model

We propose a model that examines the effect of search algorithms for creative tasks The domain we consider is the domain of the creative task being searched for by the focal actor. In a somewhat recursive fashion, we conceive of the concepts relating to the creative task as the real-world phenomenon that the search algorithm system is representing. Our level of analysis is the individual social actor in their process of ideating. Our theoretical framework is built upon two main components that we define below: (1) Concept spaces (the overall concept space and the focal actor's concept space) and (2) algorithmic design logics, the logics that drive search algorithms.

Scholars have previously explored the connection between information encountered during the innovation process and the variety of ideas iterated upon by the focal actor (Austin et al., 2012; Jarvenpaa & Standaert, 2018). As we are theorizing a model of ideating with search algorithms, we start our model by defining ideas and the process of generating new ones. According to the creativity and innovation literature, ideas can be defined as "provisional and communicable representations… that are identifiable, discrete entities, such as specific mental representations or mental states, propositions or proposals, concepts or solutions." (Hua et al., 2022). This definition aligns with representation theory which defines a representation as "a sign[that] represents information about an object for a particular observer." (McKinney & Yoos, 2010). Our focal actor has their own concept space, which is the space of ideas that they have generated and accumulated over their lifetime, activated during their ideation process. Beyond this individual actor, there is a mass concept space of humanity, composed of the totality of human knowledge instantiated in concepts, including various forms of knowledge such as tacit, embodied, implicit, and explicit forms of knowledge (Nonaka, 1994). Over time, various domains accumulate, modify, and diffuse these knowledge core concepts. While most of the knowledge production and problem-solving take place within disciplinary boundaries, innovation usually takes place by importing or combining core concepts from a different domain than the original one (Henfridsson et al., 2018; Jeppesen & Lakhani, 2010).

Search algorithms

To elucidate the search process and connection between the search algorithms and humanity's concept space, we define the digital data space that search algorithms act upon. The units of the digital data space are digital information units or signs (Burton-Jones & Grange, 2013). These signs include text, images, and any other information that is digitized and accessible to the search algorithms. Each information unit may represent one or multiple concepts in the overall concept space. Search algorithms take in a subset of the digital data space (signs) as input and return an output that is a series of digital information units as well, termed algorithmic search output in our model. As stated above, the actor interprets the output of the search algorithm, representing it in their concept space. To date, search algorithms have been predominantly designed for informational search purposes, seeking to achieve search effectiveness. Larry Page, Google's founder, stated their search design vision, 'The ultimate search engine… would understand exactly what you mean and give back exactly what you want… Google didn't want to return thousands of links, it wanted to return one, the one you wanted." (Pariser, 2014)

Most search algorithms' filtering functions are based on semantic similarity between the user's query and the digital information unit. However, as there are frequently many results that are similar to the user's query, there are usually algorithmic design logics that narrow down and prioritize the results that are presented to the user. Google is designed based on popularity and personalization logics. Google originally relied on PageRank, an innovative internet-indexing algorithm that crawled the web and indexed results by popularity. Throughout the last decade, Google has bolstered the personalization of results, providing context-sensitive, user-specific results (Pariser, 2014). Each of these logics acts in a recursive, time-amplified fashion. As the personalization algorithm learns more about the user, it can provide more personalized results. Similarly, as the popularization algorithm pushes popular results to the top of users' results, these results become more popular.

Algorithms can be designed in various ways: there is no technological determinism on how they are designed and used (Bijker & Pinch, 1987). Recent research on algorithms and innovation has started exploring algorithmic tools that can enhance innovation. For instance, algorithmic tools to find new combinations of knowledge entities to produce innovation, including chemical and biological entities from academic papers (Foster et al., 2021), and patent databases (Youn et al., 2015), as well as building algorithmic tools that search for analogous knowledge components across knowledge domains (Hope et al., 2017). With this spirit, we argue that the way to counteract the effects of the popularity and personalization forces is by purposefully designing exploratory search algorithms. The search literature has broadly described exploratory search algorithms, as "more open-ended, persistent and multi-faceted" search that is needed when "searching for inspiring ideas about new products or services" (Taramigkou et al., 2017). We believe that there could potentially be multiple ways to design an exploratory search algorithm that will enhance creativity and overall create an augmented innovation process. In this paper, we compare the effect of the informational versus exploratory search process on idea creation. We propose that engaging in an exploratory (rather than informational) search process will enhance novel idea creation.

Methods

To test our theoretical model, we designed an exploratory search modification to Google's search algorithm. Specifically, many of the results that are indirectly related to the original search query are returned in the back pages of Google. Building on this logic, we created a tool based on Google's search results but reindexed them to be more diverse. We name our tool GoogleXYZ (developed for academic purposes only), because it promotes the "tail end", or the "XYZ" of the search results.

To promote diversity, we use natural language processing techniques to semantically cluster similar search results returned by Google. We then reindex the results using our algorithmically identified semantic clusters. Promoting diverse results, Google XYZ aims to inspire the user with a variety of concepts by diversifying the set of digital information units that they are exposed to. We note that there is a need to balance this logic with the baseline relevance logic (as complete diversity would essentially be randomness). Therefore, we balance relevance and variety. In addition, to counteract the personalization logic, our algorithm presents the same results to each user regardless of who they are and their past searches. Specifically, we generated a generic request for Google's information that is not sent from any historic Google user. Hence, the algorithm does not learn anything about the user, and therefore does not tailor results over time and biases the user towards their prior searches. Here, we generate results based solely on the underlying digital information units and the user query, without using metadata about the user. This logic has been gaining traction; DuckDuckGo, a privacy-based search engine has been growing exponentially in popularity in the last few years (Barry, 2021) and other competitors, such as Qwant getting growing attention.

We designed the search engine interface to present users with results from multiple semantic clusters on the first result page. We mimic Google's search interface to make sure differences stem from the search process, and not the user interface. To generate GoogleXYZ's search results, we take the user's initial query q, and retrieve the search results for a non-personalized user querying Google with query q. Each result, r, has both a title and description, which we concatenate into sentence strings, removing unnecessary punctuation. For each result, we tokenize its string, numericize it and then generate a semantic embedding using Sentence-BERT (Portenoy et al., 2022; Reimers & Gurevych, 2019). We then use a community detection algorithm based on cosine similarity to assign the embedding to a learned cluster (Reimers & Gurevych, 2019). All results that are not assigned a cluster are preliminarily put into one bigger cluster, noting that they are often a mixture of obscure paraphrases of existing clusters, foreign languages, or legitimately different results that could be their own clusters, but our algorithm does not find them another similar result to cluster with. When presenting GoogleXYZ's results to the user, we generate the list of results by drawing one result from every cluster (including the outliers) and adding them to the list of results. Once a cluster runs out of results we continue drawing from the remaining clusters until all clusters are empty. Usually, there is a less "informationally" relevant series of search results at the end of Google's returned results.

In the experiment, we manipulate the search process participants engage in. Specifically, participants are randomly assigned to one of two experimental conditions: informational versus exploratory search. Using a between-subject design, each participant has a chance to perform an online search related to a given problem we provide. After conducting the online search, participants are asked to generate a novel idea related to the problem they were exposed to. We will assess how different search processes (informational versus exploratory) affect the novelty of ideas participants create.

References

Austin, R. d., Devin, L., & Sullivan, E. e. (2012, January 1). Accidental Innovation: Supporting Valuable Unpredictability in the Creative Process. *ORGANIZATION SCIENCE*, *23*(5), 1505–1522.

Barry, P. (2021, April 5). *DuckDuckGo Usage Stats for 2022*. Backlinko.

Bijker, W., & Pinch, T. (1987). The Social Construction of Facts and Artifacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other. In *Social Studies of Science* (Vol. 14, pp. 17–50).

Burton-Jones, A., & Grange, C. (2013). From Use to Effective Use: A Representation Theory Perspective. *Information Systems Research*, *24*(3), 632–658.

Chitra, U., & Musco, C. (2020). Analyzing the Impact of Filter Bubbles on Social Network Polarization. *Proceedings of the 13th International Conference on Web Search and Data Mining*, 115–123.

Foster, J. G., Shi, F., & Evans, J. (2021). Surprise! Measuring Novelty as Expectation Violation. In *SocArXiv* (No. 2t46f; SocArXiv). Center for Open Science.

Haidt, J. (2022, April 11). *Why the Past 10 Years of American Life Have Been Uniquely Stupid*. The Atlantic.

Henfridsson, O., Nandhakumar, J., Scarbrough, H., & Panourgias, N. (2018). Recombination in the open-ended value landscape of digital innovation. *Information and Organization*, *28*(2), 89–100.

Holone, H. (2016). The filter bubble and its effect on online personal health information. *Croatian Medical Journal*, *57*(3), 298–301. https://doi.org/10.3325/cmj.2016.57.298

Hope, T., Chan, J., Kittur, A., & Shahaf, D. (2017). Accelerating Innovation Through Analogy Mining. *ArXiv:1706.05585 [Cs, Stat]*.

Hua, M.-Y., Harvey, S., & Rietzschel, E. F. (2022). Unpacking “Ideas” in Creative Work: A Multidisciplinary Review. *Academy of Management Annals*.

Jarvenpaa, S. L., & Standaert, W. (2018). Digital Probes as Opening Possibilities of Generativity. *Journal of the Association for Information Systems*, *19*(10), 982–1000.

Jeppesen, L. B., & Lakhani, K. R. (2010). Marginality and Problem-Solving Effectiveness in Broadcast Search. *Organization Science*, *21*(5), 1016–1033.

McKinney, E. H., & Yoos, C. J. (2010). Information About Information: A Taxonomy of Views. *MIS Quarterly*, *34*(2), 329–344. https://doi.org/10.2307/20721430

Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization Science*, *5*, 14–37.

Pariser, E. (2014). *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin Books.

Portenoy, J., Radensky, M., West, J., Horvitz, E., Weld, D., & Hope, T. (2022). Bursting Scientific Filter Bubbles: Boosting Innovation via Novel Author Discovery.

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *ArXiv:1908.10084 [Cs]*.

Storm, B. C., Stone, S. M., & Benjamin, A. S. (2017). Using the Internet to access information F

Taramigkou, M., Apostolou, D., & Mentzas, G. (2017). Supporting Creativity through the Interactive Exploratory Search Paradigm. *International Journal of Human-Computer Interaction*, *33*(2), 94–114.

Wang, K., & Nickerson, J. V. (2017). A literature review on individual creativity support systems. *Computers in Human Behavior*, *74*, 139–151.

Youn, H., Strumsky, D., Bettencourt, L. M. A., & Lobo, J. (2015). Invention as a combinatorial process: Evidence from US patents. *Journal of the Royal Society Interface*, *12*(106).