Transformations issue 34 (2020)
www.transformationsjournal.org

ISSN 1444-3775

Faster than the Speed of Thought: virtual assistants, search and the logic of pre-emption

Paul Atkinson & Richie Barker

AUTHOR BIO

Dr Paul Atkinson teaches within the Academic and Professional Writing program at Monash University, Australia. He has published widely on a range of media, from cinema and animation to dance and comic books, in addition to his work on philosophy, technology and aesthetics. In general, his work examines how time operates in different media including new media environments.

Dr Richie Barker is a researcher and lecturer in communication with School of Communication and Creative Arts at Deakin University. His research considers contemporary promotional and search cultures with a focus on creative practice.

ABSTRACT

The rapid development of predictive technologies from simple pre-emptive text to voice-activated virtual assistants raises questions about how we engage with bodies of knowledge mediated by algorithms. Predictive technologies with increasingly adaptive algorithms supported by machine learning, have the capacity to learn alongside us, gleaning information to better understand behavioural patterns and predict human action and intention. These technologies are often promoted in terms of how they assist human users and are evaluated in terms of their speed and relevance. This valorisation of speed is underpinned by an algorithmic means-end logic that is not subject to the durational constraints of human perception and attention. Indeed, the inhuman time of an algorithm has to be adjusted to fit the lived time of human thought and action. Drawing on the work of Henri Bergson and Bernard Stiegler among others, this paper argues the quest for speed in the development of search technologies constructs a future in which time is reduced to discrete possibilities and disregards the lived delay immanent to human thought.

KEYWORDS

pre-emption, internet search, voice assistants, predictive technologies, time, thought, cognition

The rapid development of predictive technologies from simple pre-emptive text to voice-activated virtual assistants raises questions about how we engage with bodies of knowledge mediated by algorithms. Predictive technologies with increasingly adaptive algorithms supported by machine learning can learn alongside us, gleaning information to better understand behavioural patterns and predict human action and intention. These technologies are often promoted in terms of how they assist human users and are evaluated in terms of their speed and relevance. Large technological firms, most notably Google, claim that the faster a decision is made or action is taken, the better these devices can serve the user. Why wait for the retrieval of information when it can be ready to hand, and why remember everyday events and practices, when a device can do it with greater reliability? According to this logic, the speed of computational decision-making not only can but should usurp slow, indecisive and fallible human decision-making.

We argue that this valorisation of speed is underpinned by an algorithmic means-end logic that is not subject to the durational constraints of human perception and attention. Indeed, the inhuman time of an algorithm has to be adjusted to fit the lived time of human thought and action, which raises issues concerning when and how information is interpolated in human experience. To be presented with an answer even before a question has been properly articulated might free up our time to address other matters or allow us to devote more time to a complex idea, however it can also break the tension immanent to thinking. Remaining within the time of an idea is actually hindered by the provision of a solution. Moreover, the facility by which search technologies are integrated into cognitive activities can also diminish awareness of the relationship between knowledge, reasoning and the formation of questions.

Drawing on the work of Henri Bergson and Bernard Stiegler among others, this paper argues that the quest for speed in the development of search technologies constructs a future in which time is reduced to discrete possibilities and disregards the lived delay immanent to human thought. The article begins by identifying speed as a primary algorithmic logic inherent in search and related applications and then explores the capacity of these applications, and their inhuman speed, to be interpolated into human thought and lived time.

Search, Algorithmic Speed and Voice Assistants

When we directly consult a search engine by typing in a query or formulating a question, we are to some degree aware of what we are searching for and why we have used the search engine. The best search engine will find relevant results that directly align with our expectations in a timely fashion and are not delayed by unnecessary buffering. The speed by which the search algorithm returns a result complements the drive to know or understand, for the quicker the response, the more the search engine approximates the retrieval of our own ideas, thoughts and memories. In this description are two types of speed, the first is based on relevance, where finding what we want in an initial request removes the need for subsequent requests, and the second refers to the

application's processing speed. Both reduce waiting and enhance the user's trust in the search mechanism and its algorithms. The technology companies do not want us to wait either, for it allows the mind to wander and could encourage the users to look elsewhere. Amazon undertook a study of waiting times on their website, particularly in light of the fact that users of digital technologies can become easily distracted. They "found that, on average, for every tenth of a second that its customers are made to wait while using its website, Amazon would lose 1 percent of revenue" (Farman 13-14). Wait times are also an issue for Google, which tested a page producing 30 search results rather than the usual 10, in which a slight delay of 0.5 of a second reduced the traffic through the search results page (Farman 14). The importance of direct speed is further illustrated in every Google desktop search result page, which lists the number of seconds the search engine took to process the results. This focus on speed has been noted by Hillis et al. who describe instantaneity as a primary principle in Google's model of relevance. Speed allows us to remain focused on a myriad of tasks ranging from online shopping to work-related information gathering by removing any distractions, temporal or spatial.

The matter of search relevance is managed by an application's algorithms. In order to produce a relevant result, the algorithm must have access to accurate information derived from a variety of sources, from websites through to online networked databases and sensors. Essentially, the quality of the output is dependent on the quality of the input. Some data is easily verified, such as time, location, and weather reports. However, most information is not derived through existing quantified data and has to be assessed through other means. The principle most commonly used to judge relevance by search engines and also by many other social media applications is popularity, where more popular sites rank higher in the list of outputs than less popular sites. A key advance of the Google PageRank system was its capacity to assess popularity in terms of the number of quality links directed to the site from reputable sources. It is a form of consensus objectivity in which popularity is moderated by reputation. Tarleton Gillespie states that the PageRank algorithm has often been cited in support of the objectivity of Google Search but notes many other factors that determine relevance that are not so readily discussed (180):

> Search algorithms, for example, once based on simply tallying how often the actual search terms appear in the indexed web pages, now incorporate contextual information about the sites and their hosts, consider how often the site is linked to by others and in what ways, and enlist natural language processing techniques to better "understand" both the query and the resources that the algorithm might return in response. (175)

These factors do not directly make the search results accurate or objective, for relevance is directly appraised by the user. It describes what is relevant to the particular individual who made the query, rather than some broad notion of relevance applicable to all users. Personalisation in Google Search marks a shift towards judging a user's intention. Pariser notes that in 2007 Google introduced an algorithm that analysed 57 personalisation factors to help it predict the search intent of each user with sufficient complexity that a user could receive different search results for the same query (1-2). Some of the

personalisation factors are obvious to most users, for example the use of an IP address to determine a user's location, while others are much less visible, such as the tracking of online behaviours. Given the lack of transparency, a complete understanding of personalisation factors remains elusive. Somewhat clearer are questions of bias they present, which, Van Couvering argues, arise from the complex interweaving of engineers' individual practices with larger industrial and cultural frameworks.

Despite its greater functionality and popularity, Google Search has increasingly receded into the background of everyday use, which is reflected in the names of two now decommissioned Google Search functions, Google Suggest and Google Assistant. Both descriptors suggest a relationship with technologies in which the application is not something you directly address but instead assists in other activities and pre-empts future needs. The names of two other retired applications, Google Instant and Google Now suggest different ways of receding into the background, in which the application disappears into the instantaneity of the user's present action. Google Suggest was launched in 2004 and autonomously completes (pre-empts) search terms in the query box as the user types. Possible search results were presented alongside the incomplete query, allowing the user to choose one of them rather than continue typing. It aimed to speed up the search process by removing the need to type out a full phrase, which was acknowledged in a Google blog post with the aside: "let's face it we're all a little lazy." The same blogpost also claimed that the list allows the searcher to "explore" and "learn" (I've got a suggestion), which is an interesting contrast. Pre-emption is associated both with a reduction of labour time – the user does not need to type nor to think any more about the query – as well as exploration. Exploration here is afforded by the visibility of the search possibilities, albeit an exploration that is limited to predetermined choices delimited by personalisation and popularity algorithms.

Google Instant, which was launched in 2010, is very similar, except that the suggested search results also appear in the list of searches – the user can actually see the full links, titles, and sometimes a brief description of a particular site. The changes further sped up the provision of results with Google arguing that their "testing has shown that Google Instant saves the average searcher two to five seconds per search" ("Search: now faster than the speed of type"), however, they claimed that this would still allow for a high degree of creativity. A 2011 Google blog post describes the benefits of Google Instant as follows: "Typing [c] will give you predictions for [chicken] or [cake] versus [craigslist] or [cnn], and typing [co] will predict [cookies] or [coconut] – and maybe inspire you to make coconut cookies" ("This week in search"). What Google does not talk about based on this same principle is that speed reduces the capacity to choose. To choose quickly precludes exploration or deliberation on the relationship between different types of knowledge. Indeed, questions of speed underpinned the retirement of Google Instant in 2017, which was better suited to the speed and size of desktops after much of the search activity had shifted to mobile devices. The official blog states "we have decided to remove Google Instant, so we can focus on ways to make Search even faster and more fluid on all devices" (Google qtd. in Schwartz). The claimed creative act of bringing together ostensibly dissimilar concepts is trumped by the speed and facility of mobile use. We have to question any claim

that search algorithms support creativity, either through the creation of serendipitous concepts or freeing up labour, because speed limits possibility and popularity drives the algorithms (Barker and Atkinson). The claims about the creativity of Google Search in producing unpredictable results only serve to disguise its instrumentality. The question is not whether we will see cookies and coconut in a pre-empted list but whether we would actually expend the effort to make the connection between the terms, never mind bake the cookies. It is worth noting that the examples of unexpected associations presented by Google may appear banal and inconsequential, which accords with their general marketing strategy and its emphasis on playfulness: a strategy that disguises the company's ever increasing role in data management. The focus on unexpected search results also speaks to Google's desire to integrate search into quotidian cognitive activities. This is evident, for example, in an industrial context with advertising creatives being encouraged to introduce associations developed on their behalf in their ideation process. Statements in the *Think With Google* website promoting the company's "creative insights" initiative, such as "advertising can be more human when it is more data driven" ("Why creativity is being liberated by data and machine learning"), suggest the growing prevalence of a particular algorithmic logic.

The valorisation of speed also underscores the integration of search into other applications. A key problem with Google Search is that it remains embedded in an interface where queries are either typed or spoken and results appear in a list. Google Now was launched in 2015 as a means of extending the pre-emptive capacity of Google well beyond search. The application provided real-time information to the user by predicting what they might require or request. Predictions are derived from the user's movements, stored personal information, as well as Google's suite of algorithms that assess and utilise Internet data. The promotion for the application states that:

> Google Now is about giving you just the right information at just the right time. It can show you the day's weather as you get dressed in the morning, or alert you that there's heavy traffic between you and your butterfly-inducing date – so you'd better leave now! It can also share news updates on a story you've been following, remind you to leave for the airport so you can make your flight and much more. There's no digging required: cards appear at the moment you need them most – and the more you use Google Now, the more you get out of it. (Google Now)

As with the text pre-emption of Google Instant and Google Suggest, it is assumed that the process of making a request effectively increases the time it takes to receive an answer and, therefore, the algorithms provide information before a request has even been made. It is still a matter of increased speed through relevance, except that the application does not wait for the slow process of formulating a definite conscious intention, or the even slower process of stopping an activity to make a search request. The user might be thinking of road traffic conditions when she receives a notification, but the notification appears irrespective of whether she actually formulates a query or request.

Regardless of how the information is utilised, Google Now's predictive algorithms operate at a speed faster than the situated, analogical, contextual and conscious formulation of a human need or intention. They provide information to the user "now" although the predictions are based on long term tracking of an individual's action, which means the information technically could have been provided earlier. For example, the application knows that the individual catches a train every Monday morning, and therefore provides information on train timetables the day before. The provision of the information is only delayed to segue with the time of human action, for only information that coincides with a user's intention is judged relevant. Timeliness is a central aspect of speed. Users often claim that the most popular search engines, Google, in particular, can "read their minds" or at least their search intention (Goldman 196). For a mind to be read, an idea has to be present to the mind of the user, occur approximately in the same "now," as when the search engine provides its results. However, the time of technological prediction depends on the capacity for the various algorithms to assimilate relevant data and the processing speed of the computer or mobile device, and not the time of human intention. Accordingly, applications such as Google Now should be able to project further into the future in a way that is not limited to the time of human thought, attention or intention. From this perspective, the development of Google Now forms part of "a general project to perfect relevance such that search will eventually predict a user's intention without a clumsy interface and almost before the user knows what they want" (Hillis et al. 56). The founders of Google, Brin and Page, envisioned that Search will eventually become a form of augmentation of the human mind (Hillis et al. 55), and there was even talk within Google of an "implant" for search: "Eventually you'll have the implant, where if you think about a fact, it will just tell you the answer" (Page qtd. in Adams). Technologies such as Google Now are the first steps in this direction, and they can only properly augment the mind if they work at or above the speed of human intention.

Even with Google Now, speed is limited by interfaces that separate the user from their activity. With desktops and laptops, the ability to attend to the screen and the world at the same time is very limited. It is also somewhat limited with mobile devices, as is evinced in the awkwardness with which users move in busy urban environments while scrolling through mobile data. The increasing interest in voice interfaces in mobile devices and smart speakers forms part of a general push to limit or remove the interface, which is one reason why Google Now morphed into Google Assistant. Google Assistant, described as "your own personal Google," can be accessed with the words "Ok Google" ("Meet your Google Assistant"). This device does not completely remove the interface as would an implant, however, it becomes less visible in mimicking human speech and inserting itself into conversational context. The user can continue to act in the world through sight and touch while listening or issuing commands, and in this respect is analogous to radio, a medium favoured by people driving, cycling, walking, exercising or even doing chores around the house. Voice assistants can recede into the background of human activity while fulfilling the role of a virtual butler or personal assistant – a role that Alex Pentland envisaged as far back as 2000 for smart clothes and rooms (37). They will regularly assist the user while disguised within quotidian activity. With regard to the other popular voice assistant, IT

developers used the term "frictionless" to describe how Amazon's Alexa should be assimilated into human contexts (Shulevitz). The device remains invisible within a conversation as long as the information is correct or relevant. We only tend to notice applications that disrupt or interfere with expectations.

This frictionless algorithmic activity has implications for human understanding. It can shape how we think by lessening awareness and diminishing conscious control. Alexander Halavais claims that search is heading in two main directions, the "conversational agent" and the "predictive mind-reader" (77). Conversational agents ensure that search is more responsive and integrated with everyday practices whereas predictive mind-readers present search results without a direct request from the user. Both will "curate" results based on the system's understanding of the user, in which case individuals will become less involved in the search process. Search based systems also provide epistemological guidance, informing the user of what they should or could know. This is demonstrated in the move towards "one-shot answers," where voice assistants such as Google Assistant or Alexa provide a single answer to a question unlike the array of possible results in textual pre-emption. It is much quicker for a voice assistant such as Alexa to provide a single answer than laboriously list each possible search result. Writing in *Wired Magazine*, James Vlahos argues that this move to one-shot answers and the "conversational web" is being promoted in terms of its "increased convenience and efficiency." However, it is important to ask whether the removal of "laborious" search actions limits curiosity. If information is always *on tap*, why seek new information, or to continue the metaphor: "Why pump water from a well if it pours effortlessly from your faucet?" (Vlahos). In short, the more facile our engagement with various search technologies, the more we will lose the ability to properly understand the context of the information, and the more likely that it is that we will come to accept the information provided by these digital assistants.

Human Thought, Lived time and Inhuman Speed

Search technologies can pre-empt our thought, synthesise information and even engage in dialogue and with the further development of the conversational web, devices will increasingly resemble human participants in discourse. The most common criticisms of expanded search technologies are political and ethical, focusing on who controls knowledge and how this control conditions behaviour, however, it is also important to question increased speed and cognitive labour saving. Is there some value in slowing down thought and reintroducing friction and uncertainty? Indeed, is the time of thought something that should be fostered or developed? Of course, technology brings many benefits, by increasing the breadth of available information and by removing many of the most mundane cognitive tasks. Andy Clark has demonstrated how technologies expand our cognitive range by augmenting our memory or even providing an external platform for reasoning and thinking, practices that occur outside the body as well as in the mind (Clark "Intrinsic"; Clark *Supersizing*; Clark and Chalmers). Calculators, computers, books, and digital devices augment cognition and also remove much unnecessary mental labour. Katherine Hayles has argued that computational

technologies complement changes in a broader information environment, for example, we have developed new ways of reading, such as "hyper reading," to assimilate a greater amount of information over a shorter time. This type of reading is linked to "machine reading," in which various computational applications are able to read and assimilate data on our behalf (*How We Think* 11-12). Both have become increasingly necessary due to the amount of information available and to the widening gap between what we can read in our lifetime and what is available (27). By assimilating this information at a speed unachievable by any person, machine reading, including search technologies, will increase the breadth of our thought and provide a much larger context for any intellectual judgement. Although many aspects of outsourcing cognition to applications and devices should be celebrated, we should question any universal or automatic adoption of principles of labour and time-saving.

Outsourcing familiar cognitive tasks to our devices, such as simple mathematical operations on a calculator, differs from the use of black box algorithms to supplement cognitive activity. A device or program that performs a familiar cognitive process still operates within a conceivable time of thinking albeit at a much faster rate. In contrast, truly pre-emptive applications, whose operations remain invisible to the user, or even the programmer in self-learning systems, only connect to human thought at the level of inputs (the query) and outputs (search results), with a corresponding emphasis on speed and a decrease in friction. Faster is better when it comes to the technical provision of information but it might not have the same value when placed in the context and time of human thinking. Lorenzo Simpson, drawing on critiques of instrumental reason, argues that this means-end logic reduces human experience to a utilitarian abstraction and the prediction of results, in which the system "is assessed in terms of the efficiency and effectiveness by which it achieves its result" (48). This assessment highlights results or outputs because they are objective and measurable, unlike the less tangible notions of meaning associated with human action (48). In this foregrounding of the relationship between inputs and outputs, the technical system decontextualises human action from its social and historical context and particular practices of meaning-making. Simpson gives the example of eating among family and friends, which cannot be accessed in terms of definite measurable ends, such as the speed with which food can be produced, ordered or delivered. The focus on measurable ends does not properly account for the different ways to meet the same end (Simpson 45) or encourage interest in varying the means. If we apply this logic to academic research, articles will only be judged according to their capacity to meet impact and citation metrics rather than in terms of the particular social value and cognitive labour associated with research and writing. As has already been noted, the idea that Google Instant creates serendipitous connections through displaying multiple search results (an emphasis on means), soon gave way to speed and relevance (an emphasis on ends). Undoubtedly users want a fast return on a query, but in the constant focus on ends, they are not really given a choice of other ways of engaging with Search. Search results could be presented in ways that allow the user to see the pathways by which a search engine derives a result and, therefore, to imagine how their thought aligns with or diverges from its epistemological principles. For example, the search result could appear in a hierarchical tree

diagram, similar to a taxonomic tree in biology, which is organised according to either disciplinary or popular knowledge. Another option would to place clusters of research results in a word cloud, which would allow the user to easily attend to the less common results operating at the periphery.

Machine thinking, or to use Hayles' example of machine reading, are quite unlike human thinking despite the similarity of the outputs. The algorithms might engage with common structures of human thought, as in natural language processing, however, they do not engage with the specific time of human cognition. At this point, it is important to distinguish between lived time, in which the past in the form of memory remains continuous with the present, and time as a background to representation. Lived time, including the lived time of our thought, describes an orientation towards the future grounded in the continuity of the past. The philosopher Martin Heidegger talks about how human time is constituted as a "spirit" or "disposition" that can be revealed through consciousness. Time manifests as *Dasein* (being) and the various modes of being in the world including "performing, effecting and completing" actions as well as questioning and "contemplation" (6-7). Heidegger draws upon Edmund Husserl's argument that lived time must appear over a definite duration and requires at its lowest level "primary memory" to ensure sensual coherence. For example, when we hear a musical tone or melody, which necessarily endures, we must have some memory of the beginning of the note or the sequence of notes for it to have any meaning. Additionally, we anticipate the future of both the note and melody in a fusion of the just past and the immediate future (37). In the continuity of any present moment, primary memory describes how the past is held over (retention) and the present extended towards the future (protention), and it is upon this that that secondary memory, representation and recollection form (37-38). Lived time as an extension of the past into the present and a corresponding orientation towards the future has been postulated by a number of philosophers. Henri Bergson's theory of duration – time as continuous, interconnected and enduring, which is most evident in conscious experience – and William James's conception of the specious present – in which even the simplest perception endures over a definite interval. For all these so called process philosophers, thinking cannot be reduced to the representations of thought or logical connections or sequences, because this does not sufficiently attend to disposition and the retention of the past. For example, when we formulate a statement in a natural language, the past of the sentence is present in each word and pushes us towards future words and ideas. The idea gestures towards this future even though the complete thought is not yet given.

The logic of the algorithm which draws upon discrete representations, procedural logic, and inputs and outputs, functions differently to the intentionality of human thought. Bernard Stiegler decries the integration of science and technics in which communication has been usurped by technical rationality. With reference to Jürgen Habermas, he argues that technical efficiency supported by automation and cybernetics replaces a human notion of purpose and the "critical distance" implicit to rational action (Stiegler, *Technics* 12). Distance here refers to the capacity to make a rational judgement, which could be represented propositionally, however, most importantly, it also implies the slowness of thought – thought must have a definite time. In *Taking*

Care of Youth and the Generations, Stiegler states that reason must also be considered through the time of attention in addition to any logical structures. He argues that psychotechnologies, from writing and reading through to techniques of memorisation, actually separate mental function from bodily function by the training of the body to sit for long periods of time and removing distraction (65). In an ongoing "struggle" against "laziness," the time of reasoning and writing, increases concentration and creates "long-circuits" of attention (60) that extend the mind beyond the present (*Generations* 67). This human process of attending to the future differs from technical pre-emption which removes the labour of thought by breaking the lived continuity of time. In pre-empting or prejudging our thought, search technologies effectively shorten the circuits of attention because the end is effectively given before the means are worked through. The faster the process and the more thorough the pre-emption, the less likely it is that an individual will develop an ability to remain within a thought and its requisite mental labour. There may be greater efficiency in the production of knowledge, however, this should not be conflated with thought.

Stiegler develops his argument by opposing new technologies of attention with more traditional technologies of attention, such as the book. This traditional vs. new media opposition does not have to be maintained, for the crucial issue is the development of attention through relatively slow processes of secondary retention rather than a question of technology. It is notable that he refers to writing and not just the book, for in addition to creating a connection with earlier generations, writing serves as a means of individuation (*Generations* 75). Writing projects consciousness forward by expanding attention beyond the immediate needs of the present as a process of thinking through and pre-empting (84). When we memorise disciplinary specific knowledge, for example tables, mathematical equations, grammars, scales, letters and words, we engage with the past of a discipline and, crucially, develop the means to anticipate new knowledge (105). Writing in this context is a means of expanding our "horizons of anticipation," which is not limited to an object of attention (*Generations* 106). New technologies that encourage this disciplined engagement with new knowledge could also expand the circuits of attention, and could apply to a wide range of activities from writing blogs to coding. If search mechanisms complement this type of activity, they could contribute to attentional development. However, if we allow search technologies to usurp human forms of pre-emption, which are cultivated through mnemonic practices, we could also relinquish our capacity to expand our attentional present in thought and writing.

Crucially, from Stiegler's perspective, the waiting immanent to attention is constitutive of rational thought. Attention is often restricted to perceptual attention, however it also can refer to an attention to thought, which is necessarily extended over a definite duration. Stiegler argues that attention requires waiting, for to attend to something is not to immediately receive it or to be immediately satisfied, but rather to open oneself up to its singularity and futurity (*Generations* 96). Stiegler characterises attention in terms of anticipation or desire, in which the subject is attentive to ideas and objects that are not yet present to consciousness. This derives from the relationship between protention and retention. The protention of thought is first grounded in

primary retentions that sit at the threshold of consciousness, but through continual development, the individual is able to develop longer circuits of attention from secondary retentions, which are further coordinated by tertiary retentions (cultural memory) (101). Any process that seeks to automate this process through the introduction of already received tertiary retentions, short circuits the futurity of attention. For example, the augmentation of attentional tasks by various computational processes replaces these circuits of attention and instead configures attention as a preparedness to act (102). The individual responds directly to stimuli and information that directly captures attention (television, pre-emptive search, voice assistants, etc.) without the slow process of consciously working through and within retention and protention. In a technical model of efficiency, greater value is attributed to fast responses.

It is no coincidence that much of the recent discussion of machine-human interaction foregrounds nonconscious cognition, for this type of cognition can be most readily assessed in terms of outputs and responsiveness. Hayles states that nonconscious thought, in particular nonconscious decision-making, serves as a bridge between human and technical cognition because it operates at a speed much faster than that of conscious awareness (*Unthought* 28). Clark talks about in-built predictive models in perception and how they can be related to what is called predictive coding (*Surfing* 26), in which essentially nonconscious top-down processes make prejudgements. Again, the issue of speed is crucial here, for the prejudgements must occur in a time below the threshold of conscious awareness. There has been a corresponding interest in the work psychologists and neuroscientists such as Benjamin Libet, who argues that it takes half a second before we become aware of sensations and even many of our decisions (33), and Daniel Kahneman, who highlights the value of what he calls System 1 thinking that responds quickly and without voluntary control (20). From an alternative perspective, the media theorist Mark Andrejevic criticises this emphasis on nonconscious aspects of cognition, in particular, forms of decision-making invested in the body because they tend to characterise thought as "thoughtlessness" and therefore appeal to such disciplines as neuromarketing (103). Undoubtedly, much of our thinking occurs below the threshold of consciousness, however, focusing too much on this aspect of cognition does not open up a critical space between algorithmic decision-making and the time immanent to human decision-making.

Examining pre-emptive search in cognition should also involve an evaluation of how conscious attention is interpolated by technological time. Technology – in particular cybernetic and self-managing systems – can liberate us from time's demands, but by so doing, creates a tension between an open, unpredictable future and a closed future characterised by prediction and control (Simpson 53-54). The relationship between technological liberation and control depends on how technical prediction is integrated into human action and thought. Hayles argues that computational systems are increasingly able to make decisions that intersect directly with our own cognitive practices, which she broadly refers to "cognitive assemblages" (*Unthought* 115). These cognitive or human-machine assemblages are sustained by particular flows of information that with widespread and continued use can lead to changes in cognition (119). She gives the example of a traffic management system in Los Angeles, which predicts the movement of traffic and continually makes

decisions to improve the flow, which in turn aids drivers in making their own decisions (*Unthought* 123). The driver is partially liberated from the need to attend to the traffic and make decisions on what route to take, which in turn frees up time for thought. Likewise, navigation aids, which choose a route, indicate where to find groceries for a meal, etc., reduce the user's need to attend to navigational decision-making. However, Hayles also notes that this symbiotic relationship will in turn affect cognition, for the user will not need to be fully aware of their location and navigational practices, which will eventually lead to a reduction in those synaptic networks associated with navigation. The same principle applies to any long term relationship with any digital device that replaces cognitive activity (*Unthought* 125). In itself, the ability to navigate may not be greatly important and it might well be beneficial to reallocate cognitive resources to a device, or a system. However, we should not focus just on the skill and its accompanying synaptic network, for learning to navigate also involves developing the capacity to attend to the futurity of autonomous movement. Attention can extend beyond the immediately visible to a much more open future that incorporates space in a temporal horizon. Automated navigational decision-making reduces the need to attend to the immediate future of location and also closes off a type of extended intentional thinking.

What needs to be addressed is the incommensurability of technical prediction and the lived plenitude of thought and intention. Henri Bergson argues that science can only predict the future, create an image of what is possible, because it reduces the past to a collection of discrete representations, objects and events. The future is as an abstraction created through discarding many aspects of the lived world (20). To describe this process, he distinguishes between two senses of the word *possible*: the "ideally pre-existent," an ideal projection of an event within systems of knowledge before it is made actual, and the "not-impossible" (120). The possible, in its positive sense as the ideally pre-existent, only applies to a closed system in which all elements are given and in which the future is little more than a rearrangement of these elements (122). In contrast, the not-impossible determines the future by removing any impediments to the realisation of an event. Unlike a positive image of the future, the not-impossible predicts through precluding events, which Bergson describes through the metaphor of a gate: "If you close the gate you know no one will cross the road; it does not follow that you can predict who will cross the road when you open it" (120-21). The metaphor can be extended by stating that the gate will also determine that someone or something must cross the road at a certain point. Bergson argues that we often imagine that we are creating a positive image of the future when in fact we are imposing the not-impossible (120-21). The problem with using such models to understand freedom and futurity is that they frame the future in terms of "competition between possibles" (122-23) rather than in terms of a true openness to the future. Our attention to the future is reduced to what can be clearly articulated within the present, in other words, what can be articulated as an output. For example, for most people living in a contemporary democracy, political action is limited to the act of choosing between representatives. Each individual certainly has a choice, however, it is a choice that precludes many other forms of political action or even a proper consideration of what it means to be political. What cannot be clearly articulated as a possible event, thought, or

action recedes from the horizon of our futurity in a shift towards a closed system.

The speed of pre-emption and personalisation in search operate within the logic of the not-impossible and ideally pre-existent to ostensibly liberate the user from mundane cognitive tasks. The cognitive value of these technologies depends on how they are inserted into the open system of human thought and intentionality. By rapidly providing results that appear to match our lived intention, we do not have sufficient time to consider what we want. If a solution is proffered quickly, we are more likely to accept that solution even though an idea could be much more open and variable. Search shapes our judgement through imposing a temporally-based agenda of what is, could be, or should be possible. With personalisation, the search results are derived in part from the individual's own actions, an attempt to create a closed system from human intention and action. Kylie Jarrett argues that Google's reading of intentions is essentially an intentional fallacy because motivations cannot truly be read from the data (21), which complements Bergson's and Stiegler's arguments that the technologies do not work within the time of intention. In addition to possibly misreading the intention, the search suggestions shape what the searcher is looking for and therefore operate as a feedback loop that confirms the system's predictions: "In this recursive logic, the potential of futurity becomes limited by past resolutions. Embedding the output of search into the logics of search in this way can be understood as a form of control" (Jarrett 24). This recursive process confirms the value of the algorithm in predicting the user's intentions, and in doing so creates a technological normative (24). In other words, the tertiary retentions of the algorithm usurp the secondary retentions of human thought. Because we are not entirely aware of how the results are obtained, we are more ready to accept them as objective. It is only when the results do not meet the user's expectations that there is cause for speculation on the process. König and Rasch state that we have become too familiar with search mechanisms, especially now that they are integrated into a range of devices, that they have become invisible through their ubiquity (10-11). In the evolution of Google Search, the increase in the facility and integration of the technologies into daily use, in particular the speed of responsiveness, means that the user cannot develop a sufficiently critical or temporal distance – it is difficult to operate within the critical time of attention. This is a feature of what Pariser and others have described as the "filter bubble" in which access to alternative perspectives is precluded and the status quo is reinforced.

The more we speed up the provision of information, including the pre-emption of search queries or human behaviour, a feature of Google Now and its current manifestation Google Assistant, the less we can operate within the proper time of attention. It is difficult to remain within the thought when a solution is readily proffered. It becomes a given, a normative truth that contracts the duration of attention, rather than something that can be contested or properly examined. Adrian Ward states that the integration of the Internet into our devices and consequently into quotidian behaviour blurs the distinction between external knowledge and our memory. The Internet incorporated into our devices "provides information quickly, virtually invisibly, and without any of the extraneous physical cues inherent in human-

to-human interactions" (344) The less it intrudes into our lives, the greater the confusion as to whether information is stored in human memory or accessed through the Internet (344). The confusion arises because the speed of access to information on the Internet is often quicker than trying to remember something. For example, psychological studies on the use of external memory on Internet enabled devices have shown that individuals share the process of memorisation with their devices (Sparrow et al. 777). One way of reinstating the distinction between human memory and computer memory is by slowing down access to information (Ward 344). This is not just a question of assessing the epistemological status of our beliefs and memories, but reflecting on how and when information is revealed in the particular time of our attention.

Conclusion

Pre-emptive algorithms incorporated into voice assistants and mobile applications increasingly deliver information that keeps time with human action and human thought. Search companies argue that increases in speed serve the user's interest by providing timely results and information, and allow the technologies to seamlessly and frictionlessly embed themselves within our activity. This increased speed is advantageous if the time of waiting is primarily imagined from the point of view of a user who has clearly articulated a question and does not want to be impeded by buffering or any other form of technological delay. However, the push towards greater speeds, assisted by personalisation and the removal of a definite interface, will start to interfere with the specific time of human intention and attention. Futurity will be tethered to a technological time in which attention is configured in terms of possibility, which can interfere with the continuity of thought and an immanent epistemological drive. Search technologies present images of what can be known through the listing of search results or the direct response to questions, and in doing so also place limits on our action. The act of knowing is reduced to the discrete set of possibilities, in which search results that do not appear in the first listing are largely ignored. This is further accentuated with the push towards one-shot answers in voice assistants, where all other possibilities are effectively effaced by the presentation of a single answer. If the answer is provided before the proper articulation of a question – the mind-reading that Google's founders hope for – then the open time of the present will be reduced to the future perfect, in which we will always have known something irrespective of the protention of a lived present. Speed could introduce a facile timeliness that colonises the time of attention while taming its uncertainty and thereby efface those ideas that flicker on the fringe of consciousness. Of course, knowledge production and dissemination are not limited to the use of personal assistants, voice assistants, wearable technology and search. This particular way of stripping back knowledge in the service of efficiency has to be looked at in a broader technological ecology. Other forms of information retrieval may simultaneously work towards the expansion of knowledge thus counteracting the inhuman speed inherent in contemporary search-related applications. However, as this article suggests, to better understand the relationship between technology and cognition, we need to highlight the algorithmic logic inherent to specific technologies.

Works Cited

Adams, Tim. "Google and the Future of Search: Amit Singhal and the Knowledge Graph." *The Guardian*, 20 Jan. 2013. http://www.theguardian.com/technology/2013/jan/19/google-search-knowledge-graph-singhal-interview.

Andrejevic, Mark. *Infoglut: How Too Much Information is Changing the Way We Think and Know*. London and New York: Routledge, 2013.

Barker, Richie and Paul Atkinson. "A New Master of the Cultural Domain? Google's Role in the Transactive Memory Systems of Advertising Copywriters and Art Directors." *Media International Australia* 172.1 (June 2019): 61-73.

Bergson, Henri. *The Creative Mind*. Trans. M. L. Andison. New York: Philosophical Library, 1946.

Clark, Andy. *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford: Oxford University Press, 2015.

---. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford Scholarship Online, 2009.

---. "Intrinsic Content, Active Memory and the Extended Mind." *Analysis* 65.1 (2005): 1-11.

Clark, Andy and David Chalmers. "The Extended Mind." *Analysis* 58.1 (1998): 7-19.

Farman, Jason. *Delayed: The Art of Waiting from the Ancient to the Instant World*. New Haven and London: Yale UP, 2018.

Gillespie, Tarleton. "The Relevance of Algorithms." *Media Technologies: Essays on Communication, Materiality, and Society*. Ed. Tarleton Gillespie, Kirsten A Foot and Pablo J Boczkowski. Cambridge: MIT Press, 2014. 167-93.

Goldman, Daniel. "Search Engine Bias and the Demise of Search Engine Utopianism." *Santa Clara Law Digital Commons* (2006): 187–200. https://digitalcommons.law.scu.edu/cgi/viewcontent.cgi?article=1112&context=facpubs.

"Google Now. The Right Information at Just the Right Time." *Google*. 1 July 2019. https://www.google.com/intl/en-GB/landing/now/.

Halavais, Alexander. *Search Engine Society*. 2nd ed. Cambridge and Malden: Polity Press, 2017.

Hayles, N. Katherine. *Unthought: The Power of the Cognitive Nonconscious*. Chicago: University of Chicago Press, 2017.

---. *How We Think: Digital Media and Contemporary Technogenesis*. Chicago: University of Chicago Press, 2012.

Heidegger, Martin. *The Concept of Time*. Trans. William McNeill. Oxford: Blackwell, 1992.

Hillis, Ken, Michael Petit and Kylie Jarrett. *Google and the Culture of Search*. New York and London: Routledge, 2013.

Husserl, Edmund. *On the Phenomenology of the Consciousness of Internal Time (1893-1917)*. Trans. John Barnett Brough. Ed. Rudolf Bernet. Dordrecht: Kluwer Academic Publishers, 1991.

"I've Got a Suggestion." *Google Official Blog*. 10 December 2004. https://googleblog.blogspot.com/2004/12/ive-got-suggestion.html.

Jarrett Kylie. "A Database of Intention?" Ed. René König and Miriam Rasch. *Society of the Query Reader: Reflections on Web Search*. INC Reader #9. Amsterdam: Institute of Network Cultures, 2014. 16-29.

Kahneman, Daniel. *Thinking, Fast and Slow*. London: Penguin, 2012.

König, René and Miriam Rasch. "Reflect and Act!: Introduction to the Society of the Query Reader." Ed. René König and Miriam Rasch. *Society of the Query Reader: Reflections on Web Search*. INC Reader #9. Amsterdam: Institute of Network Cultures, 2014. 10-15.

Libet, Benjamin. *Mind Time: The Temporal Factor in Consciousness*. Cambridge, Mass.: Harvard UP, 2004.

"Meet your Google Assistant." *Google*. https://assistant.google.com. 23 March 2019.

Pariser, Eli. *The Filter Bubble: What the Internet Is Hiding from You*. London: Penguin Press, 2011.

Pentland, Alex. "Perceptual Intelligence." *Communications of the ACM* 43.3 (March 2000): 35-44.

Schwartz, Barry. "Google has dropped Google Instant Search." *Search Engine Land*. 26 July 2017. https://searchengineland.com/google-dropped-google-instant-search-279674.

"Search: now faster than the speed of type." *Google Official Blog*. 8 September 2010. https://googleblog.blogspot.com/2010/09/search-now-faster-than-speed-of-type.html.

Shulevitz, Judith. "Alexa, How Will You Change Us?" *Atlantic* 322.4 (2018): 94-104.

Simpson, Lorenzo C. *Technology, Time, and the Conversations of Modernity*. London: Routledge, 1995.

Sparrow, Betsy, Jenny Liu and Daniel M. Wegner. "Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips." *Science* 333 (2011): 776-78.

Stiegler, Bernard. *Taking Care of Youth and the Generations*. Trans. Stephen Barker. Stanford: Stanford UP, 2010.

---. *Technics and Time, 1: The Fault of Epimetheus*. Trans. Richard Beardsworth and George Collins. Stanford: Stanford UP, 1998.

"This Week in Search". *Google Official Blog*. 29 March 2011. googleblog.blogspot.com.au/2011/04/thisTweekTinTsearchT42911.html

"Why Creativity is Being Liberated by Data and Machine Learning." *Think With Google*. October 2019. thinkwithgoogle.com/marketing-resources/experience-design/machine-learning-creativity/

Van Couvering, Elizabeth. "Search Engines in Practice. Structure and Culture in Technical Development." *Cultural Technologies: The Shaping of Culture in Media and Society*. Ed. Göran Bolin. New York: Routledge, 2012. 118-32.

Vlahos, James. "Amazon Alex Search for the One Perfect Answer." *Wired* (18 Jul., 2019). wired.com/story/amazon-alexa-search-for-the-one-perfect-answer

Ward, Adrian F. "Supernormal: How the Internet Is Changing Our Memories and Our Minds." *Psychological Inquiry* 24.4 (2013): 341–348.