Algorithmic subjectivity and the need to be in-formed

Author Neal Thomas

Department of Communication Studies

CB# 3285, 115 Bingham Hall

The University of North Carolina at Chapel Hill

Chapel Hill, NC 27599-3285

neal.thomas@mail.mcgill.ca

Abstract Computer algorithms support collective judgment by organizing it into repeatable and iterable structures. They make communication more consistent and efficient by mechanizing certain aspects of thought, most notably chains of reasoning. With these features in mind, this paper asks the following: under what intellectual assumptions do algorithms also wind up representing a more ineffable aspect of thought, human need?

Keywords Google ; algorithms ; search ; information ; representation ; need.

Citation Thomas, N. (2012). Algorithmic subjectivity and the need to be in-formed. In G. Latzko-Toth & F. Millerand (Eds.), *TEM 2012 : Proceedings of the Technology & Emerging Media Track – Annual Conference of the Canadian Communication Association (Waterloo, May 30 - June 1, 2012)*, http://www.tem.fl.ulaval.ca/www/wpcontent/PDF/Waterloo_2012/THOMAS-TEM2012.pdf

1. Introduction

Discussing the future of the search engine giant in an August, 2010 interview, Google's former CEO Eric Schmidt said the following to the Wall Street Journal:

> We're trying to figure out what the future of search is […] I mean that in a positive way. We're still happy to be in search, believe me. But one idea is that more and more searches are done on your behalf without you needing to type. […] I actually think most people don't want Google to answer their questions […] they want Google to tell them what they should be doing next (Eric Schmidt, quoted in Holman, 2010).

Marketing experts, technology writers and critical communications scholars alike were quick to react to his pronouncement: what future horizon for the web was implied? How might users be surveilled to achieve such a capacity? Such questions highlight the intensifying daily use of social information technology, as well as the need to more critically assess how network devices mediate our signifying practices. The philosophy of technology advises us, for example, to focus on the implicit forms of world-representation and subjectivity that come bundled with our machines and systems. Andrew Feenberg (1999) writes that,

> Technical design is not determined by a general criterion such as efficiency, but by a social process which differentiates technical alternatives according to a variety of case-specific criteria […] Competing definitions reflect conflicting visions of modern society realized in different technical choices (p. 83-4).

Certainly this is the case with Google: there is a particular theory of the subject latent in its operations, in Schmidt's proposal for the service's future, and in the technologies of industrial social computing more generally. Indeed the company's designs on the future serve as a helpful backdrop, to account for how information search intersects with algorithmic technique through its technical representation of *need*. Before opening up this line of thinking, it's important to ask first what it means to say that Google is an algorithmic medium.

An algorithm can be broadly defined as an *effective procedure*; a way of solving a problem or getting something done through a finite number of discrete steps (Berlinski, 2000, p.xvi). If you've ever arrived at a party thanks to convoluted directions, assembled an Ikea bookshelf from start to finish by looking at their infamous stepwise diagrams, or followed a recipe for baking a cake, then at some level your actions have been governed by the logic of an algorithm. That is to say, you have acted by following rules that rationally presume some closed final success condition, or *state*; this is one important understanding of an algorithm.

As they relate to the materiality of computers though, algorithms have a more precise, mathematical definition: they are a recursive function. That is, they apply sets of formal rules to symbols, to produce deductive systems that can iterate, or operate on themselves. The efficiency for human beings is to be found where thinking can literally be made mechanical, by fixing the signs we encode into computers in logically precise ways. We do this in spite of the fact that, by many accounts, signs are not inherently stable; they are rather an expression of our continuous struggle over meaning in everyday life. With this in mind, the paper asks: under what procedural assumptions do algorithms organize discourse, to represent something as seemingly ineffable as a *need?*

Second, it asks: how have procedural schemes for the representation of need replaced one another over time? On this point, what follows is also a brief speculation as to whether we are well served by current schemes for the representation of need. Despite industrial social computing services now being a medium for everyday global communication and personal expression, at the level of technique, need is still largely conceived according to certain key assumptions from the library and information sciences, the academic field from which information retrieval has developed. Is it possible to adopt other perspectives, where a theory of need could be seen more generically as a theory of *desire*? By exploring how algorithms currently produce the conjuncture whereby a user moves from a state of not-knowing to one of knowing, might we conceive of other models of succession that, through their adoption, could *modulate desire* on the border between subjects and objects, in a different way (Deleuze, 1995, p. 44-5)?

Searching for answers to such questions enables one to see how the contingent history of information retrieval processes brought us to the contemporary moment, of Schmidt's more deeply open question of '…what one should do next?' To give a sense of the progression, three strategies concerning need in information systems theory will be outlined. Here they are in point form:

1. A focus on an *instrumental* need for a *specific document*, that follows a simple 'best-match' engineering principle. The approach is developed in traditional computer science, where interfaces satisfy precisely formulated queries. The theoretical framework is a formal-semantic correspondence between subject and object, and historically its focus gives way to:
2. One based on a *cognitive* need for *obtaining knowledge*, where systems help an inquirer model their socially contextualized 'problem situation'; one resolved by providing transformative information. The approach is from the library and information sciences, with a theoretical framework of epistemic correspondence between subject and object. With the rise of industrial social computing, this focus gives way to
3. One based on the *intersubjective, cognitive-existential need* for perpetually obtaining and communicating knowledge; its transmission and reception among all users. Contemporary network interfaces like Google rely on the collective posing and satisfaction of ongoing, socially contextualized 'problem situations', so that they can act as an intermediary for transformative information. The focus is on perpetual, collective sense- making, and the theoretical framework is a *utilitarian-economic* correspondence between subject and object. More simply, it is based in rational choice theory.

The length of each point roughly approximates their prominence in the paper, with the central focus being the current moment of industrial social computing, along with some concluding consideration of its future forms.

2. First strategy: Instrumental need. Document best-match and a focus on planned syntax

Initial strategies for electronic document retrieval unsurprisingly hail from computer science. Following a tradition begun by Shannon and Weaver (1963), information has long been conceived as something 'with extension'—that is, substantialized like *matter*: quantifiable in unit form, and possessing a kind of causal effect of 'reducing uncertainty' through the exchange of messages (Balnaves & Willson, 2011, p. 22-23). Under this paradigm, need is approached from the basic cybernetic mindset of information feedback, with computer systems responding to a user's issued command for a known document by presenting it to them. At this point in the development of information systems, no attempt was made to model cognition socially; search was simply a matter of correct encoding and decoding. To put it in a phrase, need was essentially expressed through the form of the *semantically precise query*: "I need to find the specific document that I believe is called *x*". Summarized in Cole (2011), the role of the user in this first case is to,

> […] (1) collect his or her thoughts on the information that is needed, (2) identify and label these thoughts with concept terms, then (3) forecast which keywords will obtain [citations] to the needed information… (p.1218).

Specific match, or a strict correspondence of terms between system and user is foremost.

Anyone encountering this first generation of search technology today would quickly discover its limitations. Living in what Halavais (2009) calls a 'search engine society', most end users now grasp that informational need is often, if not almost always, partly inchoate: knowing *exactly* the document—or nowadays, person, historical fact, song, street or restaurant—being sought is uncommon; often we may have no idea as to what information object is available to answer our question. Imprecision towards syntax was also quickly determined to be a problem at the level of system design: concepts and the things they represent are naturally expressed through words that are polysemous. A book on cars may be titled "The Big Book of Automobiles", for example; were this simple difference not to occur to a needy user or system designer, the book would never be found. It was in the face of problems like these that librarians and other information experts began to articulate a more complex theory of informational need.

3. Second strategy: Cognitive need. Modeling individual knowledge acquisition as part of a larger 'problem situation'

Improvements to systems occurred as information experts and software engineers became more attuned to the underlying cognitive, or epistemic *context* for need; known by some accounts as the 'problem situation'. Two archetypal theories—Taylor's account of question-negotiation and information seeking, and Belkin's theory of anomalous states of knowledge (or *ASK*) exemplify this tradition in the library and information sciences.

Writing back in 1968, Taylor was concerned to model the lived interactions between a reference librarian and an inquirer. He especially sought to account for how the inquiring process went through communicative *phases of adaptation* between the actors, the librarian seeking to discover the document that answered the knowledge-question of the inquirer. Rendering Shannon & Weaver's initial insight—that information is a measure of "the reduction of uncertainty"—into a more sophisticated framework, Taylor suggested four different phases for informational need: visceral, conscious, formalized and compromised.

Taking the last phase first, *compromised* need is akin to applying the 'best match' principle outlined above. Once an inquirer's question has been formulated into precise terms that "compromise with" what is likely available in an information system, they issue a command for a document that will (hopefully) satisfy the need, thus ending the search. Stepping backwards through Taylor's account, the three prior phases represent an additional prefiguration, or 'focusing in' of this final compromise.

In the third phase of formalized need, for example, the inquirer has established a rational statement of their problem; they can describe their area of doubt in concrete terms, though not in the terms of the information system. In the second, quoting Taylor (1967), "…there is a conscious mental description of an ill-defined area of indecision. It will probably be an ambiguous and rambling statement. The inquirer may, at this stage, talk to someone else to sharpen his focus (p. 127)." Here is where the line between informational need and existential need begin to

blur; read philosophically the line seems to hint at important lived elements of thinking. And finally, in the most primordial and intriguing first phase of his theory, visceral need is described as an unknowable black box; following Taylor it is "only a vague sort of dissatisfaction", "probably inexpressible in linguistic terms (p. 127)."

Belkin's ASK theory follows a similar trajectory. It argues that documents sitting somewhere on a library shelf or in a database are said to possess a stable conceptual state of knowledge, containing beliefs and intentions that form coherent statements. Incoherence on the part of the searcher (need) is resolved when the document that contains the information they require is found. ASK argues that,

> The most general thing that one can say about such a circumstance is that the user, faced with a problem, recognizes that her/his state of knowledge in inadequate for resolving that problem, and decides that obtaining information about the problem area and its circumstances is an appropriate means towards its resolution (Belkin, 1982, p. 63).

In both of these attempts to account for informational need in a more sophisticated way, retrieval was improved by considering and contextualizing the circumstances that lead to a deficiency of information in a user's mind in the first place. The theories try to anticipate the process of producing *sense* through a more socio-epistemic formation, especially communication between two people. It has served as an influential conceptual basis for the development of modern search engine algorithms.

4. Third strategy: Perpetual social need. The intersubjective organization of knowledge

Although originally conceived along the lines of an embodied interaction between librarian and patron, Taylor's ideas have since been reformulated for industrial social computing to serve as the basic premise under which one searches the Web for information, and as important, how one seeks to communicate knowledge to others online. To give just a few examples, Google has automated the vast citation networks of academic papers. It can guess, often correctly, on results for search queries instantly, once a user has keyed in just a few terms. And it uses social relations and the contents of email to personalize results. How

does it achieve such feats? At the level of retrieval technique, Google succeeds based on a crucial insight: that the dynamics of the four-stage path leading from Taylor's visceral need to compromised need (or in Belkin's theory, of the path that lead to the resolution of an 'anomalous state of knowledge') are worth capturing, analyzing, and constantly *feeding back* into the information system itself as a shaping *signal*. In simpler terms, Google stores and analyzes the past behavior of prior users to steer the informational needs of future ones.

Originally it did so by way of its PageRank algorithm, observing the hyperlinking structure of the web. Today it now analyzes hundreds of additional 'signals' emitted by users: how they phrase their needs in the search bar, their geo-location, how they move back and forth from Google's services, and what their friends have retrieved while logged in to its other services like Google+ and GMail. In terms of iterability, past informational needs now encounter, shape and resolve the needs of the present socially, in ways that resemble *markets* more than the planned relationship between librarian and patron. Following Schmidt's account, informational need, understood in epistemic terms, has been reframed into a perpetual daily energy of *intentionality itself*: always metabolizing, displacing and refining the context for thought, making sense of new ideas and new events by comparing them to the prior decisions of other Google users.

Indeed, the rise in importance of services like Google leads others who theorize information system design to expand Taylor's model of need to existential proportions. With systems like Facebook, Reddit and other recommendation engines increasingly playing a 'steering' role in daily life, like Schmidt many claim that information retrieval is a kind of ur-phenomenon that can account for *all* significance in experience. In other words, retrieval drops from an epistemological register down to an ontological one, by way of a loose conceptualization of the term 'information'. Retrieval takes on the import of timeless questions like "Who am I?", "What should I do?" and "How should I live?", conditioning our very survival as human beings through its terms of reference. Schmidt's predictions for Google dovetail with such information theorists as Cole (2011) for example, who writes that,

> This symbiotic dialog between humans and the world is a continuous, existential-level or macro-level quest that does not stop when the individual utilizes an information system to conduct an information search for a specific, micro-level problem or task. This existential-level question forms part of that user's information need and, we assume, is what constitutes a major part of the deepest [visceral] level of the information need (p.1226).

It can be helpful to illustrate this expanded conception of information by referencing an actual algorithm. In today's systems, Cole's existential 'quest' to perpetually resolve informational need is helped along by algorithms like *k*-nearest-neighbour, or *k*NN. The algorithm and others like it are at work mainly in collaborative filtering (CF) services: Netflix's movie suggestion service, Amazon's Recommended for You, and news aggregator sites like Digg and Reddit. *k*NN is a good example of how, through the modern environment of industrial social computing, informational need can be algorithmically structured as continuously intentional, and at least nominally intersubjective: perpetually reorganizing a 'neighborhood' of records for present users according to paths laid down by prior ones. Many items fit the format for this style of organization, especially information-objects for consumer products, people, movies, songs and so on.

CF systems 'watch' users taking steps towards the satisfaction of informational needs, and then uses their paths as training data for positioning them into a more defined space of *prior* users who've taken similar paths. As they participate in the system, the user quickly starts to belong to a neighborhood of people like them, with the relation yielding relevant information-objects more efficiently than might otherwise be uncovered by groping along alone. The dynamic resembles a flock of birds following three basic rules, to coordinate both the realization of their individual needs and those of their neighbours: 1) separate, 2) align, and 3) cohere (Boids, n.d.).

In other words, each bird contributes to governing the entire flock as an emergent effect. As an abstract organizing principle, Terranova (2004) names this the *soft control* of an acentered population (p.100). Each bird steers to avoid crowding others around it, pushing itself away from neighbours. But each simultaneously steers towards the average heading of its neighbours, keeping the group on track towards a location. Finally,

to stay clustered relatively well together, each bird steers towards the average position of its nearest flock-mates, ensuring that the entire population hangs together spatially as a topology (Boids, n.d.).

These are the terms under which users now communicate informational need to one another, producing salience as a useful side effect of the social dynamics of a flocking population. Based on how similar neighbours resolved their knowledge-state in the past, the latest user is recommended items of interest that may resolve their present problem. Selecting those suggested items—downloading them, spending time reading them, buying them, commenting upon them, rating them, putting them into a "wish list" of some kind—is analogous to adopting the average position of the other users in one's flock, leaving a trace that will help steer the *next* user like them. Choices steer a user closer to her neighbors, while also inflecting the wider flock's overall direction through the total space of information, influencing what will be of significance to the next users who join.

Because these systems thrive under conditions of *instability*, immanent and novel difference becomes more important than prior schemas of planned difference. In the previous strategy for informational need, a professional librarian guided an inquirer along socio-epistemic lines. With the rise of industrial social computing, this relationship expands to the whole population, and becomes more functionalistic in scope: *everyone* is potentially guided by their pairwise relations with *everyone else*. Informational need becomes more about an economics of attention, with the user conceived as a member of an aggregate audience-commodity. With a reliance on algorithms like *k*NN, and the perpetual resolution of need through systems of soft control, the user begins to exhibit an *economic* subjectivity; one that is best understood in the terms of rational choice theory. An economic understanding of the subject comes to define not just need for the purposes of information retrieval, but also existential need for the purposes of expression, of being-in-the-world. In other words, at the level of medium it also comes to define the parameters for communication.

A problem here is that rational choice theory gives an account of the subject that focuses, to the exclusion of most other aspects of experience, on individual autonomous choice. The theory has little to say, for example, about the interpretive or situated *substance* of decisions,

nor their social or cultural dimensions, nor how decisions may bear on reproducing normative structures of power. Instead, actors simply hold beliefs, desires and states of mind as kind of operational "portfolio" of possibility, and act through those by deciding what to do. Hindess (1988) paraphrases the rational philosopher Martin Hollis in giving a succinct list of the theory's features:

> First, actors are rational and their rationality is understood in strictly utilitarian terms. Actors have a given set of ends, they choose between them in a consistent fashion, and they select from the available means of action those most appropriate to the realization of their chosen ends. In this sense of rationality, the ends themselves are neither rational nor irrational, they are simply there. Secondly, actors are assumed to be narrowly self-interested. Thirdly, they are social atoms: 'they could be picked at random from their groups, because it made no difference *who* they were' (p.29).

With the embedding of rational choice theory as a basic mediating condition of life online, by way of algorithms like *k*NN, the user in need of information becomes an actor who makes choices from the position of privately chosen ends. Needs are defined through a desire to strategically optimize one's outcomes, however those needs may be meaningfully conceived by the individual. *Collective* ends are comprehensible only on the basis of this private position: making choices mostly establishes the user into a conceptual neighborhood of agreeable resemblance. What happens to communication? Thicker theoretical accounts of difference, and arguably even the possibility of *dissensus,* risk being flattened or obscured by the medium, as communicative agency is defined in terms of abstract logical choice. To put it another way, the potential for mutual criticism does not persist in these systems; instead, disagreement energizes the recalibration of the system, so that agents at odds with one another 'veer away' into their respectively more agreeable information spaces. Contestation in communication converts into a shallow 'numbers game', as when political shills swamp an online poll, or a news site's 'thumbs up/thumbs down' sorting system, for example.

It's at this point that critical concerns towards Schmidt's prediction from the beginning of the paper become more pronounced. Once informational need and the existential realization of need *per se*—including public communication—merge together under the umbrella of

rational choice, experiential significance gets subtly redefined through algorithmic media as an information channel, while the expression of significance gets reduced to a behavioural side-effect of private decision-making and goals. When Schmidt talks about searches done on your behalf, or Google telling you 'what to do next', he is effectively arguing for the user to *delegate their judgment* to a Google-driven device, structured by a behavioural-economic view of what it means to have judgment, or to be rational. Via contemporary forms of algorithmic iterability like *k*NN, the effect is for social information systems to apply a subtle but pervasive form of what Habermas called functionalist reason to all electronic discourse. At the level of critique, the risk is that experiential significance and expression fall prey to an intensified level of bureaucratization.

References

Balnaves, M., & Willson, M. A. (2011). A new theory of information and the Internet: Public sphere meets protocol. New York: Peter Lang.

Belkin, N.J., Oddy, R.N., & Brooks, H.M. (n.d.). ASK for Information Retrieval: Part I. Background and Theory. The Journal of Documentation, 38(2), 61–71.

Berlinski, D. (2000). The advent of the algorithm: The idea that rules the world. New York: Harcourt.

Boids. (n.d.). Retrieved from http://www.red3d.com/cwr/boids/

Cole, C. (2011). A theory of information need for information retrieval that connects information to knowledge. Journal of the American Society for Information Science and Technology, 62(7), 1216–1231. doi:10.1002/asi.21541

Day, Ronald E., & Lau, Andrew J. (2010). Psychoanalysis as Critique in the Works of Freud, Lacan, and Deleuze and Guattari. In G. J. Leckie, L. M. Given, & J. Buschman (eds.), *Critical theory for library and information science : exploring the social from across the disciplines.* Santa Barbara, Calif.: Libraries Unlimited.

Deleuze, G. (1995). *Negotiations, 1972-1990*. New York: Columbia University Press.

Feenberg, A. (1999). *Questioning technology*. London: Routledge.

Habermas, J. (1984). *The theory of communicative action, vol. 2*. Boston: Beacon Press.

Hindess, B. (1988). *Choice and rationality in social theory.* London ;Boston: Unwin Hyman.

Ingram, D. (1987). *Habermas and the dialectic of reason*. New Haven: Yale University Press.

Jenkins, W.H. Jr. (2010, August 14). Google and the Search for the Future. *The Wall Street Journal.*

http://online.wsj.com/article/SB10001424052748704901104575423294099527212.html

Shannon, C. E. (1963). *The mathematical theory of communication* (Illini Books ed.). Urbana: University of Illinois Press.

Taylor, R. S. (1967). Question-negotiation and information-seeking in libraries (no. 3). Lehigh univ bethlehem pa center for information science.

Terranova, T. (2004). *Network culture : politics for the information age*. London ;Ann Arbor, MI: Pluto Press.

Thirumuruganathan, S. (2010, June 17). A Detailed Introduction to K-Nearest Neighbor (Knn) Algorithm [blog post]. Retrieved from http://saravananthirumuruganathan.wordpress.com/2010/05/17/a-detailed-introduction-to-k-nearest-neighbor-knn-algorithm/.