

From Digital Dumbness to Digital Consciousness: The Know-How of Information Technologies

Hakim Hachour, Safia Abouad

Laboratoire Paragraphe EA349, Université Paris 8, France
hakim.hachour@univ-paris8.fr

Abstract. This paper aims to show how digital agents impact the quality of information practices. It reports key results of a five years study on the copy/paste practices of students at a Master degree of information sciences and librarianship. This work shows how routine use of information technologies goes with some reduction of explicit understanding of their effects on works. Thus, digital agents influence the decision process, action and may provoke unexpected consequences. It is then discussed how digital consciousness determines digital intelligence.

Keywords: information practices; knowledge acquisition; sociotechnology

1 Introduction

Recent researches converge toward a consensus: the use of information technologies, which are widely spread in industrialized societies, is becoming more and more crucial for a large spectrum of human activities, both professional and recreational ones [1]. Thus, information practices [1, 2] have evolved and multiplied, integrating various digital strategies, digital resources and computer-based methods. Digital agents (DAs) are not only resources exploited by users, they are entities (interacting systems of software and hardware) which are able to interact with humans through computational treatments of information in order to achieve objectives according to some rules of interpretations and sets of constraints; they can be distinguished from human users because of their artificial nature. From this perspective, users and digital agents tend to form overlapping sociotechnical systems of information. Here, the problematic concerns the potential diversity of objectives pertaining to agents (users and digital agents). If users' cognitive processes take into account contexts, situations, biological memories, informal knowledge and complex motivations, DAs' computational processes are purely based on technical sensors, formal operations, and algorithms. Moreover, cognition and computation processes are not always observable because of their complexity for the former and confidentiality for the latter. Since DAs interact with one another and with users, we chose to apprehend this kind of "*sociotechnical ensembles*" [3] from a communicational point of view and a common model. In order to provide useful concepts to discuss the impact of digital agents on the effectiveness of sociotechnical system, we have analyzed the copy/paste practices of five groups of students enrolled into a course on the basics of computer sciences from 2009 to 2014. The main hypothesis suggests that the increasing usability of digi-

tal agents' products and services, and their simulacra of immediacy, affect the conditions whereby users and DAs can reach a mutual understanding on their objectives; this effect could lead to implicit consequences on cognitive representations that counteract "digital intelligence."

2 For a Comprehensive Approach of Sociotechnical Systems

From a socio-phenomenological perspective, we focus our analysis on the relevant phenomena which are produced and denoted by actors of a sociotechnical system, either digital agents or people as long as these participate to their production, can interpret them and act according to this interpretation [4]. Following ethnomethodologists [4, 5], we define 'relevant phenomena' as 'accounts.' Those are meaningful facts, expressions or objects which are indexical and reflexive [4, 5]. Indexicality defines the pragmatic property of the interpretation process: two types of meaning are associated to a phenomenon, a formal meaning (the encyclopedic or explicit definition) and an indexical meaning which potentially refers to different kind of tacit knowledge (routine, contextual, biographical, and cultural, etc.). The other property of accounts is reflexivity, which means that meanings are built during interaction; predetermined schemes of interpretation are actualized according to the ongoing events, they are defined according to routines and are transformed by the necessary adaptations which have to be made in order to fit the "*situated action*" [4]. Here, digital intelligence defines the intelligence of sociotechnical ensemble, which is determined by the quality of its productions and individual skills certainly, but moreover by the level of mutual understanding between the agents. This pragmatic approach is in adequacy with the objective of this research, which is to describe sociotechnical system of information. Ideally, a high level of mutual understanding and effective actions, according to practical purposes, correspond to agents' good coordination and the concomitant satisfaction of their specific interests and a common objective.

3 Copy/paste Practices and the Limited Mutual Understanding between Digital Agents and Users

Through an active participant-observation, we collected data from synchronous and asynchronous communications (emails, documents, reports, audio recordings, pictures) of an ongoing course (30 h/year) on the basics of "computer sciences" addressed to master's students in information sciences and librarianship that mostly have a naïve conception of digital technology. To graduate this course, they had to produce a report about one of the course's themes (hardware, software, internet technologies, computer networks, history of computer sciences, cryptography, safety, digitalization of media, databases, etc.). Each report (Type A) had to content between 10 to 15 thousands characters (~ 2000 words). From 2009/2010 to 2012/2013, students were explicitly informed that their reports would be scrutinized, manually and automatically with an "anti-plagiarism software." Students were told how to quote

references (textbooks, webpages, pedagogical contents ...) following the APA methodology. Moreover, each year, the teacher showed them how it was possible to identify similar contents with this kind of software and told them the different subjective indicators that could alert him (changes in fonts, layouts, registers of language, high complexity, metadata of content's fragments ...); it was also demonstrated, on the basis of scientific data on natural language processing techniques, that it is unlikely to produce a text containing more than 10 consecutive strings of characters that is not original, excluding general language formulas like "In-the-beginning-of-the-twentieth-century-, -main-researches-focused-on-(...)." It was also clearly established, from the point of view of the teacher, that unreferenced 'copy/paste' contents were considered as the manifestation of 'dumbness,' either because it would impact students' assimilation of knowledge or because it may be understood as mockery. The last promotion was asked to produce another type of report with different conditions (Type B). This promotion had to elaborate an oral presentation and a short visual support (5 slides or pages), each group of two students then had to present their work and answer to improvised questions; this oral communication lasted at least 30 minutes for each group. 99 reports were analyzed (similarity and discourse analysis), followed with the conversational analysis of in situ interactions between the teacher and students and of mediated communication with students that were concerned by the presence of "plagiarism" in their reports.

Table 1: Key results on the copy/paste practices of five classes (Master degree) in information management and librarianship (99 reports)

	2009-2010	2010-2011	2011-2012	2012-2013	2013-2014
Number of analyzed reports (a)	30	25	26	10	8
Average grade	14,1/20	15,5/20	13/20	14,9/20	14,6/20
Lowest grade - LG (nb of words)	3 (9896)	10 (1665)	3 (3600)	10 (2965)	14 (413)
Highest grade - HG (nb of words)	19 (10078)	20 (13105)	20 (10684)	19 (5661)	17 (586)
% of copy/paste of LG	34%	9%	38%	13%	0%
% of copy/paste of HG	2%	1%	12%	<1%	24%
% of unreferenced copy/paste of LG	34%	7%	38%	13%	0%
% of unreferenced copy/paste of HG	0%	0%	0%	0%	9%
Nb of references of LG	0	10	10	16	0
Nb of references of HG	13	12	24	16	2
Type of report (b)	A	A	A	A	B

Notes: evaluation of reports differed according to their type. (a) From 2009 to 2012, individual reports were required. In 2013/2014 8 groups of 2 students were asked to give a report. (b) Type A, asynchronous evaluation: students had 3 weeks at least to choose a subject and write a textual report based on a template. Type B, synchronous evaluation (2014): groups of students were asked to produce a visual presentation on an imposed subject in 3 hours max., to give an oral communication and to answer improvised questions from the teacher.

As it is mentioned into the specifications of the anti-plagiarism software, the results do not define the quality of the content but only similar content comparing to accessible datasets (the web for the most part and private corpus of students' works). Even for a knowledgeable person that understands the processes involved in the computation, it is not *a priori* possible to know all the 'programming tricks,' moderations or choices made by software. It is then necessary to complete the analysis of copy/paste elements manually and qualitatively on the basis of the subjective indicators mentioned before and random tests. Table 1 cannot contain all the results, nevertheless, key results are reported. Mainly, we can pinpoint the heterogeneity of the interpretations of the teacher's instructions. Some students asked for the authorization to produce more than 15 thousands characters, and they were allowed to. All the course's themes were chosen at least one time, the most selected themes refer to hardware and internet technologies, but it is interesting to notice that 7 reports explicitly cover themes which were not studied (videogames, quantum computers, e-books display devices, Voice Over IP, e-business, digital audio station), corresponding with high grades. Concerning copy/paste practices, no clear findings could be expressed for the reports with average grade, nevertheless, highest grades and lowest grades coincide with distinctive information and discourse practices. The high quality reports were made by students which reformulated their sources: very few copy/paste contents were found and when it was the case, there were all well referenced (0% from 2009 to 2013; even for figures and graphics). These reports mostly cover original themes and contain significantly more content. The structure of high grade reports is also more developed, some students used the provided outline, and others produced an original template, but these reports also satisfy optional instructions as the edition of indexes and detailed tables of contents (titles, figures and table ...). Their discourses are more original, following sequences of concepts and descriptions that were not exemplified in pedagogical supports or courses. The time response is irrelevant; reports of both high and low quality were produced with various amount of work time. Low quality reports were not necessarily shorter than the average (e.g. LG of 2009-2010), as the average grade ones, they tended to respect the frame of instructions (1500-2500 words). Some of the low quality reports refer to as much sources as the high grade reports.

The main correlation is made between the high rate of unreferenced copy/paste contents and the following factors: lack of originality, poor layout and confusion between concepts. When students with high rate of copy/paste reports were informed, most of them (excepted one student) denied it at first. It was necessary to provide accounts of the similarity analysis and to discuss precise cases; it was needed to repeat the demonstration made in classroom. All conversations evolved towards a justification process where new accounts were produced. When students were asked about their conditions of academic work, they all referred to complex activities where they had to switch between different tasks (multitasking, managing children, being at the workplace), combined with a "writer's block syndrome." They did not give accounts of the essential role of search engines into their practice and did not think how their production would be used in other contexts. In contrast, students with high grades have mastered multiple levels of indexicality and have projected multiple applications

of their work; their reports contain accounts of their intelligent collaboration with other agents (teacher, other students, search engines and word processor).

4 Understanding How to Collaborate with Digital Agents

Students with low grades did not take into account other potential uses of their productions. The sociotechnical ensemble wherein they participated (teacher, students and DAs in interaction) has produced reports that satisfied only some of the interests in stake, omitting how the report could be a resource for other sociotechnical ensembles. All high grade students have considered multiple contexts where their report could be used: professional work, other courses, their final dissertation, etc. In the case of low grade reports, agents did not manage to satisfy the multilayer objectives: e.g., the search engine could have satisfied its prerogatives by providing relevant contents (as they have been used) according to student's requests and by trying to make the user converge toward certain web pages (through algorithms and advertising), and that, without considering the overall goal of the ensemble. These kinds of collaboration constitute 'dumb' sociotechnical ensembles. Intelligent sociotechnical ensembles have satisfied all agents (DAs, the student and the teacher at least). Intelligence, as "the ability to learn or understand or to deal with new or trying situations" (Merriam-Webster), does not imply for users to master the basic technical explications of digital agents and their products. Users rather need to be aware of their 'next-to-be' sociotechnical ensembles and think about how these relate with the knowledge built upon other experiences. On the one hand, the multiple interconnections between overlapping sociotechnical ensembles provides the means for new accounts to emerge, thus enhancing experience, on the other hand it can create confusion. In conclusion, the mandatory condition of digital intelligence is the capacity for agents to elaborate their representations upon a set of meta-information that transcends ongoing activities; as DAs do not always exhibit this capacity, the efficiency of the collaboration relies on the level of mastery of other agents. Therefore, for digital intelligence to be a reality, agents must acquire not only digital skills but digital consciousness, or as Nelson proposed: "media consciousness" [6].

5 References

1. Savolainen, R.: *Everyday Information Practices - A Social Phenomenological Perspective*. Lanham, Md, Scarecrow Press (2008).
2. Chaudiron, S., Ihadjadene, M. : De la recherche de l'information aux pratiques informationnelles. *Études de communication* n° 35, n° 2, pp. 13-30 (2010).
3. Bijker, W.E., Law, J.: *Shaping Technology/Building Society - Studies in Sociotechnical Change*. New edition. Cambridge, Mass., MIT Press (1994).
4. Garfinkel, H.: *Ethnomethodology's Program*. Rowman & Littlefield (2002).
5. Heritage, J., Clayman, S.: *Talk in Action - Interactions, Identities, and Institutions*. John Wiley & Sons (2011).
6. Nelson, T.H.: *Computer Lib/Dream Machines* (1974). In Wardrip-Fruin, N., Montfort, N., *New Media Reader*, MIT Press, pp. 303-338 (2003).