Naturality and Non-Transparency of Technology in the Age of Intelligent Voice Assistants

Jaroslaw Kowalski[1], *Cezary Biele*[1], *Kazimierz Krzysztofek*[2]

[1] National Information Processing Institute

[2] SWPS University of Social Sciences and Humanities

ABSTRACT

This article describes the psychosocial consequences of using voice assistants, based on a look at the interface as a tool that allows the use of another tool. In analysing the history of interfaces and considering them in terms of naturalness (ease of use) and transparency (visibility and understandability to the user), we observe that they long evolved in the direction of unnaturalness, but have recently reverted towards naturality. A late stage of interface development is that of the voice interface; one that is highly natural (controlled by natural speech) and extremely non-transparent. The combination of these properties results in such interfaces exerting a significant influence on human functioning.

Keywords: Intelligent Voice Assistants, Internet of Things, Human-Technology Interaction, Smart Environments

INTRODUCTION

Amidst the development of computer techniques, the term, 'interface' began to appear. Initially, such equipment was used to describe equipment that allowed the linkage of two or more devices that were unable to cooperate without it. The term has since broadened in its definition to cover the types of technology that allow humans to communicate, program, and manage the internal processes of computers. Currently, the Cambridge Dictionary defines an interface as 1) 'a connection between two pieces of electronic equipment, or between a person and a computer'; and 2) 'a situation, way, or place where two things come together and affect each other'.

Emphasis has been placed on various aspects of this definition, including on the interface's use as a metaphor (Hutchins et al. 1985), its embedding in interaction and user goals (Norman, 1988), its perception as a relationship with technology, rather than a technology in itself (Hookway, 2014), and even on its function as a performance piece (Laurel, 2014). In this article, we wish to direct the reader's attention to the notion that an interface can be considered a tool that allows another tool to be used more easily. This definition offers a fresh perspective on the entire history of technology and asserts the existence of interfaces long before the invention of computers. The development of technology can be viewed as a process in which successive levels of technological interfaces have been added. For example, one rudimentary tool was once required to sew clothes: a needle (human–needle). Next, machine interfaces were introduced and the sequence altered (human–sewing machine–needle). Following the digital revolution, the sequence has grown more complex (human–GUI–computer operating system–processor–electrical circuits–sewing machine–needle). When considering the evolution of the interface, two useful terms might be introduced to the discussion: naturality and transparency.

INTERFACE CHARACTERISTICS: NATURALITY AND TRANSPARENCY

In this article, the term, 'natural' concerns technology to which humanity has evolutionarily adapted; those we, as humans, have adapted to our bodies and cognitive abilities. For example, using a hammer is highly natural due to humans' ability to catch items; conversely, the machine language of programming (of the low order) is less natural, as using a processor is not a skill with which humans have been equipped by nature, but one that must be learned through long and laborious training. An order can be observed when analysing the subsequent stages of interface development. The earliest interfaces were characterised by a high degree of naturality. Prime examples might include knives, hammers, and spears: tools that were designed to complement the manner in which the human body functions. Later, other types of technological interface were gradually introduced to control such tools—including different varieties of lever, embankments, and electrical circuits—which decreased

the tools' naturality. As technology has developed, the naturalness of interfaces has declined; from a certain point, however, additional layers of interfaces began to be added, which were more natural. Text interfaces, based on computer programming by the language of orders understandable by CPUs (machine language) are those that have diverged furthest from their origin. Subsequent text interfaces were tasked with issuing orders to CPUs with the assistance of special overlays: first, artificial commands that were specifically designed for programming languages (in which written code was compiled to machine language), and later to operating systems. The next steps are a further approximation towards naturalness: graphic interfaces (e.g. Windows), mouse control, and touch interfaces that employ styli. Each of these constitutes a step towards increased naturalness. Writing commands in an artificial language is more abstract than using a mouse (during which, the movement of the hand synchronises with that of the cursor on the screen). Bypassing the mouse and touching the screen with a stylus is even more natural; discarding the stylus and replacing it with a human finger is discarding the last prosthesis between the body and the effect of the action. This was elegantly summarised by Steve Jobs, who said: *'God gave us ten styluses. Let's not invent another'* (Isaacson, 2011).

Aside from naturality, another dimension that is worthy of consideration is transparency. Analysing the interface sequence presented above, it can be observed that newer interfaces are becoming progressively less transparent. This means that the relationship between our actions and their outcomes is the result of laws, principles, and interactions that are becoming increasingly difficult to observe or control. This can be illustrated by the example of a bow – a tool positioned near the down of the interface hierarchy. Each stage of the causal sequence of a bow's operation is observable and controllable. This becomes considerably more difficult in the case of a radio receiver, for instance (although it is possible for specialists): the user presses the 'on' button, an electrical circuit is shorted, current flows to the appropriate circuits, and the speaker begins to emit sounds. A history of augmenting additional layers of technology onto existing ones is one in which humans have progressed from natural and transparent devices and interfaces, through limitedly natural and limitedly transparent ones, to natural, but extremely non-transparent ones (as shown in Figure 1.).

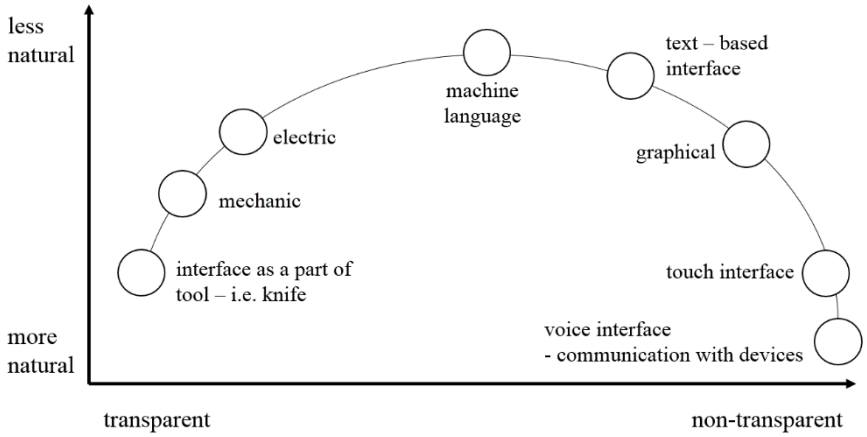

Figure 1. The development of the interface: naturality and transparency

VOICE INTERFACES

Voice interfaces that employ natural language (such as Google Assistant, Siri, and Alexa) are already in everyday use. In terms of naturality, the initial levels have been achieved – interacting with an electronic assistant is almost as natural as conversing with another human; the mechanism located between the issuing of a command and the reception of the final result, however, remains extremely non-transparent. Let us consider the following case: we ask the Google Assistant to direct us to the nearest café. After a moment, the assistant provides us with the name and the address we requested. What really happened between our query and our receiving the results? We know that our request must have been analysed by Google software. What happened next? Is what the Google Assistant did simply analogous to what it does when we enter the words, 'café' in the search engine with a marked location? Did Google suggest the cafés that we have previously visited? To one layer of non-transparency associated with the human brain's lower ability to intuitively discern the abstract laws of physics and magnetism (which informed inventions such as the radio, the television, and a variety of electronic devices), the voice interface introduces another that is connected with a characteristic of data processing in artificial neuronal networks using artificial intelligence algorithms. Voice interfaces possess the characteristics of black boxes – meaning that tracing the process by which solutions were devised is challenging. By definition, these are not algorithmised processes.

In recent decades, a process has occurred in which the ways humans utilise various types of technology have become increasingly natural and, simultaneously, increasingly opaque. The voice interface is the final stage of this process. Using it is almost as easy as conversing with another human, yet is simultaneously extremely non-transparent. The next section considers the consequences of using such devices.

NEW POSSIBILITIES

Voice interfaces are highly likely to become more widespread. A key reason for this lies in their ease of use – which is rooted in their naturality (Kowalski et al. 2019). In recent decades, interfaces that were more natural superseded those that were less so. A prime example can be found when considering the influence of the introduction of graphical interfaces on the widespread adoption of home microcomputers. The introduction of tablets and touch interfaces constituted the second threshold of the availability of computer technology. More natural methods of issuing commands and queries are always more accessible than less natural ones, and widen the potential target group of products. The popularisation of the voice interface will entail major consequences on the functioning of humanity.

The voice interface is, first of all, flexible; it adapts to its user. Data on behaviour, searches, tastes, preferences, purchases, and much more—in short, the entire user profile— has become a valuable currency (Zuboff, 2019). This information is presently gathered and processed by internet service providers, which has resulted in users being presented tailored offers for use of internet services, and search results that differ depending on what kind of information particular services or search engines are able to access on their users. The popularisation of the voice interface and its convenience means that voice assistants process huge amounts of information on their users. In combination with artificial intelligence, this will create new and previously unknown characteristics and operational possibilities. These features will make voice assistants similar to human agents. In the future, machines will best know their users, and will be best positioned able to predict their actions and intentions (possibly even motivations). Another step involves enabling such interfaces to become even more natural and anthropomorphised, bringing the technology closer to how humans operate (human assistants). The institution of secretaries will once again become popular, helping their supervisors to manage considerable numbers of tasks. Efficient, intelligent concierges will be available to everyone. The knowledge a tool possesses on its user will facilitate the adoption of another characteristic: proactivity. Voice assistants will be able not only to issue commands, but also to actively propose a variety of solutions. For example, an assistant might direct a user's attention to a conference about which they have forgotten, or remind them to close the windows before an incoming storm; perhaps, when a user grants it special authorisation, it will even close the windows itself via a smart home installation. At a conference, it will be able not only to inform the user of a selection of clashing lectures to attend, but will also guide them to the one that has the highest probability of being well received. Another feature stems from the proactivity of this interface: autonomy. As what is comfortable and time-efficient will always win against something that is less so, pressure will arise for users to cede autonomy to the assistants. If an assistant knows our needs and likes as well as we do, allowing it to shorten the decision-making process—for instance, by choosing the best restaurant in an area—and providing an answer immediately seems a logical progression. This constitutes a fundamental paradigm shift: until now, it has been the user who searched for information and

selected the most suitable parts. Sometimes, they would request others' opinions and add them to a pool of choices; in other words, they created their own action possibility menu – and only then did they select from that menu. The prevalence of voice systems will mean that users can be presented immediately with a solution. This phenomenon, which already exists in the case of popular search engines, will be amplified. Discovering the best answer among the top ten results will no longer matter; but occupying the top result will.

The effectiveness of voice assistants presents an opportunity to make the closest technological layer to humans a reality; one that is analogous to a trusted secretary, who plans and actively manages a user's schedule. The user, in this case, achieves gains in efficiency (they can have more meetings, learn more, and act faster), but simultaneously cedes control of their own schedule – in much the same way as a prime minister, a president, or the pope do.

THREATS

It is also imperative that we consider the dangers and negative consequences of using voice interfaces. Such complex tools will not belong to their users (on the same basis that they are owners of computer mice, for instance), but will be the property of private companies; such interfaces will implement the agendas of their owners, not their users. We are already experiencing this phenomenon as we use websites and applications in which the design of the interface influences the choices users make; users often focus on what is available on the menu and they fail to wonder why their options are so limited. For example, when using Google search, they often mistakenly believe that the whole internet has been searched for them, ignoring or remaining unaware that swathes of it remain unindexed. Moreover, they most often select consume information contained in the top ten search results. Another example can be found in the transformation of a user's original aim into a secondary one: the aim of the program or application. Tristan Harris (Harris, 2016) offers an example of a situation in which a group of friends, wanting to find a place where they could talk, turn on an application that recommends bars. The application is designed in such a way that photographs of the drinks served at different bars occupy a significant portion of its space. Later, the group of friends compares various restaurants based on drink photographs. Their original aim (finding a place to talk) has altered to finding a place available in the application using the criterion of which boasts the prettiest drink photograph. The content of menus influences what we expect of technology and facilitates the redefinition of our original goals.

This feature is of particular relevance in the use of voice interfaces. On one hand, using such interfaces is incredibly easy; on the other—as a result of their non-transparency, and their tendency to offer singular and definitive answers to our questions—analysing the cases in which the objectives of the owners (corporation) differ from those of the user entails great difficulty. Examples from life show that this is often the case. Introducing a new layout to a bank application might be aimed not at accelerating the activities that its users perform (such as authorising bank transfers),

but at slowing them, so that users have ample time to familiarise themselves with the other offers, such as insurance and loans, presented on the screen. In effect, to make a transfer, a user has no choice – they must realise the objectives of the technology supplier or corporation that designed the interface. Technology frequently forces us to make inconvenient choices. This usually occurs in a way that enables certain actions (those desired by the interface's creators) to be simplified, and others to be undertaken with greater difficulty. For example, on some websites, paid content can be accessed with little more than a click; withdrawing from such arrangements, however, is possible only via a convoluted process, which includes a personal conversation with an operator to be initiated and for the reasons of the withdrawal to be specified. Many webpage interfaces and applications also employ the 'foot in the door' strategy; humans demonstrate difficulty estimating the entire cost of clicking on something, and technology designers often capitalise on this trait. Having invested five minutes in an activity, users find it difficult to resist investing an additional five or ten minutes.

The example mechanisms described above (in reality, there are certainly many more) demonstrate how a tool, via its interface, specifies an area of interaction in which it influences a user, directs their attention, provokes them to make particular choices, and increases or decreases the likelihood of them behaving in particular ways; thus, influencing how the tool is used. Agency is no longer exclusively an operator trait or a user trait, but one that has shifted onto the technology itself.

As voice interfaces will use artificial intelligence and possess large amounts of data on our past behaviours (including our consumer and behavioural profiles), the application of the invisible techniques mentioned above will be much simpler. We need interfaces to communicate with technology in our own language and on our own terms, and we are also subject to all of the limitations that they impose on us. The question arises: who will devise the objectives of the interfaces? In transparent technology, either we set the aims, or we see them (because they are transparent). With websites and algorithms of graphic interfaces, this task becomes more challenging. Aside from the obvious objectives of users (e.g. 'make a bank transfer'), those of interface owners also function (e.g. 'ensure that a user's attention is directed towards an insurance offer and they buy it'). In completely non-transparent voice assistants, or in the artificial intelligence systems running behind them, the difficulty this entails will increase.

The issue of voice interface placement is also one that is worthy of consideration. Currently, we think of the technology as something related to a device that listens to us. This could be a speaker (e.g. Alexa or Google Home), but also a cell phone or a smartwatch. In reality, a virtual assistant is something more than a speaker or a device that incorporates a microphone. The commands that users issue are processed in the computing cloud – far beyond the confines of the speaker. The assistant, additionally, is not assigned to a particular speaker. It is possible to link a few speakers that stand in different rooms of the same home into one system. A smartphone and speakers will then comprise parts of the net that surrounds a user in such a manner that means they have the most convenient access to it. In this case, the interface is invisible and built into the external world. Voice assistants will, therefore, become the invisible and

omnipresent interface of everything.

CONCLUSIONS

New type of interface will present new and unprecedented features. The price paid for the technology's naturality is its extreme non-transparency; its obscurity, complexity, and convoluted nature. This state of affairs, which is being faced in the development of current information technology (the implementation by interfaces of their own agendas, which occasionally stands in contradiction to users' goals) will intensify. It is to be expected that the amount of information on a user that interfaces currently possess and process will not decrease in comparison to the existing solutions. This will sow the seeds of a condition of far-reaching asymmetry, under which interfaces will be trained on neural networks and hold knowledge on millions (maybe billions) of users. Interfaces will know how to influence a particular user, to select what is in the interest of the interface's owner, or to undertake a desired action.

In the modern world, knowledge translates into power (Foucault, 1975), and this will be the first time in history that humanity faces a situation in which the apparatus that runs behind interfaces can secure an advantage over the users on such a scale. The non-transparency of such interfaces will cause users to understand their tools progressively less; in extreme cases, the asymmetry in relations might go so far that it will catalyse a reversal of the tool-user relationship. Human beings will be the tools employed to achieve particular aims; a means to an end; a mere extension of technology. A rudimentary example that has already been employed can be found in the use of urban road traffic control – applications that display traffic density and navigate users to their destinations. Drivers rely heavily on the information they receive from such systems. They must trust that a particular road constitutes the most efficient journey. It is sufficient for an application to have a different objective—such as directing the traffic to avoid accidents—to create a divergence from the users' objective (to get somewhere as fast as possible); human users have become a tool to further the system's objective. Some features of voice interfaces will serve as an example of Postman's 'invisible technology' (Postman, 1993): systems that humans fail to see, and that we suppose are neutral by definition. As Martin Heidegger (Heidegger, 1977) noted, however, many of the technologies that have exerted the greatest influence on humanity are widely perceived as neutral.

REFERENCES

Foucault, M. (1975) Surveiller et punir. Naissance de la prison. Gallimard. Paris

Harris, T. (2016) How technology is hijacking our mind – from a Magician and Google Design Ethicist. https://medium.com/thrive-global/how-technology-hijacks-peoples-minds-from-a-magician-and-google-s-design-ethicist-56d62ef5edf3

Heidegger, M. (1977) The Question Concerning Technology and Other Essays. Harper Torchbooks
Hookway, B. (2014) Interface. Cambridge, MIT Press
Hutchins, E. L., Hollan, J. D., Norman, D. A. (1985) “Direct Manipulation Interfaces” in: HUMAN-COMPUTER INTERACTION, Volume 1, pp. 311-338
Isaacson, W. (2011) Steve Jobs. Simon & Schuster
Kowalski, J., Jaskulska, A., Skorupska, K., Abramczuk, K., Biele, C., Kopeć, W., Marasek K. (2019) “Older Adults and Voice Interaction: a pilot study with Google Home” in: CHI ’19 Extended Abstracts CHI Conference on Human Factors in Computing Systems; 6-9. 05. 2019 Glasgow
Laurel, B. (2014) Computers as a theatre. Adison-Wesley
Norman, D. A. (1988) The Psychology of Everyday Things. Basic Books
Postman, N. (1993) Technopoly: The Surrender of Culture to Technology. Vintage Books. New York
Zuboff, S. (2019) The Age of Surveillance Capitalism. Public Affairs