

Computational data, Measurement and the Double Bind of Audience Information

Aleksi Aaltonen

The London School of Economics and Political Science

v.a.aaltonen@lse.ac.uk

Abstract

This paper discusses the central role of audience measurement in fabricating the sellable asset in the media industry, that is, the advertising audience. I will argue that the taken-for-granted assumptions regarding the nature of information implied, for instance, by statistical modelling are inadequate to understand how technological information shapes media organizations and the industry in general. The paper shows how the commercial media exists in a sort of double bind between evolving measurement technologies and the fragmentation of audience behaviour. The arguments are substantiated by making a simple conceptual distinction between data and information and a short excursion to the history of audience measurement arrangements. Finally, I use an empirical case study to identify novel attributes of computational media environment and juxtapose them against the operations needed to generate relevant audience information out of the computational data tokens.

A paper presented at the 27th European Group for Organizational Studies (EGOS) Colloquium, July 7–9, 2011, Gothenburg, Sweden

Introduction

A great deal of business and new services in the new digital environment is predicated on large-scale advertising (Brown & Duguid 2000) – the apparent freeness of internet is to a significant degree based on either real or projected advertising revenues. Advertising-funded companies such as search engines, social networking sites, and, more traditionally, newspapers, television channels and radio stations operate on what economists call two-sided markets. They subsidize content and services for consumers, whose attention in the media space is then packaged and sold as audiences to advertisers (Parker & Van Alstyne 2005). This is not to say that media companies are solely funded by advertising, yet understanding the industry by its predominant business model has certain advantages over other types of definitions (cf. Hoskins et al. 2004; Küng 2008). With some notable exceptions such as the BBC, the media industry is organized as commercial companies that generate significant revenues by incorporating third party marketing messages into their consumer offerings. Advertising remains as the most important source of income for media companies in general, and the definition based on business model has the important advantage that it includes also the variety of companies not involved in traditional content production activities while still competing for advertising revenues. The empirical part of this paper will discuss one such case, an advertising-funded mobile virtual network operator (MVNO) company.

The technical capability to relay advertising messages to individual consumers constitutes, however, barely a necessary precondition for the existence of a viable advertising medium. This is because advertisers do not, in the first place, pay for the access to the media space but for the attention of a specific audience it attracts. No advertiser is ready to make a significant investment for sending advertisements to an audience of unknown quality and quantity, that is, without knowing the audience product. Every new electronic medium faces thus the problem that the audience it tries to sell is initially invisible. In contrast to live audiences, there is no inherent way to verify the reception of messages relayed by an electronic medium and therefore knowing if the consumers are paying any attention to the advertisements. Solutions to the problem of how to observe people's media consumption and turn the measurement data into information about the audience have historically had a significant impact on the media industry (Barnes & Thomson 1994; Napoli 2003; Webster et al. 2006). The need for valid and reliable audience estimates serving commercial purposes has given

rise to a whole industry of research companies offering a variety of information products and services. Furthermore and most importantly from the perspective of this paper, measuring and analysing elusive media consumption patterns are fundamentally technological operations crucial to fabricating the sellable asset in media (Ettema & Whitney 1994).

Information elicited by audience measurement operations is the basis for making informed economic decisions in the industry, and yet the standard view of information as a simple resource for decision making would seem to capture only a narrow view of data-driven dynamics in the industry. Given the fundamental role information has in economic decision-making, it is surprising that the concept is routinely assumed but indeed rarely explicitly defined. I asked for a definition from a friend with a recently completed PhD in microeconomics, who replied along the following lines: “What do you mean? It is like red. There is nothing to define.” Indeed, even an introductory text to the economics of information does not define the concept at all (Macho-Stadler & Pérez-Castrillo 2001). In the context of econometric modelling, information is for all practical purposes the kind of data available. McKinney and Yoos (2010) note that the concept is rather poorly defined even in the field of information systems. Static, taken-for-granted assumptions about information may serve analytical purposes well in the context of statistical models, but, as I hope to show in this paper, a broader perspective is needed to grasp the data-driven dynamics taking place in the media industry today. More specifically, contemporary developments in information and communication technology are disrupting the arrangements and practices under which the crucial relationship between the institutionally effective entity, that is, the measured audience, and the ultimately unknown actual audience has been traditionally established (Bermejo 2009; Napoli 2001, 2003; Precourt 2009).

This paper argues that the commercial media exists in a sort of double bind between evolving measurement technologies and audience fragmentation, a condition that conventional approaches to technology and audience measurement are not geared to analyse¹. The double bind of audience information is suggested to run from the condition that the organizational capability to generate facts about the audience is not independent of the behavioural patterns the measurement arrangements are supposed to capture. In other words, while striving for more complete information about the audience to enable more efficient decisions, the

¹ I am indebted to Jannis Kallinikos for bringing the idea of double bind to my attention.

economic actors unwittingly but systemically contribute to the decomposition of audiences. Over time, the phenomenon that is being measured changes by virtue of measurement making some organizational competencies, practices and processes obsolete, while calling for new ones. For instance, the traditional approach to audience measurement and pricing based on estimating the number of times the advertisement may have been seen by consumers is perceived in the computational media environment largely inferior to the pricing models associated with the number of actions triggered by the advertisement.

Metcalfe (2010) points out that mainstream economic theory has little to say about circumstances in which the boundaries between what is technologically possible, efficient and impossible are unclear or shifting, which is exactly what is happening in commercial media today. In order to shed light on these matters, it is necessary to first look more carefully into the role of technological information in the media industry. More precisely, the analytical interest is on how information generated out of computational data mediates the central measurement activity in the commercial media. I will start from the perhaps somewhat obscure idea of double bind.

The idea of double bind

The idea of double bind was introduced by Bateson and al. (1956) in a theory of schizophrenia. The term depicts a pattern of contradictory situations, in which an individual has to make a decision or to solve a problem by choosing between alternatives that will both result in unwanted outcomes. The recurrent experience of contradiction becomes then internalized into the individual psyche and predisposes his behaviour toward such situations. Obviously, transposing an old psychological concept into the analysis of institutional field is not a straightforward task; in particular, I do not intend substituting organizations for individuals in the context of such theory. If nothing else, however, the idea of double bind functions as a metaphor that offers a starting point for analysing how technological information shapes a certain institutional logic.

The argument to be explored in this paper can be summarized as follows. It is of utmost importance for any advertising-funded company to know its audience, yet the audience is always inherently elusive. A standard solution to this problem has been to introduce

progressively more sophisticated measurement technologies and analytical procedures to capture people's media consumption, but instead of disappearing the problem of pinning down coherent audiences seems to have paradoxically grown along with the measurement arrangements. Better measurement, while being the only feasible solution in the short term, would at the same time seem to contribute to the perpetuation of the very problem it is supposed solve. This suggests that perfectly reasonable decisions at the organizational level feed systemically back to the industrial condition against which those decisions take place. The metaphorical double bind between practical local solutions and the global dilemma can be understood in terms of technological information, that is, how audience measurement information generates the need for even more information.

In order to work through the argument, it is necessary to define the concept of information and its relationship with another central entity – data. Following Kallinikos (2006), information is understood here as a semantic event – an apprehension of novelty against a socially constructed background or a local setting. The position contrasts in an illuminating fashion with the token, syntax and representation views of information put forward by McKinney and Yoos (2010). In contrast to the aforementioned categories that would seem to espouse objectivist epistemological positions, information is considered here as observer-dependent and a categorically different kind of phenomenon from the data tokens. It is about being informed, which is always a temporally limited and agent-specific event; information cannot be frozen into a database or any other medium². From this perspective, the question of information enters the analysis in the sense of qualitatively different ways the actors are being informed instead of deviations from the ideal state of possessing complete information. In the taxonomy of McKinney and Yoos (2010), the position comes closest to the adaptation view that may look like a philosophical curiosity from the other three perspectives.

Data, in contrast, is a carrier of sign tokens that embody distinctions with potential to produce “difference that makes a difference” (Bateson 2000), that is, information. Simply put, character “0” is different from “1”, but whether that distinction stands out as relevant or not depends on the context. The example is simplistic and perhaps even naïve, but, as I will show later on, it has some striking resemblances with the empirical case. Digital data is of course merely one of the many possible bearers of signs whose signifying potential is nevertheless

² In a sense, information could perhaps be understood as a resource, but much more complex than is usually assumed.

partly defined by the operations a particular bearer enables. Information can be said to be medium specific (Lanzara 2009). Following from this, information produced by recourse to computational processes is referred here as technological information. The point in talking about technological information is thus to emphasise that, although information is necessarily an observer-dependent phenomenon, the operations by which the relevant distinctions are amassed and manipulated are specific to the medium on which the sign tokens exist³.

The rest of the paper is structured as follows. I will begin by explaining the industrial dilemma of audiencemaking business in three subsequent sections. First, the next section describes the basic operating logic of advertising-funded companies. Second, this discussion leads then to observing key historical patterns in the evolution of audience measurement arrangements that support the business of commercial media. Third, computational network infrastructure stands against this background as a potentially significant technological discontinuity. Finally, the rest of the paper analyses an attempt to innovate a new solution to the industrial dilemma, i.e. what does it take to generate information about the audience in the context of mobile communications systems.

The business of commercial media

Advertising is an intrinsic part of contemporary society. Its pervasive character in our everyday environment suggests that a whole range of products and services we take for granted rely on revenues from advertising. The basic institutional logic underpinning advertising-funded media can be understood by looking at the business model that captures the commercial prerequisites for sustaining the existence of media companies. All in all, the media industry is a typical example of platform business (Evans 2009). The companies create value by bringing consumers and advertisers come together in a coordinated fashion. A review of business model literature by Rajala (2009) identifies a number of definitions most of which incorporate a combination of two aspects into the business model: the mechanism for value creation and, consequently, its capture in a particular industrial setting. A commercial firm must produce something that others consider valuable and be able to monetize the value as a stream of revenues. How, then, are valuable assets created and exchanged by commercial media companies?

³ Indeed, it is quite a different matter to manipulate zeros and ones on a computer than on a paper.

The generic media business model entails that a company creates value by attracting consumers' attention to a space where it places advertisements paid by advertisers. Advertising-funded media companies can be understood as intermediaries operating on two-sided markets (Parker & Van Alstyne 2005). They subsidize content and services for consumers, whose valuable attention in the media space is then monetized by selling it to advertisers wishing to promote their products and services. Attention constitutes the scarce resource from which media companies manufacture audiences mapping specific lifestyles, consumption patterns and product categories. This generic business model has important implications on the structures of media industry and is closely associated with historical developments that exemplify the role of measurement and sophisticated analytical operations in the process of manufacturing the audience product. Also, from this perspective the communicative function of media (e.g. news, entertainment, user-generated content) is merely an intermediary input to the production of audiences.

The idea of selling media space to advertisers sounds deceptively simple. The complexity of media audiences is usefully theorized by Napoli (2003) who distinguishes between three dimensions of the audience product: the predicted, measured and the actual audience. First, advertising space is generally sold on the basis of *a priori* predictions on the amount and type of consumers attending the media content during a specific period of time. This is the predicted audience. Second, the final composition can, however, be known only after that period by making statistical inferences from a sample of consumers enrolled to the industrial measurement activities. It is this, the measured audience, which constitutes an institutionally effective entity with social meaning and economic value on the media industry (Ettema & Whitney 1994). Third, given the inherent shortcomings of any existing method for observing people's media consumption, the measured audience may always differ from the ultimately unknown actual audience. The substance of audience product is attention that, unlike many other types of industrial goods, is particularly difficult to quantify and to price. The head of research and development of NBC Universal, Alan Wurtzel (2009, p. 263) argued recently "you can't sell what you can't measure, and, unfortunately, our measurement systems are not keeping up with either technology or consumer behavior". The constitutive relationship from the actual to the measured audience is thus mediated by measurement arrangements as illustrated in Figure 1.

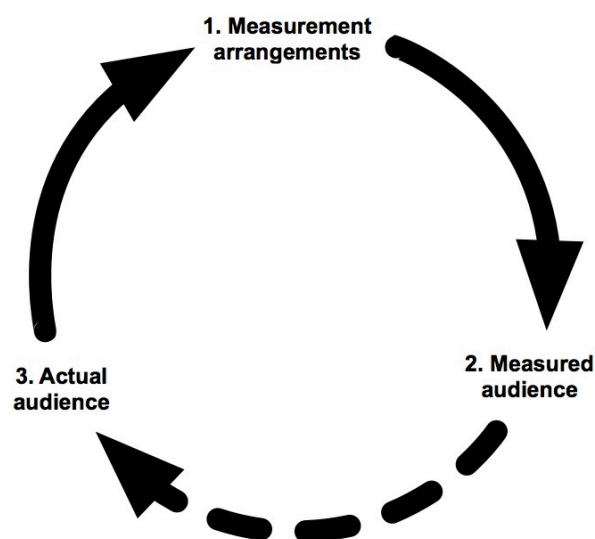


Figure 1. The audiencemaking cycle

The business model of commercial media can be executed in a number of ways, while Yoo, Henfridsson and Lyytinen (2010) point out that the changing product architectures can alter the dominant logic of organizing in the industry. For instance, the economic viability of platforms for user-generated content such as YouTube or Facebook is more often than not based on the capacity to capture the resulting consumer activity and attention as an advertising audience. This effectively outsources the content production to the advertising audience itself. In contrast to a traditional, vertically integrated content production function, the consumers become increasingly part of a disintegrated value chain by providing not only their attention but also the content of service.

The inherent difficulty of knowing the audience and industrial solutions to this problem have been an important factor in the evolution of media industry, or, as Napoli (2003, p. 83) puts it: “Changes in the techniques and technologies of audience measurement significantly affect the structure and behaviour of media organizations and media industries, as well as their relatively economic vitality.” Barnes and Thompson (1994, pp. 91-92) note “without that [sic] data, the audience has no reality for advertisers and, consequently, no value”, while Napoli (2003, p. 33) largely agrees that “the measured audience represents the central economic currency in the audience marketplace”. Recording, analysing and making predictions about the actual audience are central practices in fabricating the sellable asset for commercial media (Ettema & Whitney 1994). To elaborate this point, I will make a brief excursion into the

history of commercial audience measurement in the United States⁴. The next section describes the original problem solved by the industrial audience measurement, establishment of basic measurement practices and methodologies, progress toward automated data capture, emerging capability to break down the overall audience into consumer segments, and increasing granularity and the implications of gradually increasing sample size in measurement activities.

The history of commercial audience measurement

Early radio broadcasters in the beginning of 20th century had no evidence about their audiences, while newspapers and magazines could provide at least subscription data to substantiate claims about the size of their readership for the advertisers. The audience estimates remained largely unsystematic and sporadic based on station fan mail, occasional mail surveys or simply on population information within the reach of the transmitter until 1930s. The marketers had an interest to reach out for consumers and the broadcasters had an interest to sell airtime, but in order to facilitate market exchanges between the two, the electronic medium had to be able to come up with rigorous evidence on its audiences. Without the evidence, an investment in radio advertising could not be entered into comparative economic calculations and, thus, the advertisers would not be able to rationalize their efforts between different types of media. Developing solutions to this problem gradually brought the industrial ratings companies and audience measurement arrangements into being (Webster et al. 2006).

The first methodological attempts to measure the size of radio audiences took place in late 1920s and were soon turned into a systematic measurement of listenership making it possible for radio stations to compete with print media for advertising revenues. In 1930, a regular ratings service for national radio networks was launched based on telephone interviews asking sample respondents to recall what programs they had heard. The telephone recall method had, however, the problem of having to rely on people's memory, and few years later a competing service was launched based on an arguably more reliable telephone coincidental method inquiring what people were currently listening. The measurement of local radio moved later

⁴ Choosing to discuss the evolution of ratings industry in the United States is mainly out of convenience as the developments have been readily summarized in the literature (e.g. Barnes & Thomson 1994; Napoli 2003; Webster et al. 2006). The US represents also undoubtedly one of the most developed media markets.

on to a diary-based data collection by yet another company that started measuring local radio and television using diaries placed in people's homes in late 1940s. The benefit of the approach was that it could capture listening habits also out-of-home and during the hours the telephone coincidental was not possible. The method remained the basis for measuring local radio until recently. Many of the basic practices and concepts in audience measurement were established during this period including so-called syndicated studies compiled by the emerging ratings companies. These studies provide regular and comparable measurement across competing media outlets and are the basis for the currency of commercial media, the ratings points, in which advertising audiences are bought and sold still today. The better the rating, the more there is audience and, consequently, the higher the price of the advertising space.

The earliest mechanical meters able to continuously capture the use of media date back to late 1920s, but the technology was mainly developed in 1930s and put into a regular use by the end of 1940s. In 1938, a metering device known as audimeter was tested for recording the station a radio set was tuned in. A syndicated study known as Nielsen Radio Index was launched in 1942 based on a sample of 800 households equipped with the device. Cutting out the human element in data collection solved a number of problems related to the quality of data obtained through interviews and diaries but introduced other methodological problems – not the least the cost of maintaining the metering devices. Technicians had to initially visit the households where the meters were installed to obtain the data and to maintain the devices. Relying on dedicated devices has put a considerable premium on the approach even after the introduction of technical capability to read the meters remotely. In addition to the sample size limited by costs, the meters operated originally on the level of household without distinguishing who was actually listening to the radio set. In 1950, the approach was expanded to the emerging medium of television and provided the original basis for Nielsen Television Index driving many programming decisions today.

In 1960s, the decreasing price of digital computers made it feasible to use statistical approaches to break the overall audience into consumer segments prompting major advertising agencies to invest in computer equipment – long before the age of layout software and desktop publishing (Assael 2011). The pioneers of audience research had already made in 1920s attempts to capture the characteristics of consumers beyond the mere size of the audience, but it was not until the available computing power made it possible to analyse the

mass audience into progressively more detailed segments and thus making the idea of specialized media a commercially viable option. This contributed arguably to the eventual decline of media outlets aiming to serve as large audiences as possible (Barnes & Thomson 1994). 1980s saw again two significant methodological advances in audience measurement. First, faced by a competitive threat from a market entry by a foreign competitor, a more advanced metering technology known as peplemeters was introduced. These metering devices could capture not just the tuning of a television set but also information who was watching it increasing the granularity of audience measurement from the household level to the level of individual people. The second important development was the increase of sample size so that the measurement enabled the reliable estimation of narrowcast cable television audiences bringing them on par with national broadcasters in the eyes of advertisers.

The historical excursion is admittedly short and selective. Nevertheless, it shows the deeply analytical nature of audience product and allows highlighting a key institutional logic in the business of manufacturing advertising audiences. Increasingly sophisticated data collection and analytical arrangements have contributed crucially to making it economically feasible to set up outlets with narrowly defined audiences that offer advertisers the opportunity to rationalize their investments by targeting only those consumers that are believed to be likely to buy their products and to avoid wasting money by advertising to the others (Barnes & Thomson 1994; Iyer et al. 2005). The progressive calculability runs to a significant degree from the quality of data that has been possible to obtain from media consumption. Better measurement has, however, also contributed to the proliferation of outlets and channels fragmenting the media landscape to a point where both academics and practitioners view the idea of mass audience increasingly problematic (Croteau & Hoynes 2006). From the perspective of industry, fragmentation means considerable consumer freedom to choose how and what medium to use making the behaviour supposed to be captured by the audiencemaking arrangements ever more unpredictable. This, in turn, necessitates even more sophisticated measurement to hold the audience product together. Once there is a new outlet it can then become a common denominator for a group of people triggering perhaps behaviour and cognitive processes that did not exist prior to the media arrangements were in place. Odd as it may sound, the measurement arrangements may in some respect contribute to constructing the actual audience by looping effects between subjects and their interactive classifications (Hacking 2002, 2006; Khalidi 2010). It is this pattern of repeatedly solving the

central problem in commercial media in a manner that makes the problem itself progressively more complex that fits the pattern of metaphorical double bind.

Audiencemaking in the online environment

Contemporary information and communication technology does not abolish the basic institutional necessity to know the audience, yet it introduces a potentially significant technological discontinuity (Tushman & Anderson 1986) into the audiencemaking arrangements supporting commercial media companies. In contrast to television, radio and newspapers, the computational media environment would seem to shift the problem of measurement from obtaining data on media consumption to bringing together coherent audiences from minute or, so to say, pulverized behavioural traces left behind by the online population (Kallinikos 2006). There is no advertising media without an audience, since the audience is the very product it sells to the advertisers. The advances in information and communication technologies are part of the media evolution not only in terms of delivery and production of media content but also in shaping what kind of audiences are possible to measure into existence.

Ratings companies have tried to extend the metering device methodology for capturing the new kind of audience, but the approach based a software application installed on the personal computers of sample users faces some considerable problem in the new environment (Napoli 2003). On the one hand, the online environment has fragmented the media landscape to a degree that even with significantly increased sample sizes it is becoming impossible to provide reasonable estimates on the audience attending a particular outlet. While the number of television channels and radio stations available to consumers has also grown significantly, the number of outlets in traditional media is nevertheless relatively small compared to the vast number of websites and digital services accessible through the internet. On the other hand, the proliferation of devices and settings in which people access media content has made it increasingly difficult to observe the totality of an individual person's media consumption (Assael 2011). Finally, computational media environment is inherently interactive, which makes it difficult to conceive the measured behaviour as a one-way reception of messages (Spurgeon 2008).

In some cases the new actors have been able to altogether bypass the rating companies in terms of measuring up the audience. For instance, the lucrative search engine marketing has established a novel approach to audiencemaking that bypasses the ratings companies and generates significant revenues from small and medium size advertisers that have been traditionally less interesting customers to the media corporations. The interests of search engine users are inferred in real-time from the keywords typed into the search box instead of static profiling data. The audience effectively selects itself for the advertisements placed among the search results (Spurgeon 2008), and there is no need for the statistical estimates on how many people might have seen the advertisement. The transmission infrastructure generates automatically a trace of every page view (Sørensen et al. 2000) making it a matter of counting the entries in the server log to come up with the exact number of times somebody has had an opportunity to see the advertisement. Finally, the advertisements can be charged on the basis of actual clicks they triggered.

These observations may not spell the demise of ratings companies, whose syndicated studies across competing outlets maintain the currency that provides the grounds for comparing and pricing audiences bought by advertisers. Nevertheless, significant changes on the industry have historically often coincided with shifts in technological infrastructures. An important technological shift would seem to be that the computational transmission infrastructures, that is, the internet and mobile telecommunications networks, generate by default a trace of every click, call and message relayed through their components. While traditional electronic mediums required the ratings industry to develop various second-order technologies (diaries, metering devices etc.) to capture people's media consumption, the new environment embeds the monitoring function into the medium itself. Raw data emanating from computational network infrastructure is also qualitatively different from traditional measurements. Contrary to the carefully selected samples and variables of interest necessitated by the economics of second-order measurement techniques, the masses of digital data tokens are not tied to a particular organizational purpose, but are characteristically open, mostly affording aggregation, alignment and juxtaposition against each other.

Turning mobile network subscribers into an advertising audience

The arguments put forward in this paper developed out of an empirical fieldwork study that took place at an advertising-funded mobile virtual network operator (MVNO) company. In 2009, I spent three months as a participant observer at the company head office gathering data while working along the employees developing and managing the business operations. Systematic data collection resulted in 62 days of participatory observation log, 34 semi-structured interviews covering the entire staff, 26 press releases and 60 blog posts on the corporate website. In addition, I gathered hundreds of other documents as the opportunities arose during the fieldwork. The company had been incorporated in 2006 by high-profile founders who succeeded in raising millions of euros in venture capital to support the development and launch of an advertising-funded telecommunications operator. The lack of industrial standards and established business practices in mobile advertising meant that the organization had to engage in an active search in ways to turn mobile network subscribers into an audience product. This made the underpinnings of a new audiencemaking process readily observable.

The idea had been from the beginning to build a new kind of advertising medium that would be first launched and developed in one European country and then expanded to other markets. The company positioned itself both on the media and telecommunications industries whose actors were perceived to share a common interest in mobile advertising while generally unable to understand the workings of each other. This was perceived as an opportunity to innovate a business platform that would bring telecommunications operators, advertisers and consumers together in a new, more effective way (Evans 2009). Operating as an MVNO but making money from advertising the organization would “have the soul of media, but the body and muscles of a telecoms operator” as one of the informants put the nature of platform business. The consumers could sign up for the service by providing profiling information and opting in to receive marketing messages to their mobile phones. In exchange the company offered free minutes and text messages every month.

The following empirical exploration will identify the distinct qualities of computational data against the historical evolution of audience measurement arrangements and briefly discuss the types of analytical operations the organization employed to turn the data into information

about the audience. It is of particular interest how the digital data as *comprehensive, open* and *granular* behavioural observations shaped the audiencemaking practices revolving on the data emanating from the mobile network infrastructure. Finally, the analysis suggests certain implications on the audience product that match the informational double bind pattern. Let us start from the core of the suggested technological discontinuity, the nature of data tokens underpinning the audience measurement activity at the research site.

Computational data tokens

A digital communications network records a trace of every click, call and message relayed through its elements generating millions of records every day. The network elements keep a log of traffic for various purposes such as the optimal allocation of resources, detecting and recovering from malfunctions, and the identification of potentially harmful usage patterns. The automatic recording of network activity into so-called server logs is not just enabled by default, but disallowing it would usually make it difficult to maintain the network infrastructure itself. The microscopic records capture the time, type of action, and originating and terminating ends of the operation, and few technical details about the exchange, whereas the resulting entries carry no reference to the social setting, intentions and activities that triggered their recording. The following example serves to illustrate the type of behavioural records the computational network infrastructure generates⁵.

```
097369D2D7372762D31080000000000000001;1;33668741168;3322208;6;20081101004923;2
0081101004923;20081101004923
```

(Advenage SMS Gateway Router 1.0 documentation)

The individual records are strings of alphanumeric characters that carry little meaningful content by themselves. They are stored as separate lines in log files, and as such they are far removed from the idea of audience or any other object of business activity and practices. Obviously, no amount of raw data as such would convince advertisers that there is indeed a coherent audience at the other end of the medium; staring at the plain records is rather uninformative and does not tell anything about the audience. The basic availability of records contrasts nevertheless with the transmission infrastructures supporting radio and broadcast

⁵ While I am not allowed to reproduce an actual record from the research site, the example illustrates all the relevant aspects of the records.

television, which do not generally provide any information if there is actually somebody at the receiving end of the medium.

First, the embedded monitoring function automates much of the painstaking work of data collection that in traditional commercial media is associated with the arsenal of second-order measurement techniques needed to obtain observations about media consumption. Such techniques effectively limit recording to carefully planned samples geared to specific purposes, whereas the massive amount of data produced by the computational network infrastructure is not a sample but the census of activity in the network. Whatever the pattern derived from that data mass, it is possible to analyse the pattern in respect to a subsection of audience members without needing to know in advance (of the data capture) what kind of audience segmentation might be relevant. As a result, analytical operations are largely decoupled from the concerns driving the data collection.

Second, the computational network infrastructure does not only automate the rudimentary data collection but it also provides qualitatively different kind of behavioural traces as illustrated by the previous example. Even if the series of numbers and letters separated by delimiter characters in the example could be mapped to other entities such as phone numbers, time, type of action etc., the composite token is still tied to an ephemeral behavioural episode (e.g. a click or a message) and is as such rather uninteresting from the aggregate perspective of audiencemaking practices. The millions of digitally recorded behavioural traces are an incomprehensible heap of data tokens that are potentially amenable to a range of analytical operations. As such they are agnostic; the data does not answer any specific question in terms of audiencemaking business. The tokens exist as an open-ended potential to be explored in a variety of ways and to different ends.

Third, the data tokens dissolve people's media consumption into discrete clicks and messages from which a coherent audience has to be assembled by recourse to data-based operations. The monitoring function itself does not record any ordering scheme or usefully classify the behavioural observations from the perspective of company business model, but it merely stacks the records onto hard disks for later exploration. The extremely granular constitution of the data mass affords mainly analytic operations on the data tokens but at the same time renders the individual behavioural observations virtually meaningless.

Characteristic	Contrast with the traditional measurement arrangements
Comprehensiveness	The data tokens are not a sample but the census of activity in the network
Openness	The data tokens are not generated for a particular organizational purpose but as an open-ended potential
Granularity	The data tokens dissolve the observed behaviour at the level of discrete clicks and messages

Table 1. The distinctiveness of computational data tokens in audiencemaking

Table 1 summarize the three key attributes of technological information in audiencemaking business: comprehensiveness (not a sample), openness (not tied to a particular purpose), and granularity (breaks down the observed reality at the level of minute detail). In the historical perspective, the comprehensiveness and ease of data capture set the computational media environment apart from the traditional mass media relying on costly sampling-based approaches to measurement. The problem of manufacturing the advertising audience shifts from obtaining the data to analysing it; the latter operation is not any longer constrained by the sampling criteria set by the predefined analytical purposes and the economics of second-order technologies used to observe people's media consumption.

Analytical infrastructure

The previous section discussed three distinct attributes of computational data tokens generated by the network infrastructure in terms of audiencemaking business. Taken together, the attributes suggest that there was a lot of informative potential yet little actual information in the data. In order to illustrate the implications of these attributes, I will identify and briefly discuss different types of operations employed at the research site to turn computational data into information about the audience. Four kinds of operations can be distinguished within the analytical infrastructure that provided the visibility and made it possible to act upon the audience: *automatic data aggregation procedures, reporting software tools, organizational reporting practices, and custom analyses* – each encapsulating and building on top of previous mechanisms. The data tokens produced by the network were raw material for automated aggregation procedures encoded into various software components and reporting software applications associated with different organizational functions, whereas the employees did not usually access the tokens directly. The two approaches of automatic data aggregation and reporting software tools reflect, however, quite different starting points with

respect to how the data tokens are processed and assembled into meaningful categories and patterns serving organizational goals.

First, there were no industrially accepted metrics for measuring mobile messaging based advertising, and it was therefore possible to try shaping the rules of the evolving market until a certain way of measurement would settle as *de facto* standard in the eyes of advertisers. In this respect, it was necessary to remove the utter malleability of the data tokens in an often-painstaking process of developing metrics that could inform various organizational practices. The automatic aggregation procedures effectively locked down a specific formula for filtering and adding up data so that the result appeared as a fact-like entity for all practical purposes. Once in place, changing such a procedure was laborious. On the one hand, it would usually require modifications to the interconnected software systems supporting many organizational functions, while, on the other hand, it was vital for the credibility of metrics that they provided consistent and comparable results over time. The measurements along the key metrics such as the rate at which consumers responded to advertisements generated background assumptions for many organizational practices. The automatic aggregation procedures were black boxes in a sense that they provided a semantic closure on the data – with one important qualification. The data itself was not limited by preconceived sampling criteria. Despite their programmatic rigidity, it was possible to apply the automatic procedures on any subset of comprehensive records representing for instance an individual advertisement, time period, or a group of members.

Second, various reporting and analytical tools were prominent among the software systems used to support the operations. The systems enabled the handful of people working from a remote location to routinely stage the relevant aspects of distributed organization and its environment, and thus single out issues with the view of acting upon them. Contrary to the automatic data aggregation procedures, the logic of reporting software tools was to strategically harness the open and granular constitution of data. Indeed, the tools could be understood primarily as user interfaces for querying multidimensional data. They enabled aggregating, filtering and juxtaposing data tokens and presenting the results in tabular as well as in visual form, often encapsulating metrics produced by the automatic aggregation procedures. The operations made it possible to relatively freely explore and uncover potentially relevant patterns in the data that nevertheless ruled the kind of exploratory

opportunities the tool provided. The use of business intelligence and data mining applications is not uncommon in the telecommunications industry, but their scope was arguably rather extensive in the organization that was built from the ground up to rely on many of these systems. Many organizationally important objects were practically accessible only through the screen and these applications.

Third, the automatic data aggregation procedures and the routine use of reporting tools were the foundation for reporting in the organization. On top of these two rudimentary layers, a number of organizational reporting practices created a third layer in the analytical cascades making the audience present in the context of everyday work. The employees produced regular reporting to support their own functions as well as to keep the other teams informed; the analytical operations and reporting culminated regularly at the Monday morning office meetings where the senior management gave brief accounts of their respective functions. In such occasions, reporting was indeed thoroughly social endeavour with all kinds of interpretive contingencies. It would be difficult to deny that the selection, timing and the framing of facts brought into the discussions were perhaps just as important as the facts themselves. Without the computational foundations a great deal of reporting would have, however, been impossible or severely limited in scope and frequency. The underlying analytical layers and ultimately the computational data conditioned the matters that were channelled into PowerPoint presentations, Excel sheets and, finally, discussed in the meetings. The advantage of manual reporting practices in contrast to technological operations was that the employees were able to selectively combine data from a variety of sources and contextualize it with tactics guiding the interpretation to address specific issues. In this respect, there were often a lot of discussion on what did a specific measurement mean for the task at hand; which numbers should be shown on a particular occasion or material. Without the data, metrics and reporting software tools such discussion would have, however, made little sense, as there would have been nothing to interpret.

Fourth, in addition to aforementioned three kinds of regular operations, there was a type of analytical task that was executed more irregularly but constituted nevertheless an important part of analytical activities. These were the manually crafted custom analyses harnessing the available sources of digital data. Often once-off or rarely needed analyses were tailored to answer specific questions emerging in the context of various operations and activities.

Apparent problems in the network infrastructure, inexplicable audience behaviour seen in the reports, or the needs of business development activities could motivate a novel cut into the data to shed light on the particular matter at hand. In this respect, the data tokens existed as a potential that attracted practical attention not unlike the case of reporting software tools. It was always possible to try looking for more information from the data. Even if these analyses were somewhat laborious, the masses of open and granular data tokens meant that there was usually no need to reach out to obtain new data. This effectively limited the required effort into the number of working hours and made the undertaking into a relatively straightforward matter that could be carried out by one or two persons at the office.

The previous section opened up and identified the technological discontinuity in audiencemaking business in term of comprehensiveness, openness and granularity of computational data, which shifts the problem of audience measurement from obtaining observations about consumer behaviour to summoning the audience out of data. Contrary to crafting surveys, metering devices or other sampling based instruments for gathering data about the audience, the organizational problem was to craft the audience using extant behavioural records. The monitoring function embedded into the network infrastructure did not impose any meaningful order or classification onto the records, and therefore the data tokens had to go through cascades of operations on their way out from the database toward practical applicability in the context of organizational practices. Since there was no inherent order or meaning in the data, it had to be created at the time of use. Weinberger (2007) calls this the third order ordering. Table 2 provides a brief summary of the observed implications of computational attributes at the different levels of analytical infrastructure.

Level	Comprehensiveness	Openness	Granularity
Automatic data aggregation procedures	<i>Removes</i> the need to know relevant segmentations and groupings prior to the data collection	<i>Requires</i> meticulously pinning down formulas for achieving closure on specific metrics	<i>Requires</i> dealing with microscopic idiosyncrasies captured by the records
Reporting software tools	<i>Removes</i> the need to know relevant segmentations and groupings prior to the data collection	<i>Enables</i> the exploration for unexpected behavioural patterns	<i>Enables</i> focusing on minuscule behavioural details
Organizational reporting practices	<i>Necessitates</i> the two rudimentary layers in order to turn the data tokens into meaningful information	<i>Enables</i> mashing up measures from various sources and framing reports with interpretive tactics	<i>Necessitates</i> the two rudimentary layers in order to turn the data tokens into meaningful information
Custom analyses	<i>Removes</i> the cost of reaching out to consumers for new data needed for the analysis	<i>Requires</i> imagining hypotheses driving the custom analyses	<i>Enables</i> focusing on minuscule behavioural details

Table 2. *The implications of technological discontinuity at the different levels of analytical infrastructure*

Table 2 includes nine different ways in which the three computational attributes shaped the analytical operations crucial to the audiencemaking business. These can be understood both as regulative and as constitutive rules specific to the computational media environment (Hildebrandt 2008). In other words, technological information both enabled and constrained the analytical operations stretching from the hard-coded procedures to the social practices in qualitatively distinct ways. The transition from using dedicated metering devices to the common information infrastructure removed some previous restrictions and enabled new kinds of operations on the data. However, it also required certain new conditions to be met and necessitated building sophisticated analytical tools to harness the informative potential of computational data in the context of everyday work practices. The four different predicates (*removes*, *requires*, *enables*, *necessitates*) and the short summaries in the table cells summarize the medium-specificity of technological information in the current case.

Reading the table by columns reveals that the implications of three attributes vary from the level to another and are understood best vis-à-vis the particular type of analytical operation. Comprehensiveness removed the need to know the relevant consumer segments and classificatory variables in advance and did away with the need to reach out to consumers in

order to gather data for custom analytical purposes. However, together with granularity, it also necessitated operating the two rudimentary layers in order to make the behavioural observations useful in the context of everyday work practices. Openness required locking down rigid formulas to manufacture facts as well as generating hypotheses driving the custom analyses. It also enabled exploring hidden patterns in the data and mashing up measurements from various sources. Granularity enabled focusing upon minuscule behavioural details but required also dealing with microscopic behavioural idiosyncrasies.

Reading the table by rows suggests how the three attributes interacted on each level of the analytical infrastructure. Automatic data aggregation procedures were not held back by preconceived segmentation criteria but had to also deal with the openness and irrelevant behavioural details captured by the data. The use of software reporting tools would seem to benefit considerably from the decoupling of data from preconceived sampling criteria and the possibility to explore unexpected behavioural patterns in detail. Organizational reporting practices were largely predicated on the smooth functioning of two rudimentary layers while enabling mashing up and interpreting data from various sources. Finally, custom analyses became much more economic due to readily available computational data but required also envisioning questions that could be posed to the data.

Discussion

This paper argued that the usual assumptions about technological information and data as representations of external states of the world are inadequate to analyse the dynamics of information on the media industry. In order to tackle this issue, I made a simple conceptual distinction between information as an event and computational data as one the possible bearers of potentially informative distinctions. The role of measurement was then theorized as a sort of informational double bind, and the argument was substantiated by making an excursion into the evolution of audience measurement arrangements. Against this background, the empirical part of paper discussed the computational media environment in terms of comprehensiveness, openness and granularity of audience measurement data. The implications of these attributes were illustrated with respect to four types of operations needed to generate meaningful information about the audience.

From a purely economic perspective, the possibility to monitor and thus to condition increasingly microscopic behaviour by contractual means should, in principle, promote efficiency and optimal allocation of resources (Varian 2010). Any such gain may, however, be in practice offset not only by social repercussions from the increasing intensity of advertising but also by mechanisms such as the double bind of audience information. Historically, the more audiences have been measured the more audiences have appeared to require measurement. More generally, the case can be understood in terms of expanding scope for the statistical modelling of human behaviour. Since data-based information is all that exists from the perspective of models, the evolution of audience measurement arrangements effectively brings new things and events into the domain of modelling. In some sense this is rather obvious and underpins the idea of computational social science (Lazer et al. 2009) as well as recent excitement around business analytics and data mining (e.g. Ayres 2007; Davenport et al. 2005; Lavallo et al. 2010; Nedelcu 2009; Redman 2008). While largely agreeing with such trends, the case suggests, however, that no amount of modelling may be able to arrest the runaway dynamic of measurement information.

At the level of organizational practices and arrangements, the medium-specificity of technological information is clearly not captured merely as straightforward automation and increasing efficiency, and indeed the extremely granular records require perhaps more than ever human interpretation and intervention to be of any use. However, the point is that the computational data is not just malleable (cf. Orlikowski 2000, p. 406), but it forced the organization to cope with this malleability in new ways. Furthermore, the constitutive relationship between the measured and the actual audience was essentially interactive as the result of direct and immediate feedback received from the audience behaviour. The operations targeting the audience members fed immediately back into the same system that was used to plan and execute the messaging. The expectation of feedback could thus become factored into the planning of operations effectively treating the member as an interactive entity. Knowing the audience meant therefore not only pinning its members down in terms of static profiles but also learning and managing its behavioural reactions to different interventions. This brings the discussion back to Figure 2 depicting the loop from the measurement arrangements to the measured audience, and through the actual audience back to the measurement arrangements.

The recursive character of cycle suggests that the audience does not only derive from the natural groupings of people, but the audiencemaking practices can bring into existence a new common denominator between people. In Figure 2 this is reflected by the enumeration of phases starting from the measurement arrangements instead of the actual audience. Without the measurement arrangements there can be no measured audience, and hence any actual audience would be irrelevant from the perspective of media business model, whereas the measurement arrangements are, in principle, able to shape the actual audience. In order to function properly, the audiencemaking loop needs all three elements, but the measurement arrangements would seem a more appropriate starting point than the actual audience. Finally, understanding the audience as a primarily interactive entity opens up a whole new domain of measurement beyond static profiling and, indeed, reproducing the double bind.

Concluding remarks

The institutional ruptures in the audiencemaking arrangement assessed in the study may, after all, turn out to be temporary and over time become accommodated into the existing industrial arrangements. Even so, the crossover period will provide a good opportunity to study the techno-economic underpinnings of commercial media in general and the suggested double bind of information in particular. Regardless what we think about the content of advertisements, advertising is an important mechanism in regulating consumption as well as supporting journalistic production in democratic market societies, and thus worth understanding on its own right. Information and data are vital entities in the media industry, yet arguably not well understood.

References

- Assael, H., 2011. From Silos to Synergy: A Fifty-year Review of Cross-media Research Shows Synergy Has Yet to Achieve its Full Potential. *Journal of Advertising Research*, 51(1, 50th Anniversary Supplement), pp. 42-48.
- Ayres, I., 2007. *Supercrunchers: Why Thinking-by-Numbers Is the New Way to Be Smart*, New York: Bantam Dell.
- Barnes, B.E. & Thomson, L.M., 1994. Power to the People (Meter): Audience Measurement Technology and Media Specialization. In J. S. Ettema & D. C. Whitney, eds. *Audiencemaking: How the Media Create the Audience*. Sage Annual Reviews of Communication Research. Thousand Oaks: SAGE Publications, pp. 75-94.

- Bateson, G., 2000. *Steps to an Ecology of Mind*, Chicago: University of Chicago Press.
- Bateson, G. et al., 1956. Toward a Theory of Schizophrenia. *Behavioral Science*, 1(4), pp. 251-264.
- Bermejo, F., 2009. Audience manufacture in historical perspective: from broadcasting to Google. *New Media & Society*, 11(1&2), pp. 133-154.
- Brown, J.S. & Duguid, P., 2000. *The Social Life of Information*, Boston: Harvard Business School Press.
- Croteau, D. & Hoynes, W., 2006. *The Business of Media: Corporate Media and the Public Interest* 2nd ed., Thousand Oaks: Pine Forge Press.
- Davenport, T.H., Cohen, D. & Jakobson, A., 2005. *Competing on Analytics*, Babson Park: Babson Executive Education.
- Ettema, J.S. & Whitney, D.C., 1994. The Money Arrow: An Introduction to Audiencemaking. In J. S. Ettema & D. C. Whitney, eds. *Audiencemaking: How the Media Create the Audience*. Sage Annual Reviews of Communication Research. Thousand Oaks: SAGE Publications, pp. 1-18.
- Evans, D.S., 2009. How catalysts ignite: the economics of platform-based start-ups. In A. Gawer, ed. *Platforms, Markets and Innovation*. Cheltenham: Edward Elgar, pp. 99-128.
- Hacking, I., 2002. *Historical Ontology*, Cambridge, Massachusetts: Harvard University Press.
- Hacking, I., 2006. Making Up People. *London Review of Books*, 28(16).
- Hildebrandt, M., 2008. Legal and technological normativity: more (and less) than twin sisters. *Techné: Research in Philosophy and Technology*, 12(3), pp. 169-183.
- Hoskins, C., McFayden, S. & Finn, A., 2004. *Media Economics: Applying Economics to New and Traditional Media*, Thousand Oaks: SAGE Publications.
- Iyer, G., Soberman, D. & Villa-Boas, J.M., 2005. The Targeting of Advertising. *Marketing Science*, 24(3), pp. 461-476.
- Kallinikos, J., 2006. *The Consequences of Information: Institutional Implications of Technological Change*, Cheltenham: Edward Elgar.
- Khalidi, M.A., 2010. Interactive Kinds. *British Journal of Philosophy of Science*, 61, pp. 335-360.
- Küng, L., 2008. *Strategic Management in the Media: From Theory to Practice*, Los Angeles: SAGE Publications.
- Lanzara, G.F., 2009. Reshaping Practice Across Media: Material Mediation, Medium Specificity and Practical Knowledge in Judicial Work. *Organization Studies*, 30(12), pp. 1369-1390.

- Lavalle, S. et al., 2010. *Analytics: The New Path to Value*, Massachusetts Institute of Technology.
- Lazer, D. et al., 2009. Computational Social Science. *Science*, 323(5915), pp. 721-723.
- Macho-Stadler, I. & Pérez-Castrillo, J.D., 2001. *An Introduction to the Economics of Information* Second Edition., Oxford: Oxford University Press.
- McKinney, E.H.J. & Yoos, C.J.I., 2010. Information About Information: A Taxonomy of Views. *MIS Quarterly*, 34(2), pp. 329-344.
- Metcalfe, J.S., 2010. Technology and economic theory. *Cambridge Journal of Economics*, 34(1), pp. 153-171.
- Napoli, P.M., 2003. *Audience Economics: Media Institutions and the Audience Marketplace*, New York: Columbia University Press.
- Napoli, P.M., 2001. The Audience Product and the New Media Environment: Implications for the Economics of Media Industries. *The International Journal of Media Management*, 3(2), pp. 66-73.
- Nedelcu, V., 2009. Data abundance – how to make intelligent use of it. *Ericsson Business Review*, 2009(2), pp. 25-27.
- Orlikowski, W.J., 2000. Using Technology and Constituting Structures: A Practice Lens for Studying Technology in Organizations. *Organization Science*, 11(4), pp. 404-428.
- Parker, G.G. & Van Alstyne, M.W., 2005. Two-Sided Network Effects: A Theory of Information Product Design. *Management Science*, 51(10), pp. 1494-1504.
- Precourt, G., 2009. Getting Metrics Right. *Journal of Advertising Research*, 49(4), pp. 395-396.
- Rajala, R., 2009. *Determinants of Business Model Performance in Software Firms*. PhD thesis. Helsinki School of Economics. Available at: <http://hsepubl.lib.hse.fi/pdf/diss/a357.pdf>.
- Redman, T.C., 2008. *Data Driven*, Boston: Harvard Business Press.
- Sørensen, C., Fagrell, H. & Ljungstrand, P., 2000. Traces. In K. Braa, C. Sørensen, & B. Dahlbom, eds. *Planet Internet*. Lund: Studentlitteratur, pp. 157-183.
- Spurgeon, C., 2008. *Advertising and New Media*, London: Routledge.
- Tushman, M.L. & Anderson, P., 1986. Technological Discontinuities and Organizational Environments. *Administrative Science Quarterly*, 31(3), pp. 439-465.
- Varian, H.R., 2010. Computer Mediated Transactions. *American Economic Review*, 100(2), pp. 1-10.
- Webster, J.G., Phalen, P.F. & Lichty, L.W., 2006. *Ratings Analysis: The Theory and Practice of Audience Research* 3rd ed., Mahwah: Lawrence Erlbaum Associates.

- Weinberger, D., 2007. *Everything Is Miscellaneous: The Power of the New Digital Disorder*, New York: Times Books.
- Wurtzel, A., 2009. Viewpoint: Now Or Never - An Urgent Call to Action for Consensus on New Media Metrics. *Journal of Advertising Research*, 49(3), pp. 263-265.
- Yoo, Y., Henfridsson, O. & Lyytinen, K., 2010. The New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research. *Information Systems Research*, 21(4), pp. 724-735.