Alma Mater Studiorum - Universit di Bologna

Dottorato di Ricerca in Fisica
Ciclo XXVI

Settore Concorsuale di afferenza: 02/B2
Settore Scientifico disciplinare: FIS/02

Novel investigation methods in Computational Social Dynamics and Complex Systems

Presentato da: **Pietro Gravino** (612855)

Coordinatore Dottorato
Fabio Ortolani

Relatore
Armando Bazzani

Relatori Esterni
Vittorio Loreto
Vito D. P. Servedio

Esame finale anno 2015

"If I have seen further it is by standing on the shoulders of giants."
Sir Isaac Newton

To my family, my giants:
Carmela, Salvatore,
Elena and Vito.

Contents

Chapter 1

Introduction

On 10 September 2008 the greatest machine ever build by mankind finally became operational. It took years of huge economic and engineering efforts to complete it. A total estimated cost of 3 billion euros, the collaboration of thousands of scientists from tens of countries and hundred of universities and research centers, a titanic, unprecedented effort of cooperation coordinated by the European Organization for Nuclear Research (CERN) but finally the famous Large Hadron Collider (LHC), started to fire protons in its 27 km long tunnel beneath the Franco-Swiss border near Geneva, Switzerland [9]. In the following years, while becoming more and more operational, LHC broke several records. In November 2009 it became the world's highest-energy particle accelerator [14]. In April 2011 became the world's highest-luminosity hadron accelerator [15]. And to these growing performances corresponded an increasing set of novel data from which new discoveries emerged, like the $\chi_b(3P)$ [21] or the famous Higgs Boson observation [12], and new questions arose. The yearly data flow has been increasing with the improvement of the operationality: the dataset size was estimated around 13 petabytes ($10^{15}\,bit$) in 2010, which would fill a stack of CDs around 14 km high, and 27 petabytes in 2014 [35]. This humongous quantity of data will grow even more after the planned reboot and upgrade of the facility. It has been foreseen that data output will soar to 110 petabytes a year by the early 2020s, and, eventually, to as much as 400 petabytes a year [41]. Even if these numbers are quite spectacular, the importance of the numeric value becomes far less impressive if we look at them in the perspective of our contemporary culture.

1.1 The Information Age of Technosocial systems

According to some historians, our current historical period can be defined as the Information Age. This period, beginning around 1970, as a distinctive feature shows an abundant publication, consumption, and manipulation of information, especially by computers and computer networks. In particular, since the spreading and the overcoming of the Internet (in which, again, the CERN had a seminal role) the volume of information exchange online has reached formerly inconceivable levels, and it is continuously growing. The world's effective capacity to exchange information through two-way telecommunication networks was 281 petabytes of information in 1986, 471 petabytes in 1993, 2200 petabytes in 2000, and 65000 petabytes in 2007 (this is the informational equivalent to every person exchanging 6 newspapers per day) [72]. Web services and information aggregator handle massive data flows:

- in 2009, the Google data centers process about 24 petabytes of data per day, on average [51];
- in August 2012, Facebook's clusters include the largest single HDFS cluster known, with more than 100 PB physical disk space [17];
- as of January 2013, Facebook users had uploaded over 240 billion photos, with 350 million new photos every day. For each uploaded photo, Facebook generates and stores four images of different sizes, which translated to a total of 960 billion images and an estimated 357 petabytes of storage [19];
- in last days of 2014, the Steam, a digital gaming service developed by Valve, delivers over 20 petabyte per day [16].

This information flows reflect the change in our societies which have merged with technological systems forming an unprecedented evolution in human history. Our very DNA as a society has irreversibly muted with consequences that will probably affect the whole mankind in its techno-social globalized system. The economic crisis of 2007 may be the most famous and historically well-studied example [11]. Another remarkable example are the worldwide virus spreading, or the large scale riots and revolutionary waves, like in the case of Arab Spring [10]. And those are just the most evident and sudden phenomena. Other processes are happening on different timescale, like the climate change, which is slowly causing a shift in people environment perception and a consequent change in their behavior. Some process, instead, take place on different spacial scale, like those happening in our town which have to face new urbanization levels and all the consequent challenges. Also, languages are evolving in ways never seen before, because of the new communication channels, influencing the way we define the world. The way we think itself is influenced by this boosted availability of information. And if some phenomena are generated intrinsically by the connections possibility enhancements, others are willingly pushed by some entities, like commercial companies or political parties, for economic or politic reason: what are marketing and electoral campaigns if not an attempt to influence our techno-social systems? It is needles to say that the understanding of this techno-social system and of phenomena taking place in its various layers acquires a crucial importance.

1.2 Information is not knowledge

An opportunity is offered by the above cited data, which represent a useful trail left by the system in its evolution. But those data, differently from the data generated by LHC, are not meant for understanding. LHC data are the result of a long and complex activity of design intending to accurately investigate specific events in a scientific way. LHC itself can be thought as one of the acme of the scientific method, as the last step in a path centuries-long, as the most advanced and sophisticated way we have to look at nature: the mankind greatest experiment. In comparison, data spilling out from techno-social system are more or less garbage. Still, this garbage has been, so far, the main way to study the otherwise almost completely unknown system. Looking at this garbage is undoubtably the first step in this novel path of scientific comprehension. And if we look at it, the first characteristic arising it is what we already pointed out: the huge quantity of information. So, the first methodological question may be: how can we understand something from this mess? Another impressive number, to be attached to the others already cited, is 295 exabytes, i.e. $2.95 \cdot 10^{20}$ bits which is the estimated cumulative storage ability of the global techno-social system in 2007 [72]. It is quite ironic, but never, in its history, mankind had that much information about something so little known. Looking from wrong perspective, to deal with those data may seem impossible. But this incredible number is more or less 4 order of magnitude smaller of the number of water molecule in a glass of water. Physics developed, in its history, a vast set of intellectual tools to deal with thermodynamic systems (like the glass of water). It is common knowledge that the first step to find the right answer is to choose the right question. In this case, the first main question may be about the identity of the global meaningful quantities describing the system and its dynamics. And physicists have proved to be good at this sort of task. And it is not just a matter of statistics. Also modeling abilities, required to overcome the mere description, are a characteristic typically belonging to this community. The strategic value of physics intellectual tools applied to, but not only to, techno-social systems give rise to a new research field often referred to as a physics of complex systems. More in general, we can affirm that natural science instruments can give a strong contribution, joining the historic effort of social sciences in the unveiling of the complexity of the new challenges.

1.3 An unexpected help

Some successful example are already there in unsuspected areas. The study of dynamics of crowd disasters made pilgrimage in La Mecca safer [29]. The application of mathematical modeling to the spread of epidemics give a fundamental help in forecast and control [45]. Theoretical modeling has been used for traffic monitoring and control in various country [18, 99]. And data mining is only a step in this scientific discovery path. The following crucial and mandatory step is the implementation of experiments. It is way over the purposes of this thesis, and it is probably unnecessary, the explanation of the epistemological, before that practical, value of the experiments for a scientific path. Experiments already have been successfully implemented in the history of social sciences but now they can be taken to another level. In fact, the techno-social system give us not only great occasion for data mining but also new opportunities of interaction with the system itself, which means new experimental opportunities. And exploiting the ICT tools to interact with wide communities to achieve useful purpose thanks to individual contribution is the essence of the Crowd Sourcing, an hot topic emerged in recent years.

1.4 Web Experiments

This methodology, in some cases referred to also as Social Computation, can be used also for scientific purposes, and it actually has been. A notable example is Galaxy Zoo[1] [116], in which players are asked to classify astronomic objects of galactic type, by browsing a catalogue of telescopic images. Another example can be reported about the 3D structure of viral enzymes that challenged scientists for years and has been discovered thanks to the efforts of Foldit[2] players [82]. Human ability to easily solve tasks that are difficult to solve by machines has been largely exploited for instance in labelling images, through the collaborative *ESP Game* [132]. In the last two example, the idea of linking playful activities with useful outcomes has led to the paradigm of Games With a Purpose (GWAP) [131], i.e. a way of engaging people in games that can extract valuable information or work as a side effect of the game. The playful rearranging of experiments, together with their appealing graphic interfaces, has also proved to be a fundamental ingredient for web-based experiments design, boosting user participation and data reliability. Even if this approach is relatively young, his tremendous potential attracted a lot of interest and a lot of other notable example are appearing all around the scientific community. So far, anyway, these separated efforts does not show any cumulative growth in methodologies, neither a standardization of tools. These features are strategic focal points for this approach to express his real potential. In other words, the next step of the evolution that take from data mining to web experiments should take from web experiments to web laboratories.

1.5 Thesis abstract

In this thesis the evolution of the techno-social systems analysis methods will be reported, through the explanation of the various research experience directly faced. The first case presented is a research based on data mining of a dataset of words association named Human Brain Cloud: validation will be faced and, also through a non-trivial modeling, a better understanding of language properties will be presented. Then, a real complex system experiment will be introduced: the WideNoise experiment in the context of the

[1] `http://www.galaxyzoo.org`
[2] `http://fold.it`

EveryAware european project. The project and the experiment course will be illustrated and data analysis will be displayed. Then the Experimental Tribe platform for social computation will be introduced . It has been conceived to help researchers in the implementation of web experiments, and aims also to catalyze the cumulative growth of experimental methodologies and the standardization of tools cited above. In the last part, three other research experience which already took place on the Experimental Tribe platform will be discussed in detail, from the design of the experiment to the analysis of the results and, eventually, to the modeling of the systems involved. The experiments are: CityRace, about the measurement of human traffic-facing strategies; laPENSOcosí, aiming to unveil the political opinion structure; AirProbe, implemented again in the EveryAware project framework, which consisted in monitoring air quality opinion shift of a community informed about local air pollution.

At the end, the evolution of the technosocial systems investigation methods shall emerge together with the opportunities and the threats offered by this new scientific path.

Chapter 2

An unplanned web experiment: Human Brain Cloud

In this chapter we introduce a dataset gathered through the massive multiplayer web-based Word Association Game known as *Human Brain Cloud*[1] (HBC in the following). We used this dataset to investigate the directed and weighted complex network of free word associations. For validation and further analysis we matched HBC with two other notable datasets: the *South Florida Free Association Norms* dataset [100] (SF in the following). The network of associations exhibits quite robust properties like the small world property, a slight assortativity and a strong asymmetry between in-degree and out-degree distributions. A particularly interesting result concerns the existence of a typical scale for the word association process, arguably related to specific conceptual contexts for each word. We also point out the basic cognitive mechanisms underlying word associations when they are represented as paths in an underlying semantic network. We derive in particular an expression describing the growth of the HBC graph and we highlight the existence of a typical scale for the word association process.

This work will be linked to an analysis of the complex organization of syntax in hierarchical structures, which is one of the core design features of human language. Duality of patterning refers for instance to the organization of the meaningful elements in a language at two distinct levels: a combinatorial level where meaningless forms are combined into meaningful forms and a compositional level where meaningful forms are composed into larger lex-

[1] http://www.humanbraincloud.com/

ical units. The question remains wide open regarding how such structures could have emerged. The aim is that of addressing these two aspects in a self-consistent way. Thus, we introduce suitable measures to quantify the level of *combinatoriality* and *compositionality* in a language, and present a framework to estimate these observables in human natural languages.

This chapter, beside its scientific value, provides two important proof of concept. In first place, it is shown that tools from Physics of Complex Systems may be successfully adopted to face social science problems, actually helping the achieving of a deeper comprehension of human language. Also, we also used data gathered by users through the Internet in a highly participate (even if unintended) web-experiment, proving the reliability and the potential of this experimental method.

2.1 A different kind of experiment

Word associations have been experimentally studied in details, especially by linguists and cognitive scientists [44, 100]. An interesting point of view, not yet fully explored, is to look at the ensemble of words and associations as a complex network, where nodes are words and links are associations, and to analyze it as such [95,122]. But previous works performed the typical analysis of the complex networks' framework on words association networks obtained from scientific experiment performed by researchers in extremely controlled environments. This makes data reliable but at a quite high cost. Typical word association experiments involve a relatively limited number of subjects (100 - 1,000), in controlled conditions. A word (called cue word) is presented to the recruited subjects, who are asked to write the first related word that come to their minds (called target word). Due to high costs in terms of time and money, the number of cue-target associations gathered in these classical experiments has been relatively limited, at maximum of the order of 10^5 associations. This is the order of magnitude of the dataset gathered during the most important word association experiment, the Nelson *et al.* "University of South Florida Free Association Norms" database (SF) [100]. SF is the outcome of a great effort started back in 1973 and lasted almost thirty years. It consists of $5,000$ words and $700,000$ associations.

The experimental data we analyze were obtained from the "massively multiplayer word association game" *Human Brain Cloud* (HBC)[2]. HBC was

[2]Dataset courtesy of Kyle Gabler [62]

designed as a web-based game in English language that simply proposed a cue word to the player, asking for a target word (e.g., volcano-lava, house-roof, dog-cat, etc.). The cue was each time taken independently and randomly with uniform probability from an internal self-consistent set of words, constructed by gathering the answered target words at the end of each game session. With respect to usual experiments, no control is performed on the number of players, the number of cue-target associations given by each player, etc... On the other hand, the obtained dataset is considerably larger than that of previous experiments, and consists of approximately $600,000$ words and $7,000,000$ associations, gathered within a period of one year (while pre-existing experiments involved teams of specialists for much longer periods of time). As each player could enter whichever word or set of words, the data contain a certain volume of inconsistent words, which have to be discarded. After a suitable filtering procedure, which we describe in the appendix of [67], we obtain a strongly connected directed weighted graph with almost $90,000$ words and $6,000,000$ associations, i.e., a dataset still considerably larger than those of previous experiments. A snapshot of a part of the HBC graph is shown in Fig. 2.1.

2.2 Size is not everything

Even if the size reached by the HBC dataset in such a short time may sound impressive, its scientific value is far to be demonstrated. To validate HBC dataset and to detect possible systematic biases, we also compared some properties of the HBC dataset with the SF dataset already introduced. Each dataset yields a graph whose nodes are the words and whose edges correspond to the associations between words made by the players/subjects. Moreover, each edge is weighted by the number of times that the corresponding association has been made. The first result of the comparison is that almost all (99%) the words used as cues in SF were present in HBC and 72% of the SF associations were present in the HBC dataset. Then we consider for each word w_i the set of associations in which w_i was proposed as a cue word to players, and define the out-strength $s_{out}(w_i)$ of the corresponding node as the total number of associations in which w_i was the cue, and the out-degree $k_{out}(w_i)$ as the number of *distinct* target words answered in response to w_i. From these definitions follow clearly those of $s_{in}(w_i)$ and $k_{in}(w_i)$.

Fig. 2.2 shows the distribution of in- and out-degrees in both HBC and

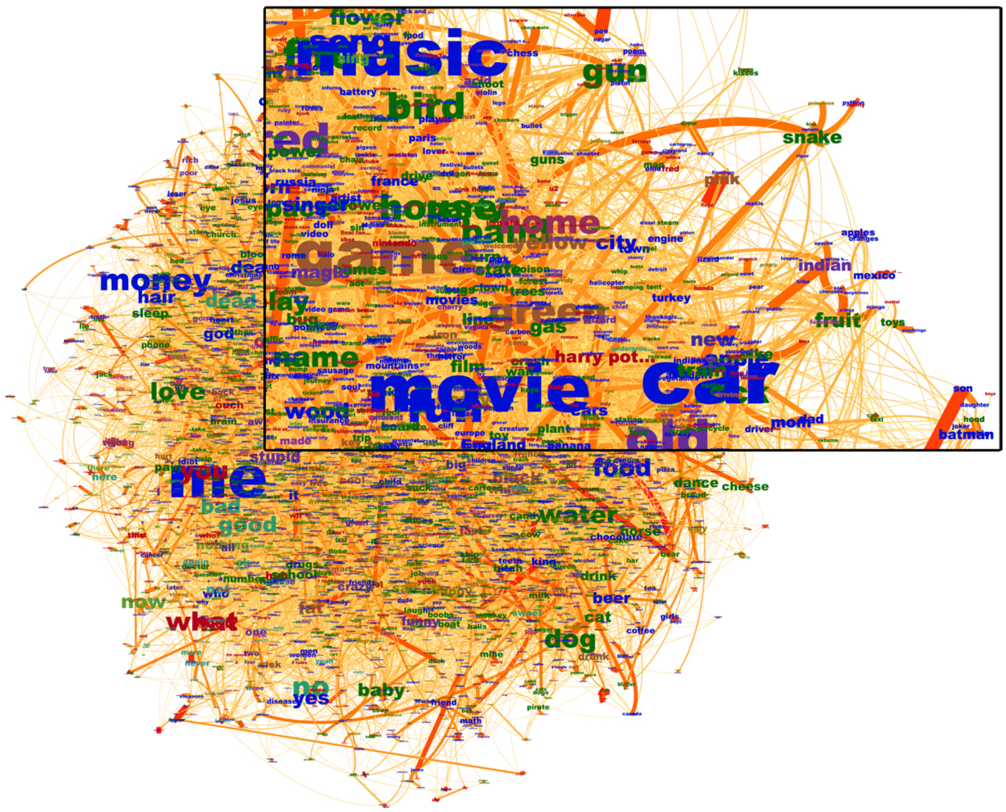


Figure 2.1: Graphical representation of the word association network obtained from the HBC dataset. Labeled nodes represent the first 5,000 words entered in the system while arcs represent word associations. Label colors distinguish different parts of speech or combination of parts of speech (adjectives, nouns, verbs, nouns-verbs, etc). Arcs' width codes the weight of the corresponding association, i.e., how many people ever made that association.

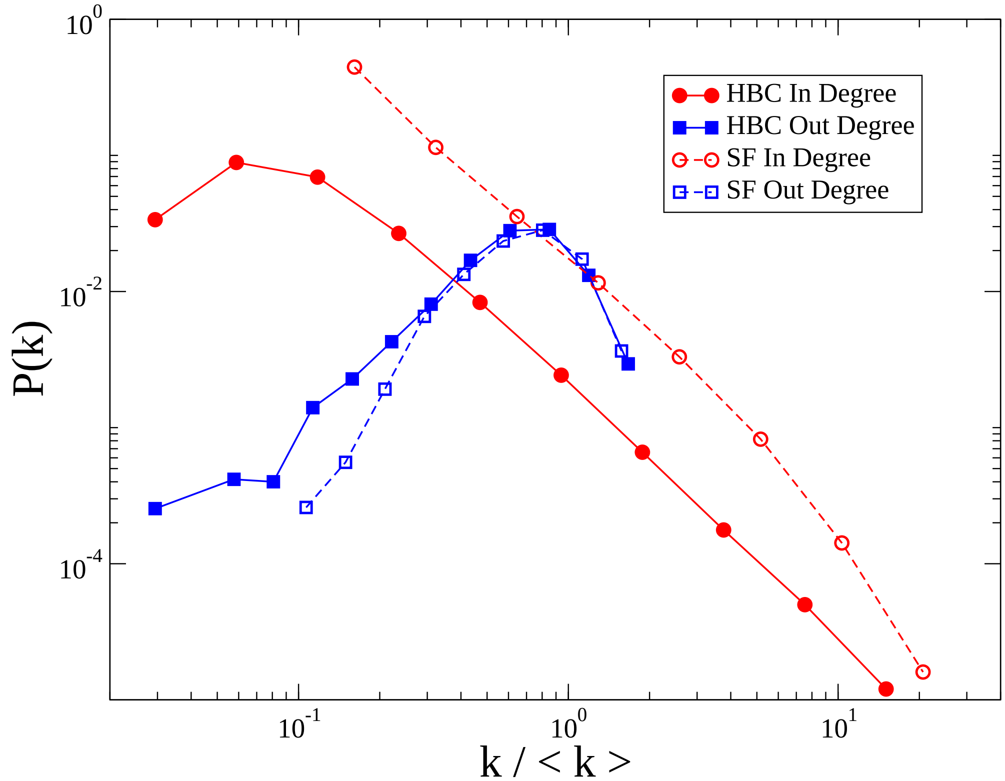


Figure 2.2: The log-binned degree distributions for the word association networks of HBC and SF, in log-log scale. Since we are matching datasets of different sizes we divided the degree by the average degree value. In red, the in-degree distributions for HBC (filled circles) and for SF (empty circles). In blue, the out-degree for HBC (filled squares) and for SF (empty squares). Both the in-degree distributions show a power law form (a straight line in log-log) for sufficiently large $k/\langle k \rangle$, i.e. for values larger than 0.1. Both out-degree distributions have a characteristic scale, showing a peak slightly before 1.

SF, as a function of $k/\langle k\rangle$ (where $\langle k\rangle$ is the average of the distribution). The distribution of out-degrees is narrow, which means that the number of distinct answers given by distinct persons to a given cue is rather limited, as could be expected. On the other hand, the distribution of in-degrees is broad: the number of distinct cues from which a given target can be obtained ranges from 1 to 10 $\langle k\rangle$ with $\langle k\rangle_{in}^{HBC} = \langle k\rangle_{out}^{HBC} = 34$, $\langle k\rangle_{in}^{SF} = 36.5$ and $\langle k\rangle_{out}^{SF} = 6.15$.

The asymmetry between the average k_{in} and k_{out} in SF is due to the asymmetry in the choice of the cue words by the experimentalists. In SF only selected words were used as cues, while many of the answered target words were never used as cues.

Note that the difference of distributions between in- and out-degrees is not a surprise: this kind of asymmetry has been studied also for other different kind of complex network, such as the Web [1], the mail network [2] and the citation networks [3].

The degree distributions of both HBC and SF exhibit very similar shapes, except for a difference caused by the filtering procedure for small in-degrees in HBC (words with low in-degree were scarcely used as answers, those were considered unreliable). It is worth noticing that even though the systems have different sizes, the out-degrees distributions show a peak at close values of $k/\langle k\rangle$, a clear indication of an intrinsic similarity of this specific kind of networks, despite their very different origin.

In order to quantify more precisely the similarity between SF and HBC, we computed the cosine similarity between the neighborhoods of each word. Given a word w_i which belongs to both SF and HBC, we are interested in how much the associations in our two datasets having w_i as a cue are similar. We define l_{ij}^{SF} (resp. l_{ij}^{HBC}) as the number of times that the association between w_i and w_j has been made in the SF (resp. HBC) dataset. The cosine similarity between the associations made with w_i in the two datasets is then defined as:

$$CS_i = \frac{\sum_j l_{ij}^{HBC} \cdot l_{ij}^{SF}}{\sqrt{\sum_j (l_{ij}^{HBC})^2 \cdot \sum_j (l_{ij}^{SF})^2}}, \tag{2.1}$$

where the sums run over all common words between the two datasets. The cosine similarity ranges between 0, no common word is associated to w_i in SF and HBC, and 1, if all the l_{ij} are equal. The average cosine similarity turns out to be 0.215. This number seems to be relatively low but this was somehow to be expected since HBC is much larger than SF and it contains many associations not contained in SF. These associations will contribute to the denominator of Eq. 2.1 but not to the numerator so they will lower the similarity. In order to avoid this kind of bias, we repeated the measure only considering those associations belonging to both datasets and we obtained in this way an average cosine similarity of 0.851. In order to evaluate the significance of these figures of the similarity, we performed 200 independent reshufflings of the SF system's nodes obtaining the following values: $\langle CS \rangle = 0.111 \pm 0.001$ by considering all the associations and $\langle CS \rangle = 0.483 \pm 0.003$ by considering only those associations belonging to both databases. This result demonstrates a strong correlation between the data obtained in the SF controlled experiment and in the web-based HBC game. This is an important information pointing to the reliability of the HBC dataset and of the potential of the method used to gather those data.

2.3 The word association network

Once acquired enough confidence on the reliability of the HBC dataset, let us now deepen our analysis only focusing on HBC, as the largest available Word Association dataset. The directed graph of HBC is composed by $88,747$ nodes and $3,013,125$ edges in total, corresponding to $6,097,806$ registered associations. As we have seen in Fig. 2.2, while the in-degree distribution follows a power-law, the out-degree distribution is peaked. We measure a power-law in-degree distribution of HBC with an exponent $\beta \simeq 1.76$ smaller than 2, which implies that the average value of the in-degree $\langle k_{in} \rangle$ is not well defined, diverging with the size of the system: the number of cues that can yield a given target word as outcome has no characteristic value. On the contrary, the out-degree shows a Gamma-like distribution peaked around a characteristic value $k_{out} \simeq 28$, corresponding therefore to a well-defined characteristic number of distinct words obtained as targets for each word inserted in the game as a cue. The distribution drops very rapidly at large k_{out} with a sharp cut-off around 100, indicating that the number of distinct words that humans spontaneously associate to a given cue is quite restricted. For a better understanding of the origin of this k_{out} scale, we can study the average growth of k_{out} as a function of s_{out}, i.e., the average number of different target words obtained from a given cue word as a function of the number of times the given cue was extracted for a game (see Fig. 2.3).

The inset of Fig. 2.3 shows that $\langle k_{out} \rangle$ can be approximately described by

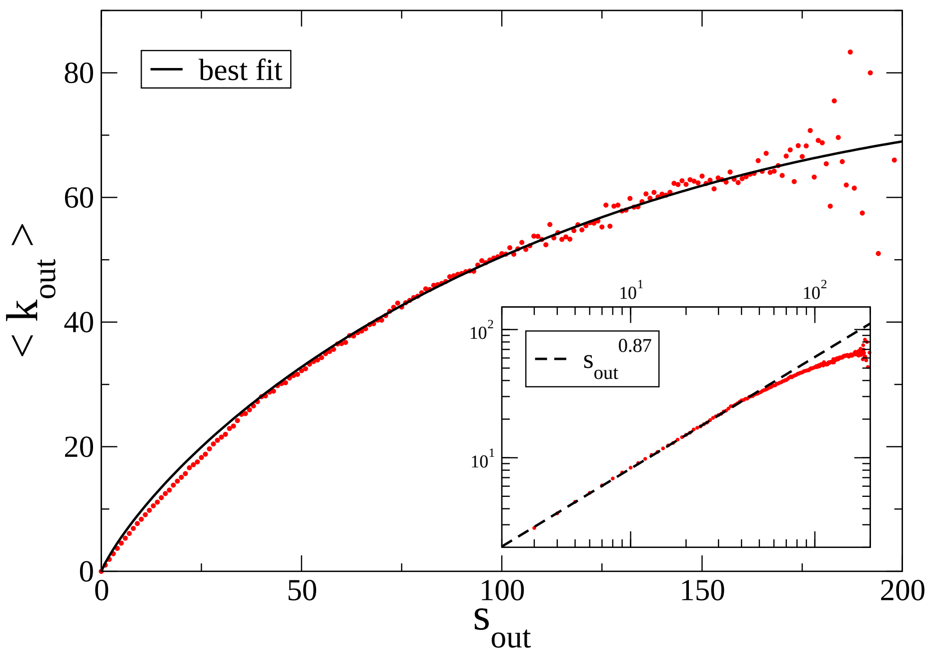


Figure 2.3: Average $\langle k_{out} \rangle$ as a function of s_{out} (red), obtained by averaging the k_{out} values of all those cue words with the same s_{out} value. The fitted curve obtained from Eq. (2.2) is shown in black (parameters: $V \simeq 74$ and $B \simeq 0.92$). Inset: same data in log-log scale together with a fit to a pure sub-linear power-law, to highlight the deviation from such a law at large s_{out}.

a sub-linear power-law with exponent of ~ 0.87, but a deviation from this law is observed at large s_{out}. The sub-linear power-law behavior is well known to appear in the dictionary growth of texts and is known as Heaps' law in this framework [4]. The Heaps' law is symptomatic of an underlying generalized Zipf's law [5], which in our case should appear in the marginal distribution of the target word frequencies, given a fixed cue word. More formally, we can say that the target word associated to a given cue is our random variable and we expect it to be distributed as a Zipf's Law. In other words, let us consider a given word X and all the words that have been associated when X has been used as cue. Some of those target words have been used just once, while some others have been answered more frequently. If we order all those target words from the most frequent to the rarest and plot their frequencies, because of the Heaps' Law observed in the inset of Fig. 2.3, we expect to observe a decreasing Zipf's Law-like power law. Figure 2.4 shows the average ranks distribution for different values of s_{out}, confirming the presence of a Zipf's law, which becomes better defined as s_{out} grows.

To estimate the functional shape of $k_{out}(s_{out})$, we investigate in **??** the impact of finite size effects combined with the above mentioned Zipf's law. We define B as the exponent of the rank distribution of the target to a given cue and assume that there exists a maximum number, V, of different target words that can be associated to a cue word. In this way we are implicitly assuming that k_{out} will converge asymptotically to V (possibly with a residual logarithmic growth that we neglect here). With these hypotheses in mind we derive an expression (see **??**) for the behaviour of k_{out} as a function of s_{out} that reads:

$$k_{out}(s_{out}) = V \frac{B {s'}_{out}^{1/B} - s'_{out}}{B-1} \tag{2.2}$$

where:

$$s'_{out} = \frac{s_{out}}{s_{out}^{lim}} \qquad s_{out}^{lim} = \frac{(V^B - V)}{(B-1)}$$

We also defined the auxiliary parameter s_{out}^{lim}: it is the value of s_{out} for which $k_{out} = V$.

We can use Eq. (2.2) to fit the experimental data and to check whether our finite-size analysis holds in this case. Then, we can also deduce an estimate for B and V. The best fit is shown in Fig. 2.3, leading to the

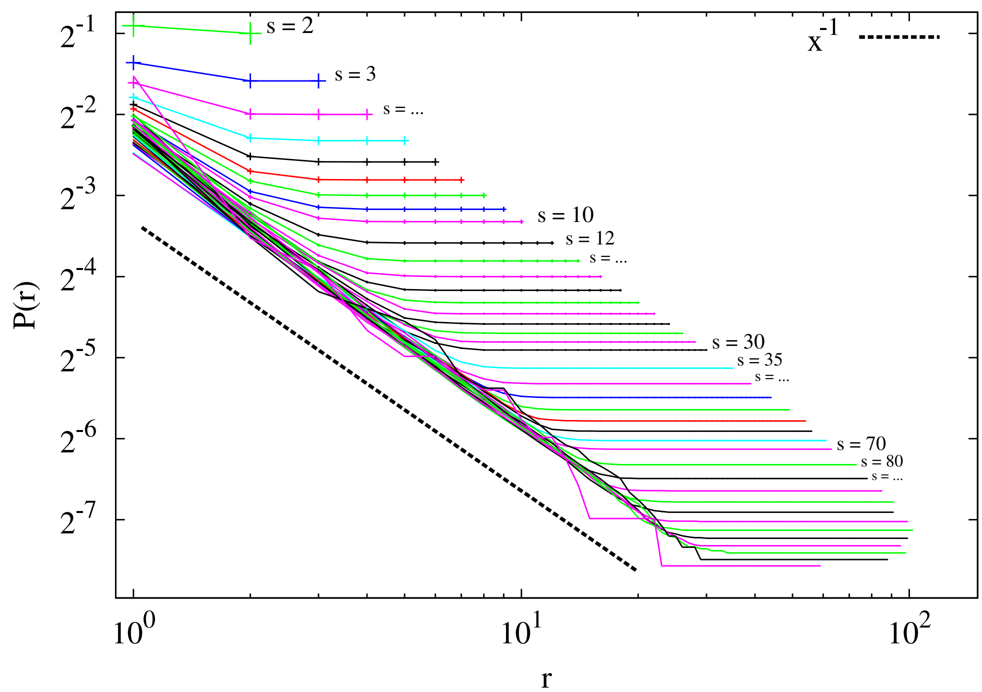


Figure 2.4: For each cue word we calculated the rank distribution (RD) of its target words. Each curve in the picture corresponds to an averaged RD plot obtained by averaging the RD curves corresponding to cue words with the same out-strength value. When s_{out} is sufficiently large (i.e., we have sufficiently large statistics) we obtain a power-law with an exponent close to (minus) unity, followed by a flat tail.

values $B \sim 0.92$ and $V \simeq 74$. It is worth to note that, while Eq. (2.2) is significantly different from a typical Heaps' law, one recovers the pure Heaps' behaviour for $V \gg s_{out}$. In this limit, in fact one has that for small s_{out} and for $B > 1$ the behaviour of k_{out} can be approximated by a power law with the correct exponent $1/B$. If instead $B \leq 1$ the dominant term of the Eq. (2.2) will be the linear one. This finding is consistent with the fact that the growth may not be more than linear since it must always hold that $k_{out} \leq s_{out}$.

It is important to remark that we cannot exclude that the actual formula fitting the data of Fig. 2.3 contains logarithmic corrections implying that k_{out} would keep growing asymptotically with s_{out} though with a logarithmic or sub-logarithmic law. Here we are neglecting this possibility.

The value of V obtained is somewhat surprising since it implies that the asymptotic number of different targets obtained for a given cue is indeed several orders of magnitude smaller than the total number of words in the system ($\sim 90,000$). Hence, we find that on average, there is a limited number of target words for a given cue, reflecting the existence of a limited semantic context associated by humans to each word. This is in contrast with the fact that the number of words (cues) that yield a given target does not show any particular scale: the semantic context is thus not a "symmetric" concept in terms of free word associations.

Other measures can help in characterizing in a more detailed way the topology of the graph. A measure of the distances between the nodes of the graph, i.e., the shortest path between nodes, indicates that the filtered strongly connected component of the HBC network satisfies the "small world" property with an average node distance of 4.0 ± 0.7. This means that with path of around 5 edges we can reach almost every node of the graph, so that we can easily and quickly explore it. This is not a surprise as most complex networks have been shown to share this property [6, 7, 38].

We also analyzed the mixing patterns [8] of the HBC graph, in order to measure the tendency of nodes with similar degree to preferably connect to each other. The assortativity coefficient r is defined as the Pearson's correlation coefficient calculated between the out-degrees of the nodes linked by an edge. If, for each edge, k_c is the out-degree of the cue and k_t is the out-degree of the target, we have:

$$r = \frac{\langle k_c \cdot k_t \rangle - \langle k_c \rangle \cdot \langle k_t \rangle}{\sqrt{(\langle k_c^2 \rangle - \langle k_c \rangle^2) \cdot (\langle k_t^2 \rangle - \langle k_t \rangle^2)}} \tag{2.3}$$

where the averages are calculated on the ensemble of edges. For the unfiltered data we measured $r = 0.0517$ while for the filtered graph we found $r = 0.0644$. The value obtained for the filtered data is slightly larger but still of the same order of magnitude of the result obtained for the unfiltered graph. Hence, it is reasonable to neglect the correlation effects that the filtering procedure may have introduced. We tested the significance of the r values found by performing the same measure on an ensemble of 100 occurrences obtained by reshuffling the edges in the network. We came up with $(1.0 \pm 5.8) \cdot 10^{-4}$, i.e three order of magnitudes lower than the corresponding value of the actual network, confirming that the values of r obtained for the real system are small but still significant. Furthermore, if we consider the number of times a given cue has been associated with a given target as the weight w of that edge, we can define the weighted average of a given function f of the generic edge e as:

$$\langle f \rangle_w = \frac{\sum_e w(e) \cdot f(e)}{\sum_e w(e)}$$

Using this weighted average in Eq. (2.3) we can measure, for the filtered values, the weighted assortativity coefficient $r_w = 0.108$, which, again, points to a slight assortativity. We performed a set of 50 reshufflings to evaluate the significance of the result, obtaining $\langle r_w \rangle = (1.3 \pm 11.9) \cdot 10^{-4}$ which, again, points out a small but still meaningful assortativity.

These results give us an overview of the properties of the HBC network. Using the HBC word association network as a proxy of the way in which our mind stores and organizes all words and related meanings, we observe that it has indeed the properties we should expect from such a network: every word defines a limited context; we can explore the network in a fast and efficient way, for example to recover meanings; and the results about the graph seems to suggest that the network is robust so that is still connected even in case we forget some word or if we do not know it.

2.4 Introducing the phonetical level and duality of patterning

In a seminal paper, Charles Hockett [73] identified duality of patterning as one of the core design features of human language. A language exhibits duality of patterning when it is organized at two distinct levels. At a first level, meaningless forms (typically referred to as phonemes) are combined into meaningful units (henceforth this property will be referred to as *combinatoriality*). For example, the English forms /k/, /a/, and /t/ are combined in different ways to obtain the three words /kat/, /akt/, and /tak/ (respectively written *cat*, *act* and *tack*). Because the individual forms in them are meaningless, these words have no relation in meaning in spite of being made up of the same forms. This is a very important property, thanks to which all of the many words of the English lexicon can be obtained by relatively simple combinations of few dozens of phonemes. If phonemes had individual meaning, this degree of compactness would be hardly possible since phonemes would tend to proliferate in order to accomplish an holistic situation (one phoneme per each meaning). At a second level, meaningful units (typically referred to as morphemes) are composed into larger units, the meaning of which is related to the individual meaning of the composing units (henceforth this property will be referred to as *compositionality*). For example, the meaning of the word *boyfriend* is related to the meaning of the words *boy* and *friend* that compose it. The compositional level includes syntax as well. but here, for the sake of simplicity, we focus exclusively on the lexicon level.

To come up with quantitative measures we considered a set of real language dictionaries. In order encompass the largest possible lists of words, i.e., including proper names and morphological derivations, we focused our attention on the list of words used by the Ubuntu spell checker. These lists are freely available on the standard repositories of Ubuntu. We chose in particular British English, French, German, Italian and Spanish [3]. In addition to the above mentioned lexica, we considered the list of words produced in HBC. In this case we adopted an ulterior and slightly different filtering procedure, adopting Wordnet [96, 97] (WN in the following), a large lexical-semantic database of English. We left out all the word that did not have match on the WN dataset, in order to eliminate slang expression, proper names, etc.

Our analysis of combinatoriality and compositionality has been performed by subdividing each word of the lexicon in its phonetic components according to the International Phonetic Association[4] (IPA) coding. In order to perform easily an automated analysis, we adopted a particular encoding of the IPA alphabet, namely the Speech Assessment Methods Phonetic Alpha-

[3] Downlodable at `http://www.debian.org/distrib/packages`
[4] `http://www.langsci.ucl.ac.uk/ipa`

bet[5] (SAMPA), a computer-readable phonetic script using ASCII characters. More in detail, we used eSpeak, a compact open source software speech synthesizer that translates any texts in its SAMPA encoding.

First, we measured the distribution of word lengths in phonemes for the lexica of the five languages considered as well as for the HBC's lexicon. The normalized histogram of word lengths is reported in Fig. 2.5 (left). For all languages we observe fairly compatible shapes and in same range we even see identical trends; it is worth to note that in the most important region (a zoom of which is portrayed in the inset of Fig. 2.5) (left) all the languages show different behavior, except for the HBC and the English language. These similarity was somehow expected since HBC is practically a sample of the English language and it is a supplementary proof of the HBC reliability for our purposes.

We now define the frequency of a phoneme in a given language as the number of words containing that phoneme normalized with the total number of words in the lexicon of that language. It can also be roughly considered as the probability of appearance of a given phoneme. Fig. 2.5 (right) reports the frequency-rank plot of the phoneme frequencies. After rescaling the frequencies with the size of the lexicon it is quite evident that the shape of the distributions is roughly the same, except for the number of different phonemes in each language corresponding to the maximal rank. The phonemes frequency has been measured also for the HBC's lexicon. The figure show that HBC distribution is quite similar to the English distribution. This fact points out again the reliability of the HBC database when used to sample the English language phonetic properties.

A complementary measure of the statistics of use of the different phonemes in the lexicon is given by the entropy. The entropy of the elementary phonemes distribution is defined as $S = -\sum_{f_i} p(f_i)\log(p(f_i))$ where f_i is the generic elementary form (phoneme) and $p(f_i)$ is the frequency of occurrence of f_i in the whole emerging lexicon. We also introduce a measure of combinatoriality to quantify the property of a communication system to combine and re-use a small number of elementary forms to produce a large number of words. Following the idea in [63], we introduce a real-valued quantity ranging in the interval $[0, 1]$ that quantifies how frequently elementary

[5] http://www.phon.ucl.ac.uk/home/sampa/

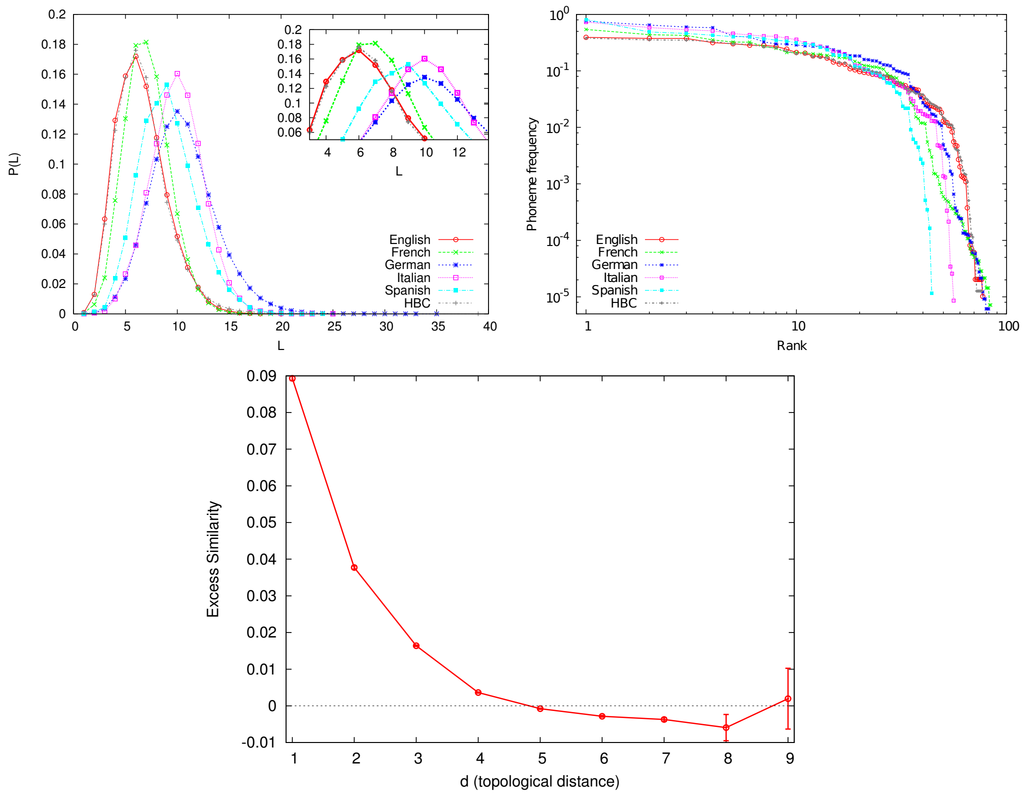


Figure 2.5: *Left* The normalized histogram of the number of phonemes composing the words of the six lexica considered. In the inset, a zoom of the region with percentages above the 5%. *Right* The phoneme frequency-rank plot for the words of the considered lexica. On the horizontal axis we report the rank of individual phonemes and on the vertical axis the corresponding frequency in each lexicon. *Bottom* Excess Master-Mind similarity of words as a function of the average distance d of the corresponding objects on the HBC word-association graph.

word list	M	F	$S/S_{\max}$	C
British English	98,326	77	0.8228	0.079
French	139,721	84	0.7644	0.077
German	327,314	83	0.7724	0.109
Italian	116,878	56	0.8087	0.147
Spanish	86,016	44	0.8206	0.167
HBC	78,954	76	0.8336	0.080

Table 2.1: Phoneme statistics of the different languages considered and of HBC. M is the number of words in the list, F is the number of different phonemes observed in the word list of each dictionary (according to the SAMPA classification), S is the entropy calculated on the normalized frequency distributions, $S_{\max} = \log(F)$ is the maximum value of the entropy reached if all phonemes were equiprobable. C is the combinatoriality defined in Eq. (2.4).

forms recur in a lexicon, according to the following expression:

$$C = \frac{\sum_i (m(f_i) - 1)}{(M-1)F}, \tag{2.4}$$

where the sum runs over all the F distinct forms present in the emerged lexicon and $m(f_i)$ is the number of distinct objects whose name includes the form f_i. The term $m(f_i) - 1$ takes into account only the forms that are used to name at least two objects, i.e., only the forms that are actually re-used. M is the number words in the lexicon. Table 1 reports the results for the values of the entropy and combinatoriality for all the lexica considered.

Let us now turn to the compositional aspects of the lexicon. The aim here is that of establishing whether, in a specific lexicon, words of semantically related concepts are expressed through morphologically similar words. To this end, we need to focus on a suitable conceptual graph, i.e. a network where the nodes are the concepts expressed by words and the connections are semantic relations between pairs of concepts. At this point, the most natural choice seems to consider the word-association graph produced by the HBC experiment as a proxy of the conceptual space and we shall measure the semantic similarity of two words in terms of their topological distance on the graph. In addition, we need to define a measure of morphological similarity between words. To this end we introduce a Master-Mind-like (MM) measure. Given two words w_1 and w_2, each composed of a certain number of phonemes, the Master-Mind-like (MM) measure of phoneme similarity is defined as follows. After the two words have been aligned, either making the left-end or the right-end coincide, we sum 1 whenever the two words share the same form in the same position and 0.5 for the same form in a different position. We take the MM measure to be the maximum between the left-end and the right-end alignment values. This MM measure conveys the idea that meaningful forms are often included in words as suffix or prefix, and in general in a well defined position.

To quantify compositionality, we measure the *excess similarity* of words used to name related objects (at small topological distance in the graph) when compared to the similarity of randomly chosen words. In order to do that, we consider the average difference between the similarity between each pair of words as a function of their distance on the HBC graph and the same value computed in ten random cases, obtained by reshuffling the associations between words and nodes in the HBC graph. Fig. 2.5 (bottom) reports the results for HBC. Ten instantiations of the null model are sufficient since the standard deviation of the distribution (represented with error bars in Fig. 2.5) is negligible for all meaningful distances. The excess Master-Mind similarity is larger for more semantically related words, i.e. for words at small topological distance in the HBC, and it monotonically decreases with the topological distance, being significantly different from zero up to topological distances of the order of 3. We interpret this results as a clear signature of compositionality in HBC and, being HBC a good proxy for English, more generally for English. Though semantically related words are not necessarily related in forms, an excess similarity between words used to name related meanings do exists in natural languages and is statistically relevant.

These measures provides empirical confirmation to theoretical predictions based on a recently introduced multi-agent modeling scheme, namely the *Blending Game* [127]. This model provides a mathematical framework to address the problem of how a population of individuals can bootstrap combinatoriality and compositionality. It is remarkable that the two sides of duality of patterning emerge simultaneously as a consequence of a pure cultural dynamics in a simulated environment that contains meaningful relations, provided a simple constraint on message transmission fidelity is also considered.

2.5 A new data source: opportunities and threats

So far we analyzed the largest, to our knowledge, dataset of word association. We measured the properties of the word association graph, in the same way we would have analyzed any other complex network. And we also used the graph to analyze language properties and to give a quantitative measure of complex entities like compositionality and combinatoriality. The application of Complex Systems tools to these open issues imposed to face the intellectual challenge to build a suitable analysis framework, which enabled us to come up with the scientific results we showed. And it is worth to note that we managed to state our scientific results with a good degree of confidence mainly thanks to the statistical importance of our dataset.

To implement a classical social science experiment aiming to gather the same amount of data would have required an extraordinary effort in terms of time, money and human resources, so from this point of view it seems extremely convenient to use web-sourced data like the HBC data. But there is obviously another side in this coin, and risks are growing along with the size, the speed and the savings. We had to face some of these threats during our work.

The first and, in the HBC case, the most important of all was the validation of the dataset, which entailed an effort at several levels. As may be expected, an open web-application gather a lot of spam, which obviously have to be filtered out. Also, there was a significative amount of junk, or simply misspelled, words that we needed to eliminate. The identification of dirt in the dataset has been an absolutely non-trivial task. We may speculate that, from a science philosophy point of view, there is only one big difference between our work and the filtering procedures of any particle physics dataset: there is considerably smaller literature to face our challenge, so our methodology choice were more risky. And, in the end, could not grant the absence of bias in the dataset. That is the reason why we double checked HBC associations with SF experiment results. That gave us the confidence to perform our analysis. And to go further, we applied what we found to other problems, providing new results and also achieving an ulterior indirect validation of the data.

There is also another important aspect we neglected so far. Our source was not intended for scientific purposes. We did not had any control at all on the data gathering, we just validated what we got from the HBC web-application. The measurements are still valid, exactly how astronomy do not lose its value because of the astronomers impossibility to interact with stars and planet. But it is worth to stress this strong experimental limitation. And it is worth to note that there is a simple way to get over it: the researcher can actually plan and implement his own web-experiment. But, again, this will comport a brand new potential and a set of new challenges that we will discuss in the next chapter.

Chapter 3

A Real Complex Systems and Social Computation experience: EveryAware and WideNoise

By introducing the HBC experience we outlined the potential of the web and its user community as a valuable data source, even if the subjects were completely unaware of that. In fact, the whole scientific value of HBC was unveiled a posteriori. In this part it is our intention to show how web users may be actively involved in a well-defined experiment of social computation, in the context of EveryAware[1], a EU medium-scale focused research project. The project was funded by the Future and Emerging Technologies program (IST-FET) of the European Commission under the EU RD contract IST-265432.

The issue of sustainability is now on top of the political and societal agenda and is considered to be of extreme importance and urgency. There is overwhelming evidence that the current organization of our economies and societies is seriously damaging biological ecosystems and human living conditions in the very short term, with potentially catastrophic effects in the long term. A lot can and must be done from the technological and policy-making perspective, for example to build passive houses, develop renewable energy, and so on, but it is only when people become fully aware of their actual environmental conditions and their future consequences that the much needed change of behaviour will truly happen.

[1] `http://www.everyaware.eu`

Public participation in environmental decision making was pushed to the fore as a result of the 1992 Rio Declaration on Environment and Development. However, the provision and production of environmental information, particularly on issues such as noise pollution and air quality, rely heavily on a 'top-down' approach in which public authorities collect the data and release it to the public. There is still room to develop better mechanisms that support citizens to not only consume but to generate their own environmental information. If successful, such processes could lead to an increased awareness and learning about current environmental issues. Furthermore, this may serve to encourage more citizens to participate in environmental decision making, and ultimately stimulate them to take steps to improve their own environment based on new observation techniques.

The EveryAware project responded to this societal need by pushing the evolution of ICT with the aim of supporting informed action at the hyperlocal scale, providing capabilities for environmental monitoring, data aggregation, and information presentation. The goal was that of enhancing knowledge, understanding and social awareness about environmental issues emerging in urban habitats through the use of ICT tools deployed to gather user-generated and user-mediated information from mobile sensing devices. The leading idea of the EveryAware Consortium has been that gathering both personal and community data, collected by citizens, processed with suitable analysis tools, and represented in an appropriate format by usable communication interfaces, has the potential of triggering a bottom-up improvement of collective social strategies.

With this aim in mind EveryAware deployed the infrastructures to support participatory sensing in an environmental framework, high-performance data gathering and storage. The resulting EveryAware platform is highly effective and represented the main backbone for all the EveryAware activities. The very same realization of the EveryAware infrastructure represents a major achievement of the project since for the first time we demonstrated a complete end-to-end infrastructure able to integrate participatory sensing, accuracy of measurements from low-cost sensors, people engagement, mobile and web technologies. This infrastructure has been successfully deployed in several case studies[2] devoted to Noise pollution and Air-quality. In addition EveryAware launched the Experimental Tribe platform[3], a general-purpose

[2]`http://cs.everyaware.eu`

[3]`http://www.xtribe.eu`

platform designed for scientific gaming and social computation whose aim is that of providing the scientific community with a tool to realize web-based experiments by skipping all the unnecessary technical coding overhead.

While the Experimental Tribe platform and the air quality case study will be described in the latter parts of this thesis, in this chapter, after a brief description of the EveryAware main activities, we will focus on the analysis of the behaviour of individuals involved in noise sensing. Citizens have been involved in noise measuring activities through the WideNoise[4] smartphone application. This application has been designed to record both objective (noise samples) and subjective (opinions, feelings) data. The application has been open to be used freely by anyone and has been widely employed worldwide. In addition, several test cases have been organised in European countries. Based on the information submitted by users, an analysis of emerging awareness and learning is performed. The data show that changes in the way the environment is perceived after repeated usage of the application do appear. Specifically, users learn how to recognise different noise levels they are exposed to. Additionally, the subjective data collected indicate an increased user involvement in time and a categorisation effect between pleasant and less pleasant environments.

3.1 The EveryAware Project

The EveryAware project expected to contribute significantly to the social goals of achieving greater awareness of localised, personalised environmental information through the implementation of novel infrastructures for bidirectional communication.

Specifically, it aimed to develop the tools and the knowledge needed to make environmental information transparent, available and easily integrated with the perceptions of people, regarded as a first-order observable. Bridging the gap between opinions and sensor data is the single factor that can make environmental knowledge actionable at the grassroots level. Current approaches to the onset of sustainable practices in citizens' environmental behaviour have been based on top-down strategies for understanding behaviour [79] and have met with mixed success [78]. The participation of citizens has traditionally been limited to opinion polls and public discussions where people have been asked to convey their needs and their opinions to

[4] `http://cs.everyaware.eu/event/widenoise`

panels of designated experts responsible for tackling emerging issues. The environmental monitoring activity, the public dissemination and discussion, and the policy making are performed in separate places and at different times, with little transparency about how environmental issues are treated by each actor throughout the whole process.

EveryAware project, conversely, has been based on the idea that citizens should be involved not only as passive receivers of pre-packaged environmental information, but also as active producers of it, by means of the networking possibilities allowed by mobile devices, pervasive Internet access, Web 2.0 and the mobile Web tools that support sharing and annotation of geo-localised content. The framework envisioned in the project allows the user to participate in all stages of environment management: by contributing to enrich its monitoring, expressing opinions, joining a motivated community, and eventually implementing best practices with the potential to improve environmental conditions.

The notion of geo-localised user-generated content is of course not novel. A number of participatory websites and Internet-based scientific projects have been successfully deployed (see [60, 65, 76] or OpenStreet Map[5] for examples and a review of the field of Volunteered Geographic Information). However, most collaborative web-based systems have bound themselves to merely visualise the data collected by users, without a scientific analysis of it. In contrast, EveryAware proposed that users participate in the scientific endeavour itself by making use of current and emerging hand-held electronic devices incorporating significant computing power. Such devices should be easily connected to sensing equipment and to the Internet without requiring specific expertise from the user. In the field of environmental monitoring and research it was, and still is, a great novelty to deal with data from a large number of mobile, randomly distributed, "uncontrolled", low-cost and therefore potentially less reliable sensors carried by non-skilled individuals, as compared to the practice of a limited number of mostly stationary and highly controlled data collection systems based on expensive high-quality measurement instruments. It was additionally novel to involve non-expert users in an end-to-end process from data capture to final output. The integration of participatory sensing with the monitoring of subjective opinions has been the key and crucial novelty of EveryAware, as it has the potential to expose the mechanisms by which the local perception of an environmental is-

[5] `http://tah.openstreetmap.org`

sue, corroborated by quantitative data, evolves into socially-shared opinions, and how the latter, eventually, drive behavioural changes. In our opinion, this approach represents a scientific and technological advance from several points of view as explained below, and EveryAware carefully addressed all the different research and technological challenges it implies.

3.2 Scientific and technologic challenges

In the following is reported a short overview of the main challenges faced in the context of the EveryAware project.

3.2.1 The EveryAware platform

A key technological novelty of the EveryAware project has been the design and the implementation of the so-called EveryAware platform that handles both sensor and subjective data acquisition. The platform is a modular system composed by several components: a SensorBox to gather objective data about the environment, a smartphone controlling the data acquisition and the user-experience, a system of data gathering, storage, analysis and visualisation and several web-services. This approach guarantees high scalability of the overall system and allows for further developments aimed at having pluggable sensors, eventually miniaturised and integrated (e.g., wearable sensors). At the same time the associated software platforms allow users to easily upload their sensor readings, and equally easily tag these with subjective information. The ICT challenge here was that of making this upload process as automatic and natural for the user as possible.

3.2.2 Community Engagement

Work dating as far back as 1969 [26] lists the possible levels of citizens' participation, ranging from non-participation to citizen control (where budgets are assigned to the citizens themselves) and more recent projects ([24, 89, 91, 107, 111]) stress the importance of the participation process and the impact that informed community members have on local decisions. Such participation can improve both the science literacy of a population [108] and offer different views of communities [119] to scientists: the real-time monitoring of opinions related to empirical observations will provide environmental sociologists with a corpus of detailed knowledge about how environmental conditions are perceived by a community: what issues are regarded as most relevant? How are novel behaviours propagated? What motivates participation, engagement and behaviour change?

Motivation for users' engagement and continuing participation in online project such as Wikipedia `http://www.wikipedia.org/` or OpenStreetMap [6] has already been extensively examined [20, 33, 104]. However, similar motivations cannot necessarily be attributed to the citizen sensing participants in the EveryAware project, which presumably requires a higher level of commitment to that of a Wikipedian (who contributes 8.27 hours per week on average [104]). Obtaining information related to encouraging initial and continued participation was therefore fundamental to the developers of systems such as EveryAware as it can be utilised to ensure that participants are highly motivated to engage with the project, and more importantly remain engaged over the longer term. Novel research has been focused on two aspects of the problem. Firstly, a number of participant recruitment techniques (such as social networking sites, flyers, posters, e-mail campaigns) has been trialled systematically to identify those that achieve greatest success and validate whether similar techniques can be applied both in cross-border situations and with groups having different interests. Secondly, still ongoing research is identifying a list of motivations for ongoing participation once recruited, with a particular focus on those users who remain engaged with the project over a longer term. The results from both elements of research not only informed all the stages of the project, but will also be of great relevance to similar participatory projects elsewhere.

3.2.3 Processing sensor data

Specific issues emerged concerning sensor data. To illustrate this point let us focus on air-quality sensors. Although in most epidemiological studies air quality is commonly defined at the level of a city, recent air quality studies have highlighted that significant differences in pollutant concentrations, and in related health effects, can occur over the day and between different locations [32, 81, 98, 135]. The measurement of air quality at a high spatial and temporal resolution can yield a tremendous advance in the characterization of the pollutants' urban concentration variability. Measuring mobility

[6] `http://www.openstreetmap.org/`

and activity patterns allows researchers to gauge the real-world exposure of citizens, and in turn the overall effect on the health of urban communities.

The use of networks of available low-cost sensors will enlarge the data coverage. In the past, the adoption of low-cost sensors for ambient air quality monitoring has always been constrained by lack of accuracy, selectivity and reliability [39]. However, new sensing technologies (arising from additional developments in the fields of semiconductors, nanotechnologies and fibre optics amongst others), will bring the detection limits of commercial sensors to the part-per-billion range needed for air quality monitoring. At the same time selectivity increases [36, 56, 88, 129]. Thanks to the integration of cheap sensors in sensor networks, increased data availability, network intelligence and advanced data mining techniques, limited accuracy and reliability can further be countered [83, 90, 128] (see also IDEA project `http://www.idea-project.be`).

Several research projects have developed or are developing low-cost portable air quality sensing tools based on commercially available sensors [25, 55, 74, 77, 92, 98, 130] (see also the Cambridge Mobile Urban Sensing (CamMobSens) `http://www.escience.cam.ac.uk/mobiledata/`). However, when EveryAware started none of those efforts had reported extensive field trials or reported full-scale validation exercises. Specific technical challenges have also to be tackled such as the precision of GPS in densely built urban environments [98].

3.2.4 Combining sensor and subjective data

One of the main novelties of EveryAware has been the strong effort towards an integration of sensor and subjective data in order to provide insights about the social perception of the state of the environment (see also below). A quantitative analysis of the gap between perceived and measured environment had never been attempted in a systematic way. Both kinds of data are affected by the procedures to gather them as well by intrinsic biases, both in space and in time. This raised new issues of data validation, calibration, interpretation and representativeness, that had to be tackled in a creative way and embedded in digital data processing procedures in an, as much as possible, autonomous, learning way.

3.2.5 Citizen science

An important challenge concerns the development of and examination of the use of web-based tools through which (groups of) interested lay people and scientific experts can interact directly, discuss provisional results of data collection and mutually enrich both the data itself and the interpretation of the data. Here the actual challenge was the presentation of complex scientific analysis in a user-friendly manner to non-specialists. From this point of view the project paid a special attention to ICT challenges that include: (i) the usability of the interface design so that users can easily find the desired information (at the individual level or aggregated); (ii) the appropriateness of the actual displaying methods: how to present results so that non-specialist users understand both the analysis undertaken and the outcomes? Will access to this information help users feel rewarded and part of a community, encouraging further participation? Thus the overall novelty of this component of the project has the development of a user-friendly manner to present complex scientific analysis (both the methods and the results) to non-specialists.

3.2.6 Opinion and Behaviour Change

The direct involvement of the users in the research as described above leads to the potential discovery of emerging behavioural patterns, as well as to an assessment of the impact of new technological solutions at the socio-economic level. Despite these benefits, none of the existing studies [24, 25, 55, 74, 77, 90, 92, 98, 107] using citizen sensors specifically evaluate individual behaviour change in any way, although Honicky *et al.* and Milton and Steed [74, 98] raise this as an issue to be investigated.

This issue is closely linked with the concept of participant motivation described above - will a participant sufficiently engaged with the project also modify his or her behaviour as a result of the personalised information presented? Lawrence [86] notes that the link between engagement and behaviour change is not yet fully established in the context of environmental change and climate change discourse. Although other studies using diverse sources of data have identified the usefulness of such individualised information ([50, 107]), many of the citizen sensor studies are still at pilot stage [74, 98] and do not state behavioural investigation as one of their direct aims.

In general, the dynamic processes underlying the formation and the evolution of opinions, uses and behaviours have rarely been investigated in experimental settings and almost never coupled to the exposure of users to suitably detected and processed relevant information. Influencing behaviour change is notoriously difficult due to the complexity and variety of factors that affect behaviour [79], and a number of alternative models have been proposed. "Expectancy-value" theories group together models where choice is motivated by the expectations we have about the consequences of our behaviour and the values we attach to those decisions [79] (for example the rational choice model). Staged models (Prochaska and DiClemente (1986) and Lee and Owen (1985) in [120]) include the fact that understanding and assimilation of the consequences of an action may be incomplete, that information may relate to events in the future (e.g. the possibility of developing lung cancer) and that a distinct cognitive effort is required to modify behaviour [79]. The basis of all behaviour models, however, is the assumption that knowledge and awareness of an issue or a problem are key requirements for a behavioural change. However, very few studies have been undertaken on changes in individual behaviour due to the provision of individual-specific information.

A theoretical contribution to the understanding of opinion and behaviour change came from recent studies performed in the opinion dynamics field [40]. Such interdisciplinary area focuses on the modelisation of opinion spreading in large social networks, with a heavy use of mathematical tools and methods borrowed from statistical physics. Many models have been developed in the literature to explain how social systems develop a consensus on a given issue (e.g., on political votes), or which social interaction favours the coexistence of multiple opinions in a community [84, 124]. However, empirical bases behind such models are still scarce, in particular for what concerns the opinion dynamics, which requires the monitoring of a social system during time. Although some of the partners of the project had already explored these problems in recent works, focussing on the emergence of semantic agreement in social networks [125], crucial issues such as the study and the modelisation of the resistance to opinion shift are still a largely unexplored field. The EveryAware project contributed to provide the empirical, computational and theoretical base for an advance in such line of research.

3.3 The WideNoise Case Study

In this section, we introduce one of the main Case Study implemented in the context of the EveryAware project: the WideNoise participatory sensing campaign [31]. This is a smartphone application developed within the EveryAware project, which was designed not merely as a measurement tool for its users, but also as a means to monitor opinions on the environment and noise, in a way as transparent to the user as possible. Hence the application has several features that allow for subjective/personal data to be acquired. Using these data, an analysis of user behaviour/opinions that may emerge after usage of WideNoise was performed. Changes in behavior are indeed visible after a user performs several measurements, which is a strong indication of increased awareness and learning.

3.3.1 People VS the noise: state of art

Noise pollution is a problem in cities across the world and is one that is likely to affect an increasing number of people with the majority of the global population now living in urban areas [136]. In Europe, this has been recognised and abatement measures have been introduced in many countries. However, noise pollution, in particular, is an environmental problem that relies heavily on 'top down' approaches, both in terms of communicating the issue, through instruments such as strategic noise maps, but also in the methods used to gather data. For example, strategic noise mapping became a requirement of all Member States under the EU's European Noise Directive (ENDS). The maps are used to estimate population exposure to noise in certain areas, to communicate to the public and as a basis for action plans [46].

Exposure to noise is not merely a case of annoyance. Researchers have provided a growing body of evidence that suggests that long-term exposure to noise constitutes a health risk hazard and can modify social behaviour, cause annoyance [106], increase the risk of cardiovascular diseases [28] and adversely affect levels of attentiveness and the ability to read in children [68]. The World Health Organisation (WHO) estimated that at least one million healthy life years are lost every year from traffic-related noise in the western part of Europe [61].

New participatory sensing applications that exploit information and communication technologies (ICT) are providing novel approaches to environmental monitoring. Simultaneously, they present an opportunity to widen citizen engagement and participation in local, regional and global environmental issues. This has been enabled, in part, by the relative affordability and growth in the number of smartphones in use, now estimated to have breached the 1 billion mark [23]. Miniaturisation of embedded sensors in these devices, such as a microphone, camera, accelerometer, and GPS receiver, combined with the increasing computation power, network connectivity and data plans has resulted in an increasing number of smartphone Apps (short for applications) designed for a range of participatory sensing opportunities.

Participatory sensing, also referred to as urban sensing, involves enabling individuals, groups and communities to gather, document, view, share, and in some cases analyse local observations and data about their surrounding environment. Not all participatory sensing relies on mobile technologies. For example, [69] comment on the use of low cost noise monitors in a citizen science project in which two communities collected noise data: one in relation to noise nuisance being generated by a local scrap yard and the other, in an objection to an airport expansion plan. However, the use of smartphones as sensory devices, either passively or actively, increases the ability to scale such activities. Cuff et al. [49] highlight a range of applications in which citizens can be engaged in mobile sensing, predicting a growth in the field and in the numbers of ways in which it will be applied.

The power of the 'crowd' has been recognised as an effective way of generating observations, which might otherwise be difficult to obtain, due to spatial and temporal limitations. This is particularly relevant in fields where traditional sensing relies either on a distributed network of expensive stationary monitoring devices across a target area of interest, or where sensors require physical placement for a specific deployment, or in cases where numerical simulations are needed. Cost and data coverage are key factors. The spatial distribution of static monitoring devices and the associated costs of hiring trained specialists to take measurements and process data reduce the amount of real-world measurements that can be taken.

Noise provides a good example for this problem, and is one of the environmental domains in which computer simulation models are used to predict noise exposure from various sources such as different modes of transportation in a given area. Measurements are often used to validate the results. However, observations and data provided through participatory sensing activities could provide added value to existing data sources.

Applications such as NoiseWatch in which the phone's microphone is used to determine the level of sounds it detects, which is then displayed in decibels (dB), have begun to emerge [57]. An earlier example is NoiseTube, which adopts a passive approach to noise monitoring. The App constantly collects environmental sound, in second-long samples, provided that the application is left running [93]. An additional feature allows users to annotate and tag measurements. The *Noise Nuisance* App, created by a group of environmental health professionals, provides a country specific design that is tailored to tackle noise nuisance in the UK. It provides users with relevant email addresses of every local authority noise investigation team which is programmed into the App, information on statutory noise, and alternative action measures that can be adopted. Users can record audio samples, make an entry into a diary each time the noise affects them and report these to their environmental health department or housing association officer [102].

All these applications are providing the tools to record the noise, with some attempting to overcome the technical limitations of smartphone and produce more accurate measurements. Other than a preliminary analysis of tagging patterns in the NoiseTube community [121], the authors are not aware of any studies that have provided empirical data that explores user behaviour/opinion patterns over time. More specifically, that points to the change of user opinions/awareness of perceived noise levels that may occur with an increase in use of such applications.

3.3.2 The WideNoise app and platform

WideNoise is a mobile application for recording, monitoring and analysing noise pollution. The application is intended to run on mobile devices and more specifically on Android[7] and iOS[8] platforms. It was originally developed by WideTag[9] and then was enhanced by the EveryAware team, who has improved and expanded the data recorded from the mobile device by adding new features. The mobile application sends anonymous data to an application server capable, through RESTful web services, of collecting the acquired data and showing the corresponding information on a map. Both sensor data and subjective perceptions are required to create a full sound report, so that the application consists of two main parts: the noise sampling

[7] `http://play.google.com/store/apps/details?id=eu.everyaware.widenoise.android`

[8] `http://itunes.apple.com/app/id657693514`

[9] `http://www.widetag.com/`

component and the perception tagging.

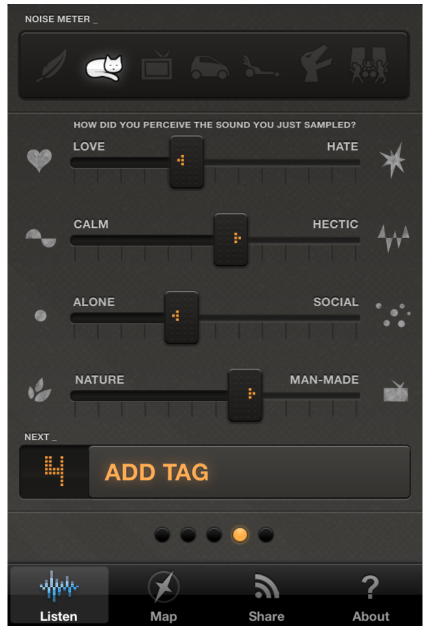
The noise measurement part gives users the possibility to take a noise sample through the smartphone microphone. When the recording starts, the user is asked to guess the noise level through a slide bar where a decibel scale is mapped. The user has also the possibility of extending the default sampling time. In this way, while the user gets more time to make the guess, the app will perform a longer measurement. After the recording phase, the noise level expressed in decibels (dB), is shown and compared to the level estimated by the user. The sound level is associated with an iconographic representation that allows the user to better understand the decibel value, by using seven intuitive noise sources: falling feather, sleeping cat, tv show, car engine, dragster, t-rex and rock concert (Figure 3.1(a)).

After the noise recording/guessing stage, the users are asked to express their own feelings about their current environment. They can provide a score by moving four different sliders associated to as many feelings and categories: love/hate, calm/hectic, alone/social, nature/man-made (Figure 3.1(b)). The users can also associate free text tags to the noise. Once the tags are applied, all the information collected by the application is sent to the web application server as soon as a working data connection is available (Figure 3.1(c)). WideNoise allows users to view a community map displaying the average noise level at nearby locations, by relying on the statistical elaboration provided by the server (Figure 3.1(d)). As an integration with social networks, users can also share their own recordings via Twitter and Facebook.

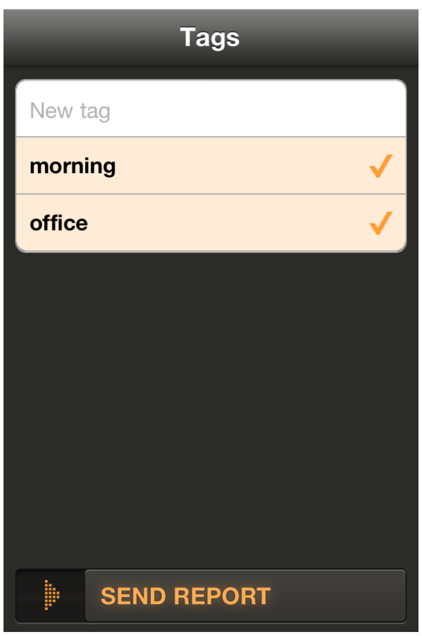
3.3.3 Participation

An obvious key integral component to any participatory sensing activity is the participants (and users). The EveryAware project has stated a clear goal to enable public participation in sensing activities and as such, face-to-face and virtual communication was adopted to recruit participants for the initial test cases. Three public test cases were initiated covering the period from February, 2012 to May 28th 2013. The first occurred during the Citizen CyberScience conference in London, which ran over a period of three days. Around 170 delegates were encouraged to download the WideNoise App and to take measurements in the conference facilities and in the surrounding area. The second test case focused on engaging communities surrounding London Heathrow Airport and was kicked-off by a launch event in the Isleworth community on the 19th of June 2012. In 2013, the test case was extended to

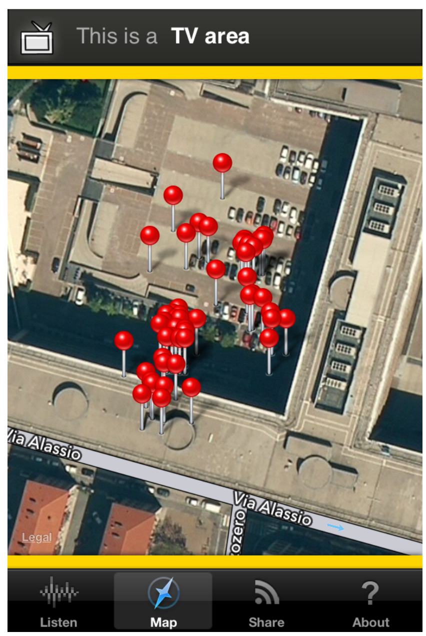
(a) Noise sample screen.



(b) Perception screen.



(c) Tag screen.



(d) Map screen.

Figure 3.1: WideNoise Plus mobile application.

the community of Windsor, with WideNoise adopted by the local authorities and a training session organised by our team on the 23rd of April. The third test case involved a one-day event held at a bookshop in Rome (9th June, 2012). Visitors to the bookshop were encouraged to capture noise measurements from the surrounding streets and these were visualised in real time on a large screen inside the store. Additionally, an internal test case was also organised in Antwerp, Belgium, on the 10th of July 2012, where members of the EveryAware team performed measurements in the city centre area. Also, a workshop with architecture students was held in Birmingham on the 5th of October 2012.

For the Citizen CyberScience conference an email was sent to all the delegates prior to the start of the conference. Email reminders were also sent every morning over the course of the three days. In addition, a short presentation was given by one of the EveryAware team members on the first day inviting people to participate. Custom business cards with links to the WideNoise application were handed out to all the delegates and project team members were on-hand to answer questions.

In order to recruit communities surrounding London Heathrow we adopted a 'hands on' approach, as well as a 'virtual' approach. Virtual recruitment included banner ads on hyper-local websites in addition to posting on Twitter and Facebook. Articles were published in several local papers and the project received coverage on BBC London news and local radio. Hand-distributed flyers were circulated and posters were placed in shop windows. The mailing list of an anti-airport expansion campaign organisation was used to inform people about the project. An official launch was held in a local community centre that was attended by local residents, a local councillor and members of the campaign group.

The third test case in Rome was promoted via Twitter, Facebook and other online web sources, as well as having a dedicated website[10]. It was also advertised in print media throughout the locality in the days before the event in order to attract participants from the neighbourhood. Several posters and flyers were placed in areas commonly frequented by local residents.

[10]in italian: `http://www.everyaware.eu/segiochifaiscienza/`

3.3.4 The WideNoise Dataset

In this section we will give a general description of the amount of data, participation patterns and coverage. The data considered for analysis have been collected up to June 7th 2013 at 3AM. These consist of 41478 sound level measurements made by 13962 unique devices.

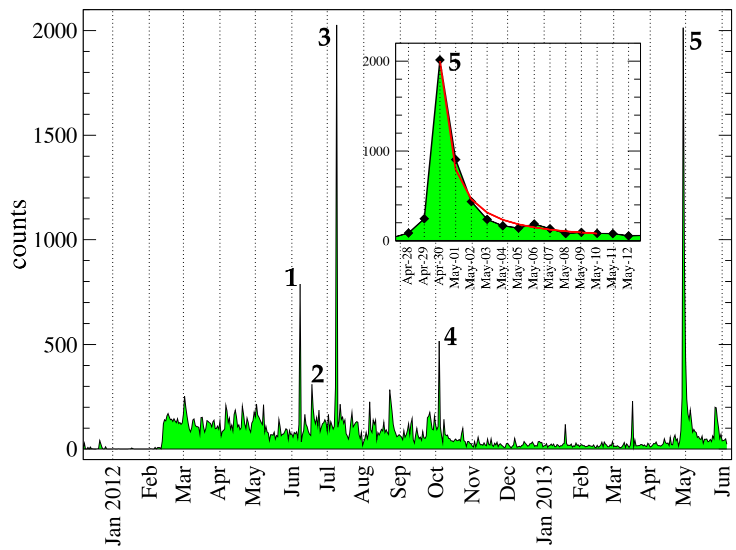


Figure 3.2: **Measurements per day.** Number of measurements collected each day from Dec. 8th 2011 till Jun. 6th 2013. The labels correspond to: (1) case study in Rome (9th June 2012); (2) launch of the Heathrow activities (19th June 2012); (3) Antwerp test case (10th July 2012); (4) Birmingham workshop (5th October 2012); (5) article in German regional newspaper (published 29th April 2013, activity peak on the 30th of April 2013). In the inset an enlarged view of event 5 is showed. The decay of user participation is consistent with a power-law of exponent $-\frac{4}{3}$ (red curve).

Figure 3.2 shows the number of measurements collected each day since December 2011. The higher spikes correspond to case studies or public advertising of our application. For instance, the first two spikes marked **1** and **2** correspond to the test case in Rome (9th June 2012) and to the launch of the Heathrow activities (19th June 2012) respectively. Another activity peak can be observed on the 10th of July, marked as **3**, corresponding to the Antwerp test case, while the peak number **4** on the 5th of October 2012 corresponds to the workshop organised in Birmingham. The activity in 2013 is boosted by the publication of an article in a German regional newspaper on the 29th of April that triggered a substantial measuring activity all over Germany, generating also a large peak on the 30th of April, labeled with number **5**. Given that this last activity spike was not produced by an organised test case, it is interesting to study the trend around it in detail. The inset in Figure 3.2 shows a possible power-law decay with exponent around -4/3 of the number of measurements in time. This means that the general interest is dropping fast, however several users maintain their activity for a longer time. Such power-law decay of human activities has been already reported for the occurrences of words related to special events in Japanese blogs [115]. In their case the power-law decay of extreme events broadcast by media (eg. *Tsunami*) was estimated with an exponent of -1.09 ± 0.45. Moreover, the power-law behavior we find, seems to be well described by the exogenous subcritical process described in Ref. [48].

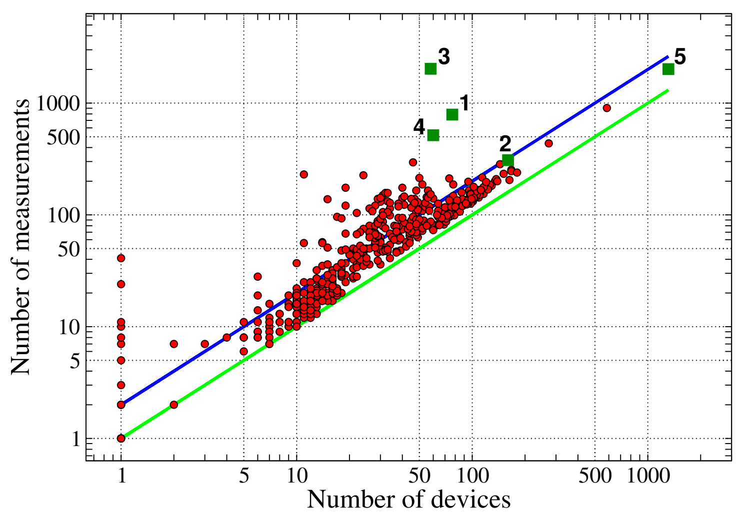


Figure 3.3: **Measurements versus devices.** Scatter plot of the number of measurements collected each day compared to the number of active devices at that day. The dark green symbols correspond to the most important spikes shown also in Figure 3.2. The green and blue lines are guides for the eye and correspond to the case of one measure per device and two measures per device respectively.

For more details on how these activities were generated, Figure 3.3 shows the number of contributing devices for each day, compared to the number of measurements. The points corresponding to the peaks observed in Figure 3.2 are clearly marked. In general, there appears to be a linear dependence between the number of devices per day and their corresponding measurements. The peaks corresponding to the Heathrow start of activity (marked 2) and the newspaper article (marked 5) fit within the linear dependence, due to the involvement of a larger amount of citizens in the measurements. Each device performs less than 2 measurements per day on average. Other test cases, such as the internal one in Antwerp, the one in Rome or the workshop in Birmingham (1, 3 and 4), did not attract as many users. However a large amount of measurements have still been obtained, shown by the deviation of the corresponding points from the larger mass of measurements. Here, each device performs much more measurements than in the previous case.

Thus, when analysing the days with higher number of measurements, we can distinguish between two types of activity trends: (i) large amounts of data generated by a small set of motivated users (peaks 1, 3 and 4) and (ii) large amounts of data generated by a correspondingly larger user base (peaks 2 and 5). The difference between the two types of behaviour can depend on several factors. For (i) the activities were goal-oriented and lasted for only one day. Users were asked to make an effort on the day with no incentives, except for the case study in Rome where the first three users with most measures were rewarded with a low value book gift voucher. This explains the large one-day activity per person. However, for the Heathrow case study as well as for the newspaper article, activities were not limited to a single day. Users did not have to concentrate all their measurements in a few hours, but were free to take samples over a longer period. It appears from the data that the natural average density for a single user is of two measurements per day. Additionally, the two types of activity trends were caused by different incentives. The activities in the first category were dedicated to covering as much area as possible and obtain a map of the daily noise. In the newspaper case, on the other hand, no incentive was given. Users were just measuring interesting noise levels. This is an event-based scenario that seems to favour a dilution of the measurements over multiple days.

General user activity patterns have been also studied. Figure 3.4 displays the distribution of the number of measurements submitted by individual users. This appears to be consistent with a power-law, with a large number of users submitting a small amount of measurements and, conversely,

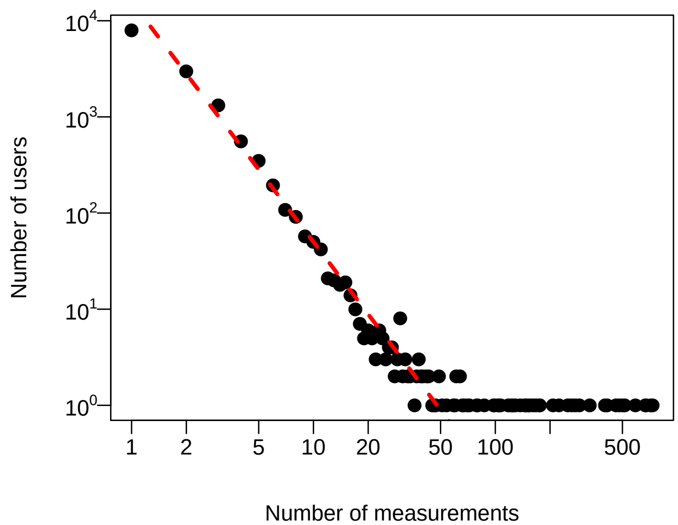


Figure 3.4: **User activity distribution.** Power-law compatible distribution of the number of measurements performed by each user. The red dashed line corresponds to a powerlaw of exponent -2.5 .

a small number of users performing a very large amount of measurements. This behaviour is the footprint of social activity, with the power-law distribution appearing in many other settings, for instance social network activity measures [64]. Hence the WideNoise user activity fits very well in the general known patterns of human behaviour.

While noise measurements can be performed with static monitors or by official agencies, distributed mapping of noise has the advantage of yielding a wider coverage in time and/or space. The data collected by WideNoise users come from across the world, with some areas better represented and covered than others. The general pattern is visualised in Figure 3.5. Europe and Asia emerge as the most active areas. While the large amount of data for Europe is understandable, since it is the area where all the test cases have been organised, the Asian activity stands out as emerging without any intervention from the researchers' side.

In order to assess awareness and learning, the subjective data submitted

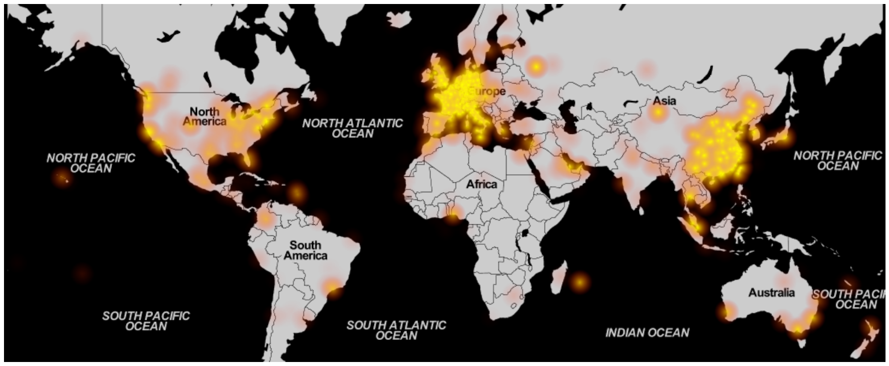


Figure 3.5: **Overall heatmap.** Worldwide sample density, including all measurements, illustrated as a heatmap.

by users are very important. However, not all measurements contain the additional data (tags, perception annotation).

3.3.5 Awareness arising

By means of the subjective data collected during measurements, an analysis of user awareness will be presented in the following. The interest is in assessing whether usage of the application leads to any change in behaviour, and whether this change indicates an increase in awareness of environmental noise and its effects. For this study, only data collected by users not belonging to the EveryAware consortium is considered (38267 measurements).

A first analysis of awareness/learning involves studying the decibel values estimated by users, in comparison with the measured values. Figure 3.6 displays the estimated vs real noise level, with light-coloured small points corresponding to early measurements by a single user, while dark large points corresponding to later measurements. Hence, the size and darkness of points displays user expertise. The figure shows larger darker points closer to the diagonal compared to lighter ones, which means that the estimation is closer to the measured value for later measurements. This indicates that during repeated usage of the application the ability of users to guess the noise level around them increases, hence the user learns in time.

To emphasise this point, Figure 3.7 shows the difference between the estimated and the real noise level as the users repeatedly perform measurements.

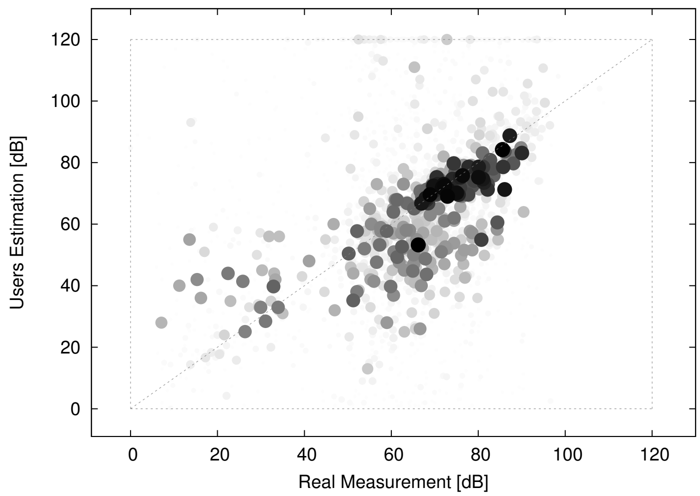


Figure 3.6: **Estimated versus measured noise.** Each point corresponds to one measurement, while both the colour scale light to dark grey and the point size represent the user expertise (small to large amount of previous measurements).

Averages and standard deviations are also displayed. This shows that as the expertise increases (number of measurements by the same user - horizontal axis), the errors become closer to zero and deviations from the mean decrease.

A different indicator of user involvement and hence awareness is the amount of tags submitted by users. An increase in repeated application usage would indicate increased involvement in data collection and hence increased awareness. Figure 3.8 displays the average number of tags per measurement, considering all measurements submitted to the platform, for increasing level of expertise (measurement number). At the same time, the number of users who have passed a certain expertise level is displayed. This shows that as the users perform more measurements, although the number of users here decreases, the average number of tags per measurement tends to increase. This demonstrates an increase in user involvement and dedication to the task, hence in the level of awareness.

A further analysis aims to compare the subjective perceptions (Love-Hate, Calm-Hectic, Nature-Man Made, Alone-Social) of the users with the measured noise levels. Out of all measurements performed, 12129 contain

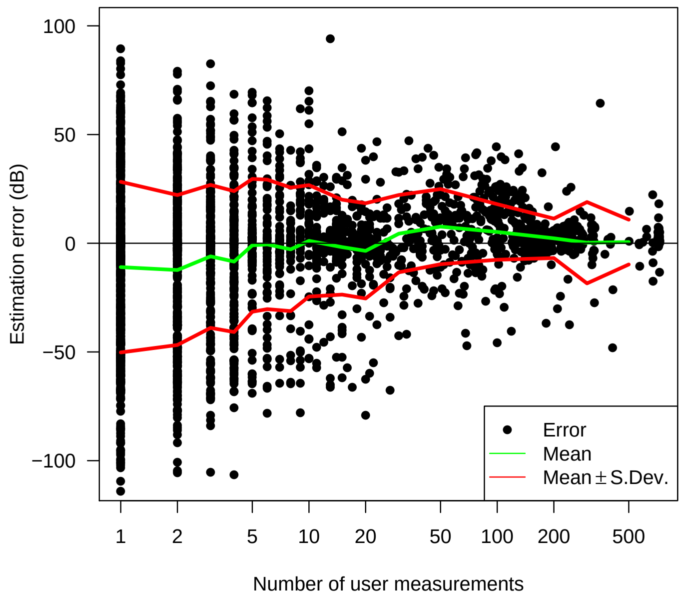


Figure 3.7: **Estimation error.** Difference between estimated and real dB value vs the number of measurements a user has performed.

perception data. Figure 3.9, shows how these perceptions depend on the measured noise levels. As expected, the perception values increase with noise. This means that, in general, users 'Love' quiet places, finding them a 'Calm' environment, while they 'Hate' loud ones finding them 'Hectic'. At the same time, high levels of noise are in general associated with Man-Made and Social environments.

To analyse the change in opinion as the user is exposed to the information from the application, i.e. the real noise level, Figure 3.9 includes two curves. One shows average perception levels for the first 5 measurements of every user, as a function of noise, while the other shows perceptions for measurements performed after some expertise has been gathered, i.e. more than 50 measurements. The two curves show a different behaviour for novice and expert users, for all perception types except for the Alone-Social evaluation. Specifically, noisy environments are perceived as less pleasant and

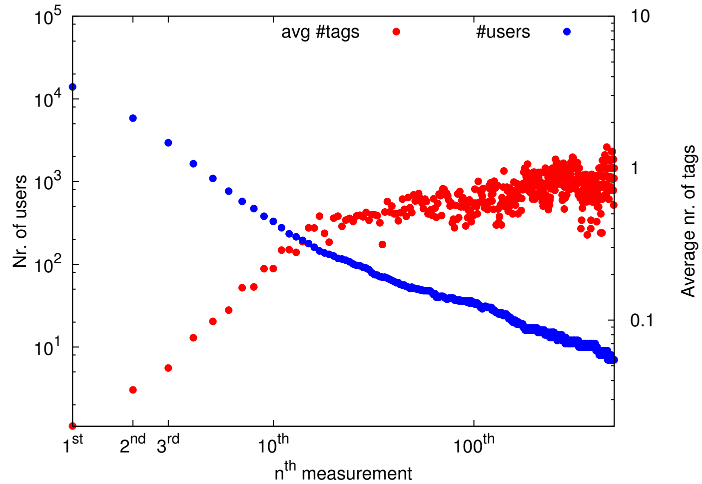


Figure 3.8: **Tagged measurements for different expertise levels.** The cumulative number of users submitting at least n measurements is displayed in blue (left axis legend), while the red points represent the average number of tags used in the n-th users' measurement (right axis legend).

more artificial as the users become more experienced, while quiet environments as more natural and lovable. A switch between the two possibilities is observed around 55-60 dB, for all three types of perceptions, indicating this as a threshold where noise becomes bothersome. This shows that indeed, exposure to information from the noise application does influence the way in which users perceive the environment. Experienced users have a more stringent evaluation of their environment, and stronger opinions about how much they love or hate the noise levels around. A categorisation of the noise levels appears to emerge, with plateaus visible for high and low levels of noise, when considering data from experienced users.

Although it cannot be excluded that experienced users might push the sliders to the extreme right or left edges so to minimize the cognitive effort inherent in judging the quality of noise, the voluntary act of modifying the slider position, by setting it away from the neutral central position, indicates the willingness in conveying a useful information. In that case, we would interprete the pushing of the sliders to the extremes as a conscious act of categorization of experienced users who got more confident with the App.

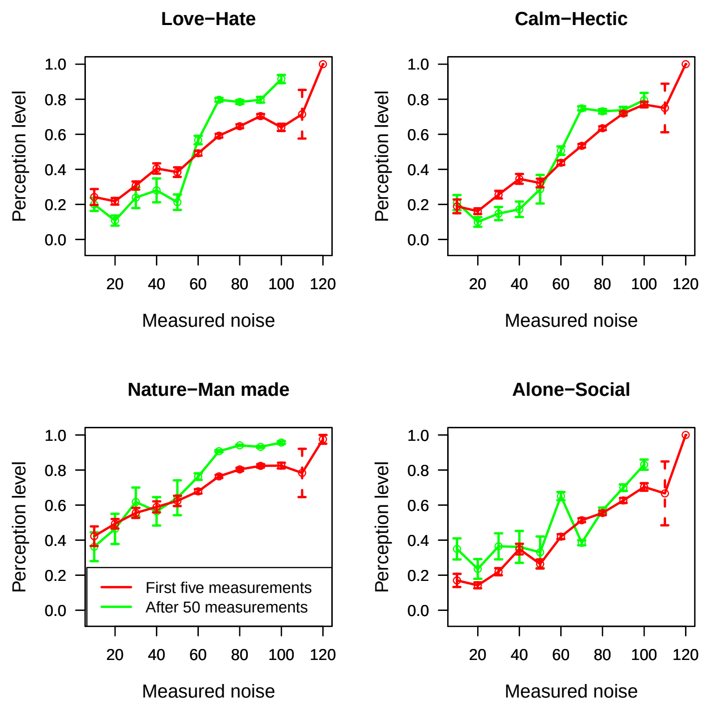


Figure 3.9: **Perception evaluation versus the measured noise level.** The red lines display the average evaluation over the first five measurements of all users; the green lines correspond to the average evaluation over the set of all measures taken by users starting from the 50th one.

As for the nature-man made indicator, we note that the typical user of our App lives in an urban environment, so that there are fewer samples collected in a natural environment and the error bars associated with the measures are consequently larger, possibly hiding the categorization effect seen in the other indicators at low dB values. The social aspect, however, does not change with repeated usage of the application, since knowing the noise levels does not affect the user's perception of how many individuals there are around. This explains why there is no definite difference between the two curves in Figure 3.9, lower right pane.

3.4 Why and how to build a web experiment

In this chapter the EveryAware project has been introduced. Other activities and results of the project will be described later. In this section, in particular, we focused on how the project provided the framework and the motivations for the implementation of a Social Computation experiment in a Real Complex System. The WideNoise participatory sensing experience helped to explain the issues to be faced when the aim is the exploit the power of the crowd, like in the HBC case presented in the previous section. The main difference is the scientific context and intention. With a clear research path in mind, we conducted this experiment gathering the data needed to analyze the emergence of the awareness, in the case of the noise pollution. For data collection, a mobile application has been developed and designed to gather objective noise data, while at the same time enabling users to contribute subjective data. To study awareness and learning, several indicators have been derived from the objective versus subjective data submitted by users, leading to the main findings we described so far:

- Guessed levels of noise, compared to the measured ones, indicate that users learn to estimate the noise level after repeated usage of the application.
- Perception rating is shown to change in time, as users perform more measurements. Hence noisy environments are qualified as more hectic and less lovable by experienced users, compared to novices.
- An increase in the fraction of tags submitted by users was observed as these became more experienced. This suggests an increase in involvement and dedication with time. Together with the change in perception, this indicated an increase in awareness after repeated usage of the WideNoise application.

This is the first study where a throughout parallel investigation of objective and subjective data has been performed, hopefully boosting an increase in awareness toward environmental issues.

Although initial signs of learning and increased awareness have been found already at this level, the usage of the application and evaluation of indicators such as those presented here will be continued in the future. Additionally, an in depth study of several data components is envisioned for future work, such as a semantic analysis of tags, which could give further important insight into both the motivation and opinion of users about their environment.

Beside the scientific findings, the WideNoise experience provide a clear evidence of the results achievable when exploiting the power of the crowd and of the ICT tools. At this point, a methodological question may emerge: what if after this experience we were going to implement a new web experiment? Then all the challenges faced in the WideNoise case should be faced again. For example, we should set up a new software platform, and we should start over the recruiting. But, since we solved that problems once, perhaps there could be a way to not solve them again. In other words, since the web, with its overlaid social structures, is becoming our lab, perhaps we should re-think our approach to web experiment, trying to find a more systematic way. And this is what we will try to do in the next chapter.

Chapter 4

An Experimental Web Framework: XTribe

So far we reported about two experience of use of the Internet as a mean to gather valuable and peculiar data: Human Brain Cloud and Widenoise. It is worth to note, once again, that the creation of equivalent dataset with traditional techniques would have been far more difficult, if not impossible. But even if the analysis and the results achieved with the data gathered are original, to our knowledge, the methodology is not innovative. There are several example of projects trying to adopt the crowdsourcing trough ICT tools in order to get scientific results [59, 82, 87, 131]. We will shortly present the state of art in the following section. Many of these initiatives represent successful experience and achieved important results. Yet, they remain isolated exploits without any cumulative growth or improvement of the methodology and of the tools.

In this chapter we present Experimental Tribe (XTribe in short), a novel general purpose web-based platform for web experiment based on social computation. Ready to use and already operational, XTribe aims at drastically reducing the effort required to develop and run web experiments. XTribe has been designed to speed up the implementation of those general aspects of web experiments that are independent of the specific experiment content. For example, XTribe takes care of user management by handling their registration and profiles and in case of multi-player games, it provides the necessary user grouping functionalities. XTribe also provides communication facilities to easily achieve both bidirectional and asynchronous communication. From a practical point of view, researchers are left with the only task of designing and implementing the game interface and logic of their experiment, on which they maintain full control. Moreover, XTribe acts as a repository of different scientific experiments, thus realizing a sort of showcase that stimulates users' curiosity, enhances their participation, and helps researchers in recruiting volunteers.

4.1 Web experiments: state of art

The Web is thus entangling in an unpredictable way cognitive, social and technological elements, giving rise in this way to the largest interconnected techno-social system ever. Social networking tools allow effective data and opinion collection and real-time information sharing processes. The possibility to access the digital fingerprints of individuals is opening tremendous avenues for an unprecedented monitoring at a "microscopic level" of collective phenomena involving human beings. We are thus moving very fast towards a sort of tomography of our societies, with a key contribution of people acting as data gathering "sensors" and with a level of fine-graining that only five years ago would have been considered science fiction. All this has deep implications for the understanding of the dynamics and evolution of our complex societies as well as for our ability to start making predictions and face the societal challenges of our era. Social Science disciplines, traditionally depending on the recruitment of test subjects to perform experiments, are for the first time experiencing the possibility to gather significant data in a very effective and capillary way, opening in this way the season of a computational social science [87].

In this context, the use of the Web for research purposes is changing the way research activities are conducted and how data are generated and gathered in many scientific fields. Despite the prediction, cast in 2009, that the new social platforms appearing on the Web might have become a very interesting laboratory for social sciences in general [87], Internet based research still lies in its infancy and methodological and procedural obstacles have to be faced in order to make it a reliable tool of investigation. Two paradigmatic examples are *Planet Hunters*[1] [59], a game in which participants can help in identifying new extra-solar planets using NASA data of star brightness and *Galaxy Zoo*[2] [116], in which players are asked to classify astronomic objects

[1] `http://www.planethunters.org`
[2] `http://www.galaxyzoo.org`

of galactic type, by browsing a catalogue of telescopic images. The above mentioned projects have in common the involvement of individual volunteers or networks of volunteers, many of whom may have non specific scientific training, to perform or manage research related tasks in scientific projects. In this sense these are two examples of *citizen science* [26, 66, 108], i.e., a long-standing series of programs traditionally employing volunteer monitoring for natural resource management.

Citizen science projects are becoming increasingly focused on scientific research [47, 103, 114] and amazing results have already been obtained. For example, the 3D structure of viral enzymes that challenged scientists for years has been discovered thanks to the efforts of Foldit[3] players [82], new candidate planets identified by Planet Hunters' participants managed to survive data verification tests [59], and brand new astronomical objects were discovered by Galaxy Zoo's users [116]. These examples show how social computation processes hold tremendous potential to solve a variety of problems in novel and interesting ways, and how amateur players are able to solve research problems, even faster than their professional researchers counterparts. Human ability to easily solve tasks that are difficult to solve by machines has been largely exploited for instance in labelling images, through the collaborative *ESP Game* [132], or in language automatic translators, through the interactive learning platform *Duolingo*[4]. In these last two examples, the idea of linking playful activities with learning processes has led to the paradigm of Games With a Purpose (GWAP) [131], i.e. a way of engaging people in games that can extract valuable information or work as a side effect of the game or the learning dynamics. The playful rearranging of experiments, together with their appealing graphic interfaces, has also proved to be a fundamental ingredient for web-based experiments design, boosting user participation and data reliability.

This idea of *crowdsourcing*, term coined in 2006 [75], is also at the heart of on-line labour markets such as Amazon Mechanical Turk (AMT), where a job is distributed by employers in small sub-tasks that on-line workers can perform in return of proportionally small monetary payoffs. Interestingly, despite its mercenary aspect, AMT has proven to be useful for scientific purposes [43,94,105], by leveraging on its ease in recruiting a potentially large

[3] `http://fold.it`
[4] `http://duolingo.com`

number of experimental subjects. This early experience with crowdsourced experiments has led to the recognition that Web experiments, despite the unavoidable partial control on the way participants are recruited and on the context in which tasks are executed, can be successfully used to study human collective behaviour and cognition and can provide elements of validation of experimental practices in the Web [123].

The tenets of social computation are being increasingly exploited, but its use in the scientific community still lacks systematization. The realization of a single project often requires substantial effort and web-based experiments are still far from being standard research tools. The lack of tools that can greatly simplify and standardize the design of Web games and experiments is a major bottleneck in the exploitation of such new research opportunities. For example, despite its versatility, AMT has not been conceived as an experimental platform, lacking dedicated infrastructures for the design of experiments, while offering some visual tools to develop simple interfaces. Beside this, AMT also show other kind of problem, linked to the economic motivation of the users. This argument will be specifically analyzed in the following chapter. Since many web experiments in Social Science or in Economics just consist in simple surveys, in recent years a lot of platform offering guided procedure to creation and hosting of web survey implementation emerged, like Qualtrics[5] or SurveyMonkey[6]. This instruments, developed initially for marketing web research and polls, have been successfully adopted in several scientific experiments. More recently, new tools and library to help build independent surveys website appeared, e.g. PyBossa[7] and PlanOut[8]. These remarkable initiative testify the need for systematization and new instruments, and represent powerful tools, but still the experimental approach may be optimized in several ways.

Experimentalists are left with the task of designing their own software solutions to manage interactions among participants and to build effective interfaces. Moreover, individual solutions to such problems often remain isolated with little or no cumulative growth of tools and ideas. Hence the need of a versatile platform to implement web-based experiments with a very small coding effort. This is the aim of XTribe[9], a general purpose platform to carry

[5]`http://www.qualtrics.com`

[6]`https://it.surveymonkey.com/`

[7]`http://pybossa.com/`

[8]`https://facebook.github.io/planout/`

[9]Active and available at `http://www.xtribe.eu`

on experiments in the form of web-games. The word "game" is here intended as a real-time interaction protocol among few players implementing a specific task, as well as a synonym of experiment on interactive behaviour, typically leveraging the ludic motivation. By providing the scientific community with a general purpose platform for social-computation and web-gaming, XTribe gathers otherwise separate efforts to use Web resources for scientific purposes and provides the community with a tool to design experiments on the Web, from simple polls to very complex multiplayer games, bypassing much of the "hard work", e.g. hosting, user registry handling and user pairing/grouping, communication protocols, exceptions handling, etc.

4.2 Anatomy of an abstract online experiment

The GWAP applications and the all the others web experiments cited above show a vast variety of features and a very heterogeneous set of targets. But even these motley experiences have several elements in common, beside the general idea of exploiting the force of the crowd. In order to introduce the necessary steps to build a web experiment, in this section we shall analyze the structural and technical components of a generic web experiment, from an abstract point of view without going into detailed technicalities. As a guide we shall consider here the structure of the ESP Game, the most famous GWAP. In this game, two players are asked to tag the same image, trying to match their tags. They will input as many tags as they want until one tag is in common to both; then they move to the next image. Within a time limit of 2.5 minutes, the players have to agree on as many images as possible, to increase their score. The goal of the game from the experimenter perspective is to obtain realistic valuable tags for online images, to be used by search engines. We shall consider this game as a prototype that will make the analysis of the typical game components more clear.

At one extreme of our abstract structure lies the *developer*, i.e. a researcher willing to create a web experiment. At the other end lies the *community*, i.e. the ensemble of users who will play the game. Depending on the experiment, this can be a wide community or a subset filtered by age, gender, language, interests or even geographical location. In the case of ESP, those are the players who tag images. Developer and users are just the two ends of a complex structure and in the following subsections we shall describe what lies in the middle and permits the execution of the game.

4.2.1 The interface: interacting with the user

In web experiments there is a flow of information that, in most cases, starts from the user, e.g. in response to a given question ("how will the other player tag this picture?", in the ESP Game case). Therefore, the application will need a user interface allowing players to insert their answers. The interface should be designed by researchers with the goal of optimizing users' experience, ensuring an easy and enjoyable interaction. The user has to invest time in paying attention to the application and the entertainment itself offered by the interface can be a reward for the user interaction, and engagement may work better than economic motivation. Moreover, a successful interface design will not only persuade the user to spend time on the application but will also stimulate her to involve other people. A well designed interface should also help her in voluntary recruiting acquaintances, e.g. by leveraging on social networks features, such as tweets about the results, Facebook sharing of the results, etc. Even if the fanciness of the interface is crucial, the designer has always to keep in mind the biases introduced by the interface. Each kind of interaction introduces biases, even the simple fact that users are interacting through a computer. As we said, the reliability of the information gathered is a fundamental point. Thus, the impact of each bias introduced has to be carefully considered in order to find a good compromise between the reliability of results and the user experience. In the following chapter we will see an example of how easy can be to introduce bias, and how to take care of the consequences.

4.2.2 The server side logic and data storage

Once the information has been gathered by the interface, in order to give feedback to users or results to the developer, it is very likely that some elaborations will be needed. So the application will need a logic elaboration part. In the ESP case, the logic component receives the tags from each of the two players, compares them and when a match is found, feeds the interface with a new image to be labeled. When the time is over, this component computes a score and sends it to both players. While the interface runs on the browser, i.e. on the user computer, the information processing should happen server side, in order to guarantee reliability and security (reducing the risk of failures, cheating and hacking). Moreover, the game logic may require complex computation involving data that the researcher cannot or does not want to make available to the user browser. Beside this, there is also a matter of control: the logic part has to be directly managed by the developer, the other end of our scheme. Hence, it should run on a machine under the developer control where all data generated by the experiment can be properly stored for further research and analysis. The logic part will also provide content for the application (e.g., pictures in the ESP Game). In other words, the logic part will take care of filling the interface with input and feedback, as well as of gathering results.

4.2.3 The rest: technical but necessary issues

The interface and the logic are the nearest neighbor of the user and of the developer, respectively. These two parts are the core of the application, the "unique" parts designed by researchers precisely on the project target. But the application itself it is still far from being complete. There are at least three missing fundamental parts:

1. a communication protocol between the two parts;

2. a user handling system;

3. an instance processing mechanism.

Communication

The communication between the interface and the logic is potentially very difficult to implement. If we consider the simple case of a client initiating the communication by sending a message to a server, the solution is quite easy to carry out (e.g., with a HTTP request). But in case of more complex communication structures, such as bidirectional asynchronous client-server communication or, in multiplayer games, client-client communication, the implementation can be quite a difficult task requiring more sophisticated technologies (e.g., web-sockets).

User handling

When dealing with users, a certain set of functionalities is likely to be useful such as user registration handling and profile management. At a basic level, it is a matter of security and reliability, because registration can provide a first filters against bots. Beside this, many experiments require a certain level of profiling of the users, to differentiate or group them depending on the gender, age, language, etc. On the other side, users may enjoy to see the result of their efforts, in the form of scores, ranks, etc. So they would prefer their "player" identity to be recorded by the game. Obviously, linked to this, there are also privacy issues: the developer has to guarantee to the user that his personal data will not be disclosed.

Instances

Once the interface has been prepared, the logic is running, they are communicating and the user is registered (if required), an instance of the game still has to be created, in order to allow the user to join the experiment. By instance we mean the single execution of the experiment task involving one or more users. This management is relatively easy for single player games, but it becomes non-trivial in case of multiplayer games. A "waiting room" has to be implemented, in order to make the users wait for others to join. This system have also to implement security checks (e.g. a users should not be able to play a two players game against himself).

These three parts have two things in common. They are needed (if not all necessary they are at least all very useful) in almost every kind of web experiment and are not particularly influenced by the specific experiment or game. Hence, since this three parts are almost unrelated to the experiment, they are the most technical and dull to implement. That is why a framework or, even better, a platform that can take care of these functionalities automatically would make it easier to create web experiments. This is where the XTribe platform comes in, to provide the technical "middleware" (i.e. Sec. 4.2.3) and allow the author of the game to focus on the game-specific interface and logic (i.e. Sec. 4.2.1 and 4.2.2). But the aims of the XTribe platform are not limited to these.

4.3 XTribe platform in detail

The XTribe platform has been designed with a modular structure so that most of the complexity associated to running an experiment is hidden into a Main Server (called Experimental Tribe Server or *ET Server* for short). In this way most of the coding difficulties related to the realization of a dynamic web application are already taken care by the ET Server and the realization of an experiment should be as easy as constructing a webpage with the main utilities for it. There are different kinds of users of the platform: the system administrator who runs the whole ET Server and provides all the necessary API's for it; the experimentalists who run individual experiments; and the players who participate in one or more experiments or games. Each experiment is conceived by the game developers/researchers who monitor the evolution through their local machines. Games have two components: the user interface (UI) and the logic - game manager (GM). The interface is what is visible to players, and will interact with them. The GM is represented by those functional parts that process the action of the players in order to implement coordination and specific game logics. *These two components (the UI and GM) are the only things on which the researchers have to focus their development efforts*, since they are highly dependent on the game itself. XTribe mediates the communication between the two and hosts the game interface. The GM part of the game is hosted by the researchers on their own server. In this way they can directly collect the data in real time and have full control over the experiment progression. It is important to remark that XTribe does not store the data coming from the hosted experiments. All scientific data collected during an experiment can be conveniently stored by the GM, so that only the researcher who developed and published the experiment benefits of the outcome of his/her work. Beside this, gathering data directly grants the opportunity to analyze them as soon as they enter the system in real time, and even use them in other instances of the same experiment.

The XTribe platform also offers the hosting and a guided coding-free procedure for the creation and the publication of a page for the description of the game rules and scientific purposes. The researcher can fill this page with all the information to be provided to the users and from this page players can access and play the game. Additionally, it handles player/user management (registration, authentication and profiling) and manages the actual instances of each experiment (creation, user grouping, error handling, feedback to users

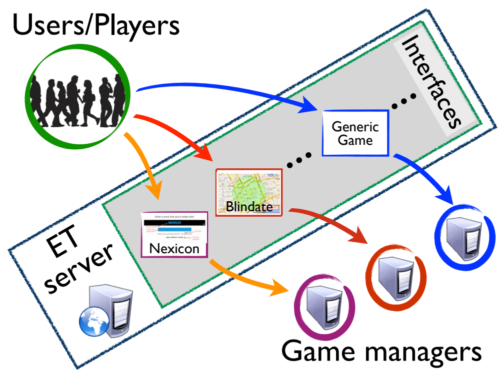


Figure 4.1: A graphical representation of the system and its interactions.

and managers, etc.). A graphical representation of the platform is depicted in Figure 4.1.

4.3.1 User management and community

Since experiments are created for research purposes, the researchers are interested in many types of statistics related to players. Beside this, they may also be interested in filtering players for specific purposes, e.g. according to their age, gender, language, geographical location, etc. To this aim, XTribe handles a user registry in which players will be allowed to register, if required, and play while the system maintains all the information about them, such as scores, ranks, game settings, leaderboards, etc. together with profile information. If needed, this information can be sent to the GM, i.e. to the experimentalist. Furthermore, based on this information, when properly configured, the system will grant the access to the game only to certain profiles. Being in charge of the handling of the user registry, the system would also spare the researcher from dealing with privacy and security issues since all data will be properly anonymized and, possibly, encrypted. However, by default, it is still possible for unregistered users to access the games. Filters are applied only if set by the researcher.

4.3.2 Communication made easy

The communication between the UI and the GM is mediated by the ET Server through a message based protocol. The general functionality of a game can be summarised with the following flow:

- Once the players have accessed the game, the system will create an instance of the game. There may be given rules for the game to start. A basic rule is the number of players. There may also be different constraints, e.g., pair players with similar scores or players playing from different geographical locations. As soon as there is a sufficient number of players satisfying the grouping constraints, an instance of the game starts.

- The interface will transmit the actions of the players to the GM, but all messages will pass through the system, which will group them by match instance number after having anonymized them.

- The GM will then receive the data, will elaborate them and will send the results of the elaboration back to the system, which in turn will transmit them to the UI of the various players. Obviously, the GM will also save the data of interest locally (as it runs on the researcher's machine).

It is important to remark that the GM can send messages to the UI either as a response of a message coming from a player (responding to that player, to the others or broadcasting to all of them) or by initiating the connection autonomously (e.g., after a given time). The platform will also handle errors and exceptions. For instance, if one of the players disconnects unexpectedly, the system will detect and notify it to the remaining players and will send a message to the GM. Since there is no direct communication between GM and interface, the GM will experience no trouble at all.

In Figure 4.2 we depicted the communication flow of a two player game: a first player joins the experiment and waits for the second one to come. When both players are there an instance is created and the player's browsers are instructed to load the game UI. When loading is completed the UI notifies XTribe which in turn notifies the GM. Up to this point everything is automatic. The GM will probably send custom data back to the players to let the game start. During the game custom data are exchanged between UI and GM, until the game is over and the instance is closed.

All these features, especially the user registry and the instance handling, usually require a lot of coding, quantified in time and money, to be realized.

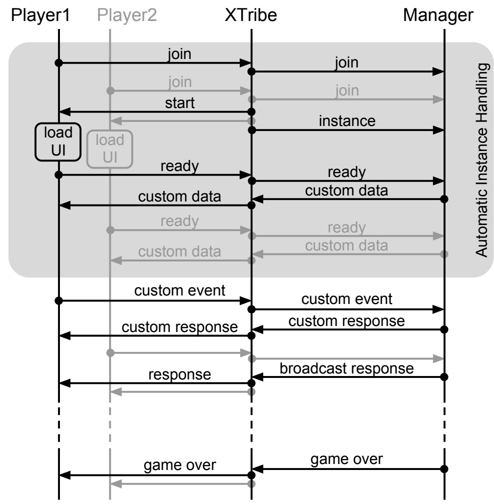


Figure 4.2: The communication flow of a two-player game on XTribe.

Within XTribe, they can be realized with a straightforward procedure. After the configuration, the system will automatically take care of all. What researchers have to do is to write the code of the UI and of the GM only.

The UI has to be structured as a web page with plenty of freedom in using HTML, CSS, Flash, etc., while the interaction between the interface and the system has to be achieved by means of the ET Server API, which are internally developed as Javascript functions. With this simple set of functions the interface will interact with the platform and, through it, with the GM. Basically, the GM has to work as a simple HTTP server hosted on the researcher's machine. The communication with the system takes place through the HTTP protocol and all messages are coded in JSON format. The GM receives the message as a POST string variable and sends back one or more messages with a JSON string in the response body. Besides a restricted set of system messages, the researcher is given full freedom to decide custom messages for the internal game protocol.

4.3.3 Social network integration

Since the strength of online games comes from large participation, the XTribe platform has been integrated with the most powerful online social network application, *Facebook*. Through Facebook the recruiting of new users is easier, since the new platform can spread through the network faster. The integration consists of the possibility to view the XTribe interface within the Facebook website and play games as Facebook games. Additionally, it provides seamless user registration, integrating the Facebook user information with the XTribe user registry. Hence, players have a better user experience connecting to XTribe without having to insert their information again, while researchers can collect more demographic information about the players of their games. Regular posts on user activity on the platform are published on user walls, and in this way additional players can be attracted to the system. Researchers wishing to build new games take advantage of this integration without any additional effort from their side.

XTribe can be used in conjunction with the Amazon Mechanical Turk (AMT) platform in order to exploit its ability to recruit users with a modest monetary investment. AMT can be used to enhance participation and possibly in the initial phase of an experiment, to provide the necessary pool of data to begin with. The integration has been implemented by simply releasing an AMT payment code at the end of every single match or experiment. This strategy has been used in an experiment described in the following chapter.

4.3.4 Community knowledge cumulative growth

One of the potential benefits of joining an experimental platform it is the community effect. XTribe offers a set of instruments to support the community of researcher. In first place, a forum has been implemented, to allow them to ask questions and share their practice with the others. More importantly, those willing to develop experiments can freely access to the set of web experiments codes shared by XTribe researcher. In fact, the platform consent to share the experiment code with a single click on a revision control web-platform, BitBucket[10]. Those features, recently introduced on the platform, are expected to boost platform usage and cooperation between researchers, by sharing of code and best experimental practice.

4.4 XTribe for developers

The XTribe platform already features several experiments mainly about language, map perception, and opinion dynamics. We developed also a sandbox

[10] `https://bitbucket.org`

XTribe platform	`http://www.xtribe.eu`
Documentation	`http://doc.xtribe.eu`
Test platform	`http://lab.xtribe.eu`

Table 4.1: Important XTribe related URLs.

version of the platform available at `http://lab.xtribe.eu`, where experiments can be tested during their development phase. All important XTribe related URLs are summarised in Table 4.1.

In the following we will briefly describe how implementing a multiplayer experiment on XTribe is a matter of hours, provided that the developers have basic knowledge of HTML, Javascript, and any server side language. We will use as a sample a well-known game in game theory, used in many experiments [113], the Minority Game, described below.

Minority Game. The game requires three players, who are presented with two choices (e.g., two numbers or two amounts of money). Each player has to choose one of the two options. If all of them agree on a single choice nobody wins and nothing happens. If only two players agree on their choice, they lose while the other player wins the amount they chose. A working version of the Minority Game can be played on `http://lab.xtribe.eu`.

In this very simple experiment a scientist may be interested in comparing user choices when numbers are shown with or without a currency sign or whether changing the ratio or the order of magnitude of the amounts results in behavioural shifts. Implementing this as a web game from scratch would require a lot of effort in managing waiting rooms for users in order to group them 3 by 3. Moreover, the players make their choices asynchronously. This means that when the last player makes the choice the game terminates and the server has to contact the other two players with the result: handling server to client communication within standard HTML pages is not easy and requires efforts in the implementation and even in the server side configuration[11]. On XTribe the experimenter is left with the only duty of implementing the HMTL interface and a server side script that, given a group of three players (identified by the platform with a unique instance number), chooses the two values to show, collects answers, and determines the winner

[11]Consider for example the fact that a user firewall may block communications on ports other than 80. To bypass this, all communication (including websockets) has to be routed through well known ports.

(broadcasting the result to the players).

User Interface. The implementation starts with creating an HTML page with two buttons, one for each choice. The developer then has to include, in the page head, the script `Client.js` which makes ET Server API available to the experiment. The game UI will be hosted on the XTribe server where the API is. In order to use the ET Server API it is enough to instantiate it:

```
c = new ETS.Client();
```

and then register a user defined callback function to receive messages:

```
c.receive('manager', myFun);
```

The UI can receive messages both from the `manager` or from the `system` (especially useful to handle errors). Each message has a `topic` (a string describing what the message is about) and a `params` field which can contain arbitrary data about the message. In our game the manager will send to the user two types of messages, one with the two choices at the beginning of the game, and one with the result at the end of the game. Consider these to be `mgChoices` and `mgResult` respectively.

A possible implementation of `myFun` would be:

```
function myFun(msg) {
  switch(msg.topic) {
    case 'mgChoices':
      play(msg.params[0], msg.params[1]);
      break;
    case 'mgResult':
      answer(msg.params);
      break;
  }
});
```

where `play` and `answer` are two user defined functions that change the HTML page according to the received values: the first one will fill the two buttons with the proper values chosen by the game manager, while the second one will show a message to the users depending on whether they win or lose.

The last thing that should be implemented on the UI is sending a message to the manager in response of a user interaction (i.e., button click). This can be easily achieved with ET Server API as follows:

```
c.send('manager', 'mgUChoice', v);
```

where `mgUChoice` is the arbitrary topic that describe this kind of message and `v` is a variable that refers to what the user chose.

Game Manager. The game manager runs on the experimenter server and can be implemented in any programming language (we use PHP in this example). It receives messages from XTribe as a POST variable exactly as a common script receives strings from an HTML form. The variable name is `message` and it is a JSON encoded structure:

```
$msg = json_decode($_POST['message']);
```

Looking at the `sender` (either `system` or `client`) and `topic` fields of the message, the manager will be able to take proper actions. Other relevant fields of the received message that have to be used are `instanceId` and `clientId`. These are two numbers generated by XTribe that univocally identify the instance this message refers to and the user who sent this message (if applicable).

In our game, as soon as three players joined the experiment, XTribe creates a new instance and notifies this to the manager with a message with `sender = system` and `topic = instance`. This is the perfect moment to generate the two values these three players will be playing with. These values can be stored, in association with the provided `instanceId` in a database table or in some persistent data structure (easier for GM written with Java, NodeJS, Python). The players will be loading the HTML interface in the mean time. As soon a each player is ready, this event is automatically notified to XTribe and in turn to the manager. The manager can then send a first message to the player, in our case a message with topic `mgChoices` with the two values as `params` (e.g., as an array). To send a message back to XTribe the manager simply writes it (as a JSON encoded string) in the body of the response: it is as easy as returning plain text:

```
$resMsg = array(
    'recipient' => 'client',
    'topic' => 'mgChoices',
    'clientId' => $msg.clientId,
    'instanceId' => $msg.instanceId,
    'params' => array(v1, v2)
  );
print(json_encode($resMsg));
```

Each time the manager receives a message with topic `mgUChoice` it stores the user choice updating the database or the persistent data structure. No response is required for the first two players, but when the third one answers the manager computes the winner and sends a broadcast `mgResult` message back to all users plus an `over` message to the system to inform it that this instance is over. Both messages can be sent together as an array.

```
$resMsg[0] = array(
    'recipient' => 'client',
    'topic' => 'mgChoices',
    'broadcast' => true,
    'instanceId' => $msg.instanceId,
    'params' => winner
  );
$resMsg[1] = array(
    'recipient' => 'system',
    'topic' => 'over',
    'instanceId' => $msg.instanceId,
  );
print(json_encode($resMsg));
```

Optionally, with the `over` message, the manager can provide a score for each player that will be used for the game leaderboard automatically managed by XTribe.

Deploy the game. Once the UI and GM are ready the experimenters will simply create a new experiment on XTribe, providing basic information such as game name, description, icon, screenshots, number of players, etc. Then they will simply upload all the UI files on XTribe and provide a URL to contact the GM running on their server. It is worth to note that, even if the procedure can be non-trivial for people with basilar coding skills, with the right development tools the whole code may be of more or less 100 lines. 100 lines to have a working web experiment without taking care of security or privacy issues, users profiling and registration and instance creation.

4.5 XTribe so far, and over

XTribe is a general purpose platform that handles all the aspects of the realization of web experiments that do not concern directly the game itself.

In this way, it allows researchers to focus only on the core of the experiment, leaving the rest to the system.

An important result of the XTribe project is to allow researchers working in different fields, who lack computer science expertise, to create web-based experiments and games with a easier and guided procedure. In order to further facilitate this, the next step is to create a set of "default" GMs for games corresponding to the most standard types of web experiment, such as surveys or coordination games. For the time being, there is a default GM available that broadcasts to all the players the messages received from each one. The following step will be the realization of a set of graphical tools that will make it possible to set up a web experiment without writing a single line of code, e.g. a drag and drop-like system that allows the development of interfaces and the creation of the relative logics.

As already stressed, the platform is expected to act as a reference point for interested users, giving a fundamental boost in facing a typical issue related to web experiments: the recruitment. It is often quite difficult to gather a critical mass of "suitable" players, and this can be an easier task for an organized and collective platform than for single games. A first step towards facilitating recruitment was Facebook integration. In time, this process will become easier for new games. Since they are hosted on the platform, and shown on its main page, other players already involved in other games would probably join, attracted by curiosity. We expect a community of players to gather around XTribe playing different games and also giving researchers feedback about their experiments. We also expect researchers to aggregate into communities, sharing advices and best experimental practices with each other.

The platform has proven its usefulness with several games already implemented by different researchers. The already existing games refer to studies in language and opinion dynamics, where the human component plays a crucial role, and are designed as web based social experiments. They show the versatility of the platform and its ability to host experiments on a diverse range of topics, as words association games, citizen mapping, response of individuals to traffic information, expressing political opinions. These are prototype experiments where issues concerning different aspects related to results reliability and to the recruitment ability of the platform, as well as of single games, can be addressed. Besides their immediate scientific interest, they are meant to open the way to the use of this online laboratory, also involving other potentially interested research groups. In the next chapter we will report about some notable experimental experience derived from the XTribe platform.

Chapter 5

An Overview of the main XTribe Web-Experiments

In this chapter we will introduce some of the experimental experience we implemented on the XTribe platform. For each, we will introduce the scientific context and we will explain the experimental setup and the analysis of the results. Beside the achievements of understanding in each framework, a portrait of the approach to the web as a lab will emerge, with its opportunities and its threats. We will not focus (if not occasionally) on the advantages resulting from the adoption of the XTribe platform, since they have been introduced at a general level in the previous chapter. At a more specific level, they consist mainly in coding simplification whose detailed explanation is far beyond the purposes of this thesis. We will simply report that each of the experiments showed here have been developed by 1-2 developer in few weeks. Without XTribe, the implementation effort would have required at least a doubled manpower for a period of months. Thus, the successful outcome of these experiments with a quite small effort represents itself an encouraging proof of the XTribe potential.

5.1 CityRace: measure of traffic-facing strategies

A first test case using the XTribe platform is introduced through CityRace, a web experiment for the analysis of driving behaviour in the presence of traffic information. CityRace constitutes a virtual social laboratory, where different aspects of human behaviour in response to available traffic data can be analysed. Although several tools for displaying traffic information exist (e.g. Google Maps[1], Autostrade per lItalia[2], Waze[3]), their effect on driver behaviour has not been at all analysed. Such a study may uncover important aspects of how much information a user needs and in what context the information is useful. We propose to use the GWAP approach for this, and describe here the results obtained after the first test cases. The XTribe platform has proven to be very effective through its flexibility which has facilitated the implementation of the web game. Beside this, the CityRace experience is meaningful also for methodological reasons. Two different communities of users were involved. The first group was composed of participants to a demonstrative event organized in a bookstore in Rome, together with other players who joined the experiment on their own by surfing the web or through Facebook. The second group was composed of workers recruited from the Amazon Mechanical Turk (AMT)[4] virtual labour market. These two groups took part in the experiment with different goals. The first group joined the experiment with entertaining purposes, while the second was primarily interested in maximizing their monetary income. Despite their strong bias towards profit, AMT users were previously proven to be sufficiently reliable for scientific purposes and to produce meaningful data connected to simple experiments [105]. Our case is interesting because, in order to play CityRace correctly, a significant cognitive effort is required, so that the question of AMT user reliability for such demanding tasks is not yet assessed.

Although CityRace has been considered here as a test case for the XTribe platform, it is connected to interesting questions. One such question is whether displaying local traffic information changes user behaviour, and under which conditions. Also, how much information is needed to trigger a change in behaviour? In real life, how much information on the state of immediate neighbourhood is required for the citizen to be able to optimize the route? For this, the platform displays different amounts of traffic information in each game, which will allow for an analysis in this direction to be performed. This is useful both from the social science point of view and for optimising future applications that offer visualisation of routing and traffic information.

[1] http://maps.google.com
[2] http://www.autostrade.it
[3] https://www.waze.com
[4] https://www.mturk.com

Furthermore, a virtual traffic dataset is generated, based on the routes selected by users. This, analysed in comparison to the real data, can enable identification of traffic features related to street network topology. Also, the overall response to the real traffic displayed can be studied, showing whether avoidance of traffic can create jams in other locations of the city.

5.1.1 CityRace Solo: a urban browsing game

CityRace[5] as been developed within the XTribe framework, and aims to analyse the effect of traffic information, and the extent of it, on the routes players choose. For this, the Google Maps API has been used to display different areas in a city. In the Solo version of CityRace the player is shown two points, A and B, and has to choose a driving route between the two. There is also a Duo version which allows two players to compete each other in real time as a multiplayer game, however it is not within the aims of this thesis.

In order to enable the study of the effect of traffic, the game consists of two stages. In the first stage, the map is shown without any traffic information, and users have to draw the route between A and B by selecting successive points on the map (within an active green area with the side of about 600m). Once the route is completed, the second phase begins. The player has to draw a route again between the same two points, but with traffic information displayed on the map, colour coded for each street segment (red - busy to green - light traffic). The user can select the same route as in phase 1, or change the strategy, as an effect of having traffic information available. Fig. 5.1 shows the user interface of the game, with its different features.

Since we are also interested in how the amount of information affects behaviour, traffic information can have different sizes (i.e. from a small to a large square area around the current location of the player). The possible widths of the traffic information area are: 400m, 900m and 1300m. Also, the points A and B are generated randomly at different distances: approximately 550m, 800m, 1100m, to analyse whether the length of the route also affects user behaviour.

At the end of the game, a score is computed for each phase. This takes into account the driving distances and times, using the traffic information available. Distance and time are compared with the best Google route, from

[5] `http://www.xtribe.eu/node/41`

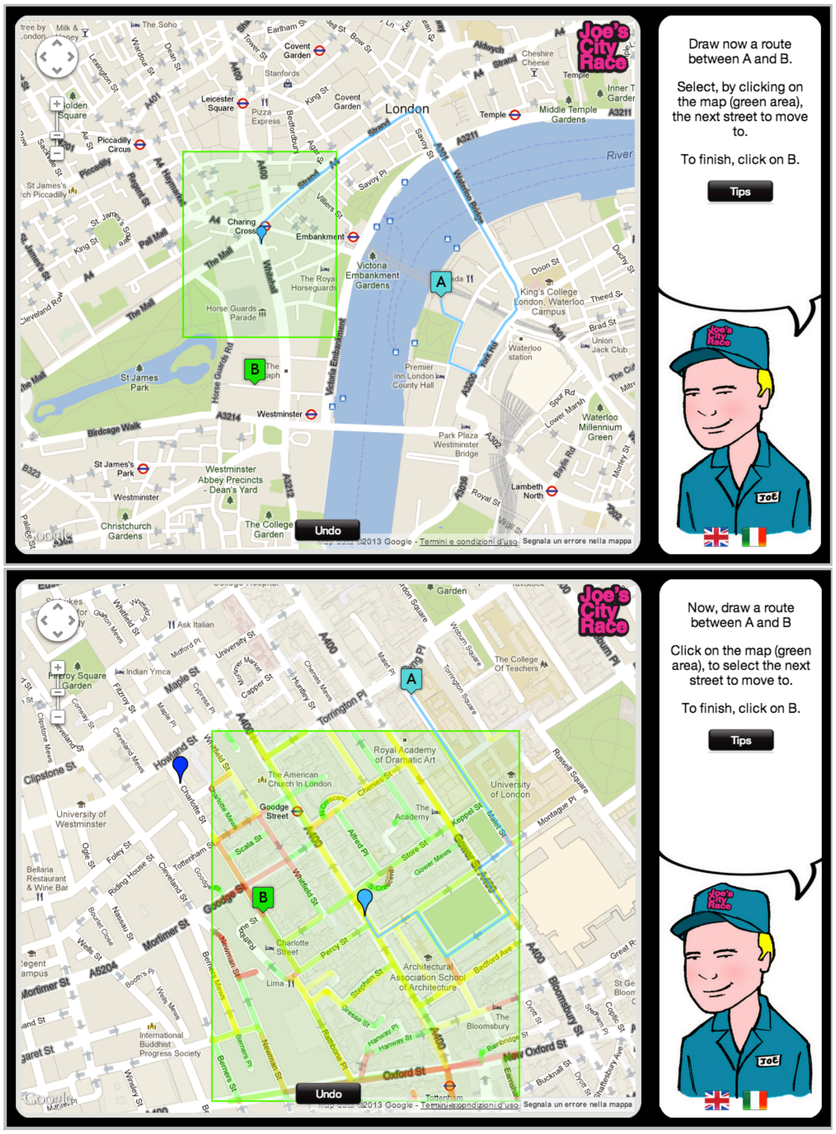


Figure 5.1: The two stages of the game interface.

the Google Direction engine. A high score (over 100) indicates that the user has overperformed Google. The final score of the game is the average over the two phases.

The game can be played in four locations: Turin, Dublin, Rome and London. For the first two, real traffic information has been obtained and shown on the map, averaged for different times of the day. Traffic data for Turin have been courteously prepared by the group of Prof. Rambaldi, at the Physics Dept. of Bologna University. These data contain average speeds recorded by GPS monitors on different street segments in Turin. Dublin data have been obtained from the Dublin City Council, through the Science Gallery Trinity College. These contain traffic counts and speed limits for different road segments in Dublin, which have been transformed into velocities using linear interpolation. For the other two cities, we generated synthetic datasets based on the Google Directions engine. Therefore, these do not fully correspond to the real averaged traffic, however they do provide a good indication of slower and faster roads (information which is included in the Google Directions framework [133]).

5.1.2 Test cases and data gathering

The City Race game has been available online since May 2012, for any player to join. To enhance participation, a special event organised for the entire XTribe platform took place in June 2012, in Rome, Italy. Moreover, we have used AMT to obtain additional data. While other users played the game without any reward, AMT users have been remunerated with an amount of money related to their score. Specifically, the incentives were as follows: players received a base pay of 0.15 US$ irrespective of their final performance provided that they would terminate the game session successfully, and collect the AMT payment code; a bonus of 0.10 US$ was assigned for defeating the Google Directions engine and a further bonus of 0.05 US$ was given for improving the score in phase 2; moreover, a linearly growing bonus ranging from zero to 0.15 US$ was added for scores from 70 up to 150 (for scores over 150, the bonus was 0.15 US$). In total, the reward could range from 0.15 US$ to a maximum of 0.45 US$, which is quite high if compared to AMT standards.

As we will present below, there are some difference between the performance of AMT and unpaid users. AMT users are performing the task as work, hence in a more stressful environment, where completing it faster in general means earning more (although we have given very large bonuses for good scores). Unpaid users performed the task just for pleasure, changing the way they react to information. We could see these two classes of users as two different types of drivers: those pressed by time and eager to get to the destination fast, and those driving without additional stress sources, e.g. for leisure activities. Additionally, the study gives insight in participation patterns for different types of users, which is important from the citizen science perspective.

5.1.3 Money vs Fun

The test cases presented above resulted in 1310 games played (as of June the 26th 2013), out of which 708 by AMT and 602 by unpaid users. In the following we will discuss the results obtained by these two groups.

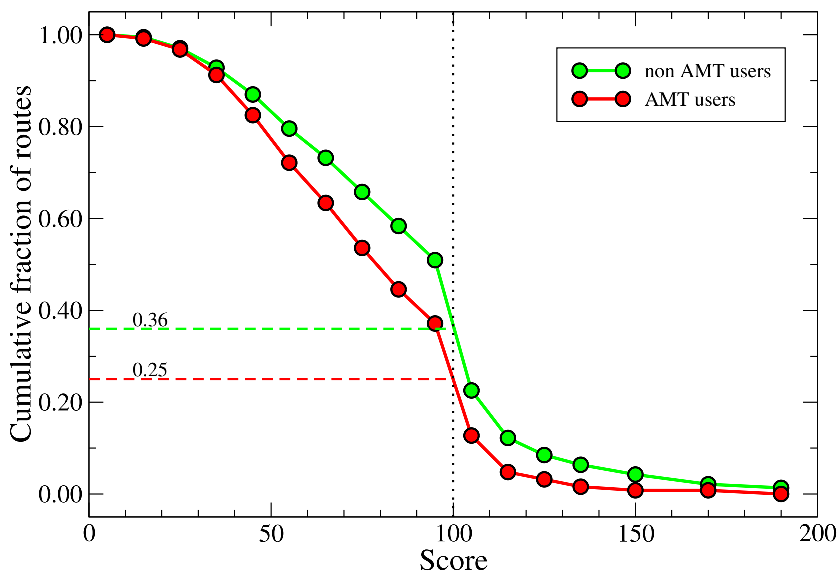


Figure 5.2: Cumulative fraction of routes compared for AMT and non AMT users. Each point corresponds to the fraction of routes that have a score larger than the values given by the horizontal axis. Note how non-AMT users perform better than AMT users since the green curve lies always above the red one. In particular, 25% of AMT users and 36% of non-AMT users scored more than 100 points, which is the score of the Google Directions engine.

First, we are interested in the scores obtained by users, as these show the overall level of performance. Fig. 5.2 shows the distribution of scores (scores from phase 1 and phase 2 considered together) for the two groups of users (AMT and unpaid). Specifically, each point shows the fraction of routes drawn by users (vertical axis), which score more than a given value (horizontal axis). This shows that in general most of the routes drawn by players score under 100, meaning that they are not better than the route indicated by the Google Directions engine. However, there are players who do achieve better performance. Also, AMT users score less than other users, indicating that their routes are less optimal. This might be on one side due to individual characteristics of the players. AMT users are also playing under different conditions, with more pressure for time, which may indicate that drivers trying to get to a destination faster actually end up using a suboptimal route, even when additional information is available [13].

The aforementioned performance may be influenced by several factors, either user related (e.g. familiarity to the city, driving experience, location) but also by the route itself. For instance, optimising a shorter route might be easier, but at the same time finding a route better than Google Directions depends on the number of possible alternatives, and these are fewer for short routes. Hence Fig. 5.3 displays score distributions for the three different route lengths available in the game. Here, the scores for phase 1 and 2 are analysed separately. The difference between AMT and non-AMT users increases with the route size. This is due to the fact that for smaller routes there are limited opportunities for route optimisation, hence making the two groups more similar. There is a small improvement for both groups from the first to the second phase (without and with information), indicating that the presence of information, in general, is beneficial and helps to find a better route. However, in general, the better route is still suboptimal. Improvement is more visible for AMT users, but this might be only due to the fact that it is easier to improve from a lower score. Additionally, the improvement seems to decrease with the length of the route for AMT users, suggesting that drivers under stress are less able to optimise longer routes.

5.1.4 The beginning of an experimental route

We introduced CityRace, a web experiment for route selection in a city. This aims at analysing human behaviour in the presence of traffic data, and study to which extent it affects player decisions, by using a citizen science

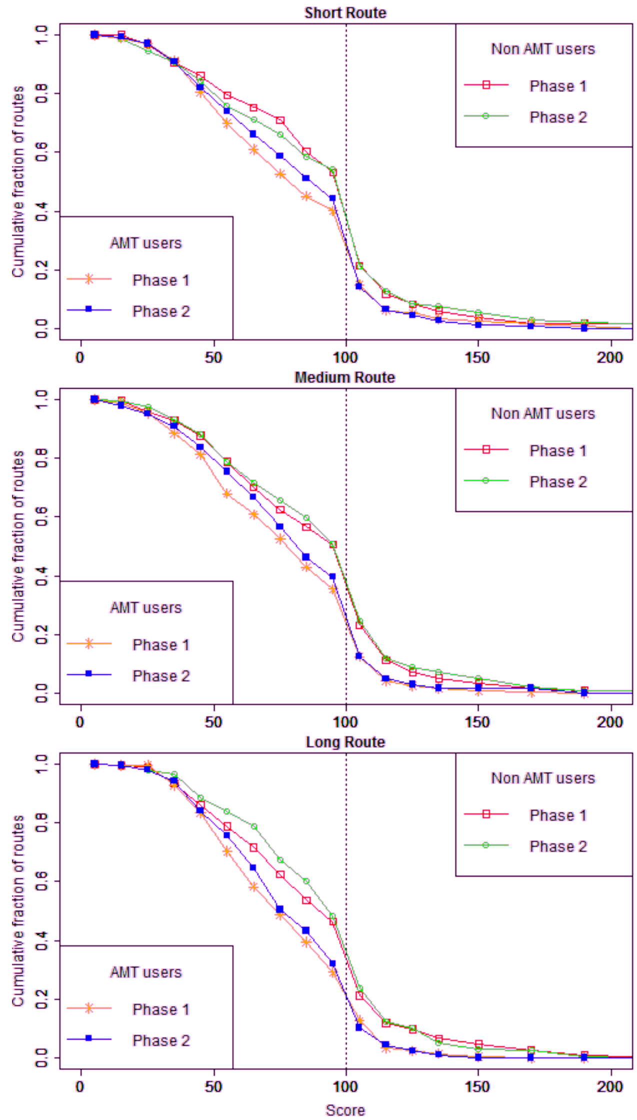


Figure 5.3: Cumulative fraction of routes larger than a given value, compared for AMT and non-AMT users on different route size, with and without information. Again, each point represents the fraction of routes with score larger than the given value on the horizontal axis.

approach, i.e. involving participants through the web. Based on the experimental setting, we have divided the players into two groups, one without time constraints (low stress) and the other with strong time constraints (high pressure) to complete their task, typical behavior of task-paid workers. Results have shown that users under pressure (AMT users) obtain lower scores in general, and seem to use traffic information less than the non-AMT users. Differences between the two sets of users also indicate that paying for citizen science projects might not give the same results as other incentives, especially when the tasks to be accomplished bear a substantial cognitive effort. Paid users will tend to maximize their revenue to the detriment of the experiment quality. We can state that, when a significative cognitive effort is required, investing in the entertaining and fun part may take to a better quality in data gathering than paying directly the volunteers.

We stress that CityRace was intended here as a test case for the XTribe platform to prove its ability to involve players coming from different settings (unpaid and paid) and to test its response to massive user participation. In particular, we are aware that many of the players never lived in the cities proposed in the game so that their strategy was based on the visual map only.

Hence, the CityRace game will remain active online and recruiting activities will continue in order to gather more data and refine this analysis. It will be interesting to see whether users respond better to traffic, which is generally perceived as time consuming, so it has an immediate effect, or to pollution, which may or not affect health in the long run. This, especially considering that highly trafficked and highly polluted areas should in principle correspond. Furthermore, the behaviour in a context with social influence should be analysed. A multiplayer version of the game is already available, where players can be aware or not of the movements of their opponents. This will allow for an analysis of imitation and peer pressure in choosing the best path, as well as for a comparison to the single-player version.

5.2 LaPENSOcosí: a web experiment to unveil political opinion structure

The act of voting is the very core of the representative democracy but, despite its importance, it is essentially an act of extreme synthesis. Almost all the occidental electoral systems allow the expression of a single positive vote per elective assembly. Thus, the voter has to compress all his interpretation of the political landscape in a single preference. In other words, the vote, by its own nature, is a very constraining act and its synthetic outcome hide almost all the information about the complex structure of the opinions about political subjects in the elector mind.

The complex structure of political opinions is a portrait of the political world as it is perceived by the citizen. Such a portrait is important not only because, obviously, it is directly linked to the vote choice, but also by itself. In fact, if we manage to measure and quantify this portrait, it may allow us to study the ability (or inability) to perceive political subjects. We could investigate the limits of this political perception ability, as it has already been done with the other normal sensing ability (sight [71], hearing, etc). This will lead to a deeper understanding of those situations in which our political opinions assignment strategy works worse or fail in describing reality. Furthermore, a snapshot of the opinion structure can also be an interesting inspiration in the understanding of the political opinions dynamics, a crucial subjects in opinion dynamics [42] and not only. In the United States hundreds of millions of dollars are yearly spent for presidential campaigns [37], but the very mechanisms through which public opinion can be succesfully shifted are largely ununderstood. Information from mass media and more focused in time political campaigns strongly influence the debate, and often in an impredictable way [34, 37]. More at the core of the question, people perception of politics and political subjects follows laws that are largely to be uncovered. Due to the practical as well as theoretical interest of understanding political opinions dynamics, available data have been studied seeking for universals [42] and experiments have been proposed aiming at highlighting different mechanisms influencing voters opinions on candidates [126]. However, comprehensive and focused databases of political opinions are lacking.

In order to perform all the described studies, data about political opinions are needed. This data cannot be the electoral results, precisely because, as we said earlier, they hide exactly the information we need. To explore and measure more in-depth citizens' opinions about political entities we have to think about something different. Given the peculiar nature of the data, the most natural way to get it seems to be a dedicated experiment in which measure exactly people opinions about political subjects in the widest sense possible. To do so, the experimental setup have to overcome all the constraints to opinion expression typical of vote but also of traditional surveys. The implementation of an open ambient for opinion expression seems a typical task for web application. Freedom of expression means flexibility of the interface, and a web application can offer wider expression opportunity respect to a traditional paper survey. Beside this, since we needed a 'photography' of the opinion, we had to perform a very fast gathering and, on equal resources, timescale of web experiments have showed to be very short respect to traditional experiments. Thus, our choice for data gathering has been a web experiment implemented on the XTribe platform.

5.2.1 The interface of the web experiment: laPENSOcosi.it

When approaching to the design of a web-experiment, the motivational aspects it is perhaps the most important. The quality and the quantity of results gathered are strongly linked to the users motivation to join and perform the experiment. Thus, we choose to implement our web-application exploiting the hot topic of political elections in Italy in the end of February 2013. The experiment we ran has been dubbed "laPENSOcosí" ("I see it this way") [6] and it introduced a generalized notion of vote, intending to override the limits imposed by its standard definition. As we said, the vote is a synthesis process in which all information about voter portrait of political reality are compressed in a single, discrete and positive preference. In our design, those limits have been broken, enlarging the opportunity of expression to all political subjects (candidates, parties and coalitions). Beside this, each single opinion could range continuously from -1 to 1. This extended and innovative opportunity of expression is the very core of our experiment. Then, the subject can change its opinion about every entity in every moment and without number limits.

The application has been made available through the website `www.lapensocosi.it`, where some general results are presented and people could join the experiment. In the main interaction flow, the user had to express an opinion through the interface reported in Fig. 5.4. In the left part we have the interface before the interaction. There is a political subject to evaluate (in the example reported there is one of the most famous italian politicians: Silvio Berlusconi) with a scale where to put a continuous opinion by a single click. The buttons in the bottom give the opportunity to report the user will to

[6] `http://www.lapensocosi.it`

Figure 5.4: The core of the interaction flow: the evaluation scale. In the left part, there is the interface before the interaction, with the subject and the scale. In the right, the user can confirm or cancel the opinion just inserted.

abstain ("mi astengo" in italian, MA in the following) or that the subject is unknown ("non lo conosco", NLC). In the right part of Fig. 5.4 we report the interface after the opinion expression, i.e. after the click on the scale. In fact, the scale reports the opinion graphically with a pointer and the color, varying between red (for negative opinions) and green (for positive opinions). The bottom buttons now allow the user to confirm or to cancel the opinion. When the user clicks on the confirm button then a new subject is presented, asking again for the user opinion and so on.

In the default mode (named "ModalitaǴioco", Game Mode, GM in the following), subjects were presented randomly from a pool composed by the 160 most famous italian politicians and parties. The main part of our data has been gather in GM but other mode were available. In fact, user were able to evaluate freely whoever they wanted, by searching the name of the political entity (in Search Mode) or by browsing the lists of the candidates (List Mode).

In order to make the experience more social, we implemented some social sharing functionalities, allowing users to publish their opinion on Facebook and Twitter, advertising the experiment. Furthermore, to improve the user experience enjoyability, the request for opinions has been formulated in the shape of a personality test: after each session of evaluation, depending on some general statistics, the user received a 'profile' describing him/her (telling if the user is extremist rather than moderate, or optimist rather than pessimist). Also this content was 'sharable' on Facebook or Twitter: in this way user participating helped us to spread word about our experiment. Beside this, we also advertised the experiment through traditional media (newspaper, TV, radio) and information websites.

5.2.2 A matter of perception

We released our web-experiment on the XTribe portal in the middle of January 2013. At the early days of March 2013 (elections were on 24-25 February 2013) the experiment gathered in GM 81508 opinions expressed by 1727 users, on our pool of 160 political subjects.

One of the main results of the laPENSOcosi.it experiment is the measurement of the opinion distribution itself. The raw distribution, calculated including all data is reported in Fig. 5.5.

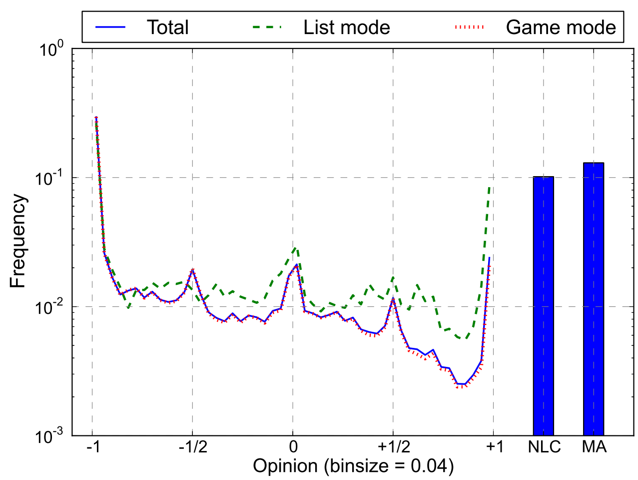


Figure 5.5: Representation of the opinion distribution as an histogram for the opinions with bin size equal to 0.04 in semi-log scale. Data are reported also for all modes. The bar on the right represent the normalized occurrence of 'Non lo conosco' (subject unknown), NLC, and 'Mi astengo' (abstention), MA.

In the right part we can see the ratio of "Non lo conosco" and of "Mi astengo", which is roughly around the 10%. In the left part, distribution for the opinion in general and for both game and list mode. They all show a similar shape, very peaked in the extremes (-1 and $+1$), especially the negative ones. Secondary peaks are also clearly visible in the middle and in ± 0.5.

Tics Effect: how the design may introduce bias

Secondary peaks corresponds to major tics in the scale visible in a previous version of the experiment interface. Those peaks represent a clear bias introduced by major tics, that we removed on the run. We repeated the analysis in order to look at data before and after the removal. Results are reported in Fig. 5.6.

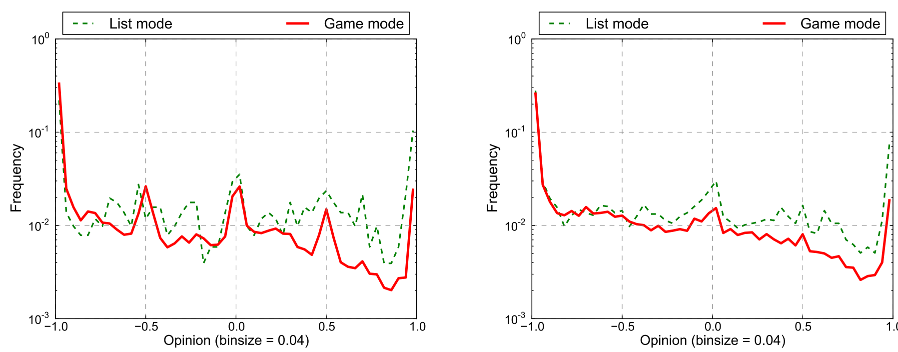


Figure 5.6: Left: the opinion histogram, calculated for opinions expressed on the scale with the major tics. Data are reported also for both modes (40522 opinions from game mode, 510 from list mode). Right: the same opinion histogram, calculated for opinions expressed on the scale without tics. Data are reported also for both modes (39154 opinions from game mode, 2565 from list mode). Again, for both graphs, bin size are equal to 0.04 and the scale is semi-log.

In the left part we can see results obtained analyzing opinions expressed before the major tics removal, while the right part is calculated with evaluation recorded after the tics removal. It is quite clear how, for game mode, the intermediate peaks at -0.5 and at $+0.5$ seem to disappear almost completely while the peak at 0 persists, even if in a more smooth form. In List Mode the situation is more unclear, since there are no clear peaks not even before the tics removal. Anyway, for the greatest part of our data, there is a clear difference that, in next analysis, have to be taken into the account. For the analysis presented in the following, we considered only data gathered with the last version of the interface (39154 opinions, 1038 users).

Starting from the opinions dataset built with the contribution of the users, we can draw a first portrait of the political world perception by measuring the distribution of the opinions, reported in Fig. 5.7. In this graph, opinions are considered without correlation with users or subjects. This portrait tells

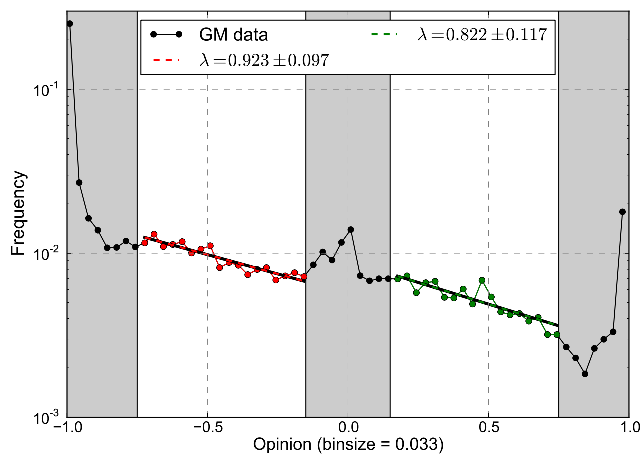


Figure 5.7: The distribution of the opinion calculated with data from GM with the last version of the interface. In green and in red the exponentially-shaped parts with relative fits (exponents are reported in the key).

us about the opinion "usage", i.e. how much a given value in the opinion range has been used, so it gives us many information about the perception of the political world. It is reasonable to assume the distribution reported as the distribution of the opinion of the single average user. In first place, it is quite clear how negative values are used more often than positive ones: the mean of -0.472. And this in unsurprising because it simply corresponds to the general negative perception of politicians in Italy.

The shape of the distribution shows peaks at the extremes (value -1 and $+1$) and in the middle (the zero value), which have been highlighted in gray in Fig. 5.7. Extremes seem to attract a significative part of the opinions: that is a consequence of the 'good/evil' distinction typical of the political context. This happens because each party tends to execrate the adversaries and to exalt itself, and obviously this take to a wider usage of extreme opinions.

As secondary effects, we should also consider the graphic capacity to draw attention of the extremes and the attractive effect of natural numbers. The extremes, in terms of usage, are an area in which the user cannot distinguish anymore between entities. Thus we shall focus, in our analysis, on what happens in the middle. In fact, if we consider the parts between the peaks (red for the negative range, green for the positive one), they seem to follow an exponential law, decreasing when the opinion value is growing. To confirm this, we fitted the distribute in these ranges with a function of the kind:

$$\rho(x) \sim \alpha e^{-\frac{x}{\lambda}}$$

Results are reported in Fig. 5.7. This exponential law, emerging in the most interesting and undiscovered part of the opinion distribution, is a quite unforeseen result. In fact, peaks at the extremes were somehow expected. Even the decreasing of the opinion usage when the value is growing is not so surprise: usually, each person appreciate few political subjects while all the others are typically despised. For this reason a negatively asymmetric distribution seems natural. But the exponential shape of this distribution is telling us something of much more important and detailed about our perception. Before exposing our interpretation of this exponential, let us report some other results, aiming to proof the robustness and the importance of this law.

Optimists and Pessimists

In order to study the stability of the opinion distribution in Fig. 5.7 we calculated the distribution of the opinions for different part of the population. The first distinction we made was between optimists (users with an high average opinion) and pessimists (users with a low average opinion). We divided the population in 4 parts according to users average opinion: the quarter with the highest averages has been considered "very optimist", the quarter with the highest average above the remaining has been considered "optimist", and so on. Then we calculated the opinion distribution for the 4 subpopulation. The result is shown in Fig. 5.8. It appears clearly how the whole structure seem the stay coherent while it seems that the only thing changing is the coefficient exponential envelop. This highlights the importance of the exponent law and seems also to point out that the exponent can be considered a proxy to measure of the optimism.

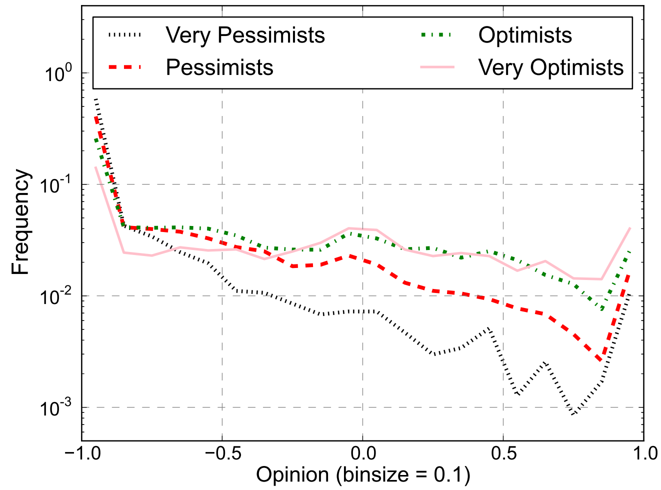


Figure 5.8: The opinions distributions for the four subset of the population, created dividing users according to their average opinion.

Political points of view

We tested also the distribution stability against difference in the political point of view of the considered population. At the 2013 political elections, the Italian political forces scheme could be divided in four main political faction: the center-right coalition (will be indicated as CDX); the center coalition (CC); the center-left coalition (CSX); the Movimento a 5 Stelle party (M5S). Since in this case there is no clear distinction we managed to define, for each user, an aggregated opinion (agg. op. in the following) about each of these poles. This opinions have been used to weight all other opinion while recalculating the opinion distribution. In other words, the get the opinion distribution of the CDX sympathizers we used the agg. op. about CDX of each user to weight his/her opinions, so that, for example, if a user had a CDX agg. op. $= +1$ then his/her opinions were considered normally while if CDX agg. op. $= -1$ then his/her opinions were not considered at all. The result of this calculus, i.e. the four poles weighted distributions are reported in Fig. 5.9. It is clearly visible how all the distributions are fairly compatible. This graph is a clear evidence that opinions distribution we measured and showed in Fig. 5.7 is quite stable and meaningful. Again, in Fig. 5.9 we can find the exponential shape (with some noise due to the reduction of the

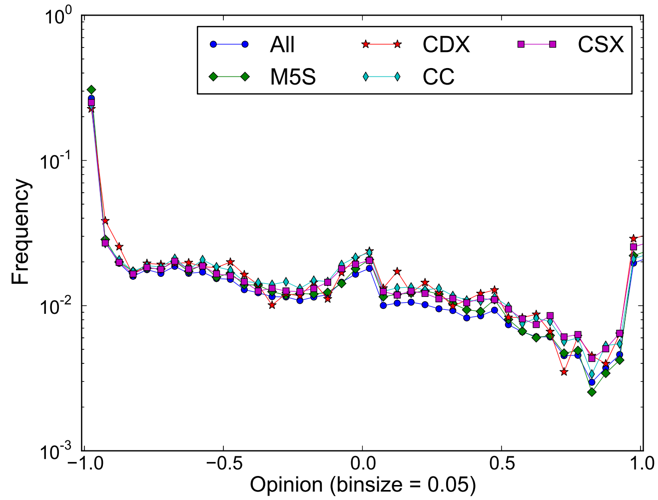


Figure 5.9: The opinions distributions weighted with the agg. op. about the four fronts.

statistics, consequence of the weighting procedure). Furthermore, its also telling that the opinions structure between different populations tends to be similar even if the political point of view is different (or even opposite). This seems to point out a universality for this kind of distribution that, at this point, seems licit to be considered as an intrinsic feature of our political subject perception ability.

5.2.3 The political opinion space and the rank space

Now we will provide the interpretation of the exponential law observed in Fig. 5.7,5.8 and 5.9. In order to understand the implication of the exponentials, let us try to reformulate our problem. To simplify, we will consider a bounded space, an opinion space, in which usage distribution follows just an exponential law. In the opinion space, each user can place political subjects according to his *perception*. Our experiment asks for opinions between -1 and $+1$ but we can also translate these opinions in some sort of *distances perceived* by the user. In this framework, entities near $+1$ can be considered as perceived near the user while those entities near -1 can be considered far from the user, according to his perception. We can thus say that users assign to each entity a coordinate in this *perceptual space*, that we will call $x_p(i)$, for entity i. Without loss of generality, we can consider 0 and +1 as the bounds of our perceptual space (0 corresponds to opinion +1 and 1 corresponds to opinion −1). The exponential law we measured and represented in Fig. 5.7 represents, in this framework, the density of the political subject in the perceptual space. We can then express the*perceived density* as an exponential with the proper normalization in the following way:

$$\rho_p(x_p) = \frac{e^{\frac{x_p}{\lambda}}}{e^{\frac{1}{\lambda}} - 1} \tag{5.1}$$

where λ is a characteristic parameter of the user (or of the community of users considered). As we saw in Fig. 5.8, this parameter is probably linked to the "optimism". What we described so far is a matter of perception. The situation is similar to what happens with our "normal" senses (sight, hearing, touch, etc). For the five senses, we can measure their efficiency in mapping reality by knowing how a certain stimulus (or a distribution of stimuli) is translated in our perception (or in a distribution of perceptions) [30, 134].

In the present case, we measured the perceived distribution, but what about the stimuli distribution? We measured the perceived density distribution of a set of political entities, but how they were distributed originally? And in what kind of space?

It is not just a matter of stimuli distribution. If the sensing process we are considering is a sort of mapping, we first have to consider exactly what we are mapping in the opinion space. From what kind of space the mapping starts? The answer to this question can easily degenerate to metaphysics but instead of speculating about an hyperuranion of politics we will try to stay inside the theoretical framework we defined. So far we considered the evaluation process as a sort of positioning of the entity in the opinion space. Positioning may also be considered a second order process respect to a ranking process. It sounds reasonable to assume that each user has his/her own personal rank of the political subjects. In this rank, we can consider the first as the nearest to the user, and the last as the farthest. While telling his opinion, the user is basically translating a rank into a position. Thus, we can define a *rank space* in which entity are placed, sorted by their rank. By definition of rank, entities are simply distributed uniformly, in a discrete sense (one in each rank position). Even if the rank space discrete by definition, to analyze the mapping, we will consider it as a continuous space, in which $x_r(i)$ will refer to the rank position of the entity i. The bound of the rank will be 1 and the number of entities N_e. As we said, we consider the density as a continuous uniform density distribution defined as follows:

$$\rho_r(x_r) = \frac{1}{N_e} \tag{5.2}$$

5.2.4 Mapping and resolving power

Now that we have a definition for both the perceived and the, let's say, original distribution we can find the transformation by imposing that:

$$\rho_r(x_r)dx_r = \rho_p(x_p)dx_p \tag{5.3}$$

This condition is equivalent to require that the transformation from x_r to x_p is essentially a bijection, i.e. to each point x_r there is one and only one point x_p corresponding. This assumption is quite reasonable for a perceptive mapping. In Eq. 5.3, by substituting Eq. 5.1 and Eq. 5.2, we get:

$$\begin{aligned} \frac{1}{N_e}dx_r &= \frac{e^{\frac{x_p}{\lambda}}}{e^{\frac{1}{\lambda}}-1}dx_p \\ \frac{dx_r}{dx_p} &= N_e\frac{e^{\frac{x_p}{\lambda}}}{e^{\frac{1}{\lambda}}-1} \end{aligned} \tag{5.4}$$

and by integrating we obtain:

$$x_r = \frac{\lambda N_e}{e^{\frac{1}{\lambda}}-1}e^{\frac{x_p}{\lambda}} \tag{5.5}$$

$$x_p = \lambda \log \frac{e^{\frac{1}{\lambda}}-1}{\lambda N_e}x_r \propto \log x_r \tag{5.6}$$

Eq. 5.6 tells us how the rank space is transformed in the opinion space: according to a logarithmic law. This law appears similar to others perception mapping law, such that observed for spatial disposition for natural numbers [52, 101]. To investigate the efficiency of the mapping, let us consider Eq. 5.4. If we substitute Eq. 5.6 we get:

$$dx_p = \frac{\lambda}{x_r}dx_r \tag{5.7}$$

which describes how the perception becomes less efficient when x_r is growing. In other words, differences are perceived smaller and smaller as we go far from the user. Hence is more difficult to distinguish among different but far (from the user) entities. Eq. 5.7 and Eq. 5.6 strongly recall a well-known law linking stimuli and perception: the Weber-Fechner law [134]. This law has been observed in several situation involving physical perceptions, e.g. for light brightness [71] or for noise intensity, but also in other context, e.g. for numerical cognition [52, 101] and recently in subjective perception on the quality of a service [54, 112]. It is at the basis of the definition of a decibel, i.e., of a logarithmic scale for our perception. The presence of a Weber-Fechner-like law in political perception is, as far as we know, a completely novel aspect. The emergence of this sort of “Political Myopia” can have a profound impact on the way models of opinion dynamics are constructed. To complete, let us calculate the resolving power. By considering dx_r as a fixed parameter, we can also find an expression for the opinion space, defining the resolving power needed to notice a difference dx_r as a function of x_r.

$$RP(x_r|dx_r;\lambda;N_e) = \frac{1}{dx_r} x_r \log \frac{e^{\frac{1}{\lambda}} - 1}{\lambda N_e} x_r \tag{5.8}$$

5.2.5 Outcomes of the laPENSOcosí experience

In this section we introduced a web experiment implemented on the XTribe platform where we explicitly investigate users opinions on political entities (parties, coalitions, individual candidates) of the Italian political scene before the political election of February 2013. As a main result, we showed, for the first time to our knowledge, that political perception follows a Weber-Fechner-like law. More in particular, when ranking political entities according to the user expressed preference, the perceived distance of the user from an entities scales as the logarithm of this rank. On the other hand, the ability of distinguish between entities decrease with rank. Our analysis points out where and how out political perception fails in describing reality: we may speculate that, since we cannot distinguish very well in the region of despised entities, we know the less who we like the less. So, perhaps, our dislike may be driven by incomplete or wrong information.

Beside its intrinsic value, these results give us valuable insights about how to design or improve political opinion dynamics models. An agent based modeling, for example, seems a very natural prosecution of this line of research, together with the improvement and the generalization of the web-experiments. In fact, so far our consideration are legitimate only for the Italian political system. The data gathering has to be repeated in different times in the same country and in different country, in order to confirm our results and generalize our conclusion.

5.3 AirProbe: monitoring a locally informed community opinion shift

In this final section we will present the AirProbe web-game experience, part of the main case study concluding the EveryAware european project. The case study involved citizens performing mobile air pollution monitoring using low cost sensing devices, combined with a web-based game to monitor perceived levels of air pollution. Both were embedded in an international competition to motivate participants. Here we will focus on the web-game implementation and result, we will leave out from this discussion the part of design, construction and calibration of the sensing devices and how they have been used to measure pollution levels across the town. The main idea was to use the web-game to observe the shift in the opinion about air pollution caused by direct measurements performed with mobile air monitoring devices.

5.3.1 The air we breathe

Air pollution has an important effect on our health, with an increasing number of studies showing higher risk of respiratory and cardiovascular diseases for people exposed to higher pollution levels [85]. In this context, keeping air pollution at bay has been a major priority for policy makers in the past decades. A lot of efforts have been put into monitoring and controlling air pollution. Large scale monitoring networks routinely monitor target pollutants. They allow for temporal trends in air pollution to be tracked. Significant efforts have also been made to make information accessible to the wider public. However, several papers indicate that official monitoring networks do not have sufficient spatial coverage to provide detailed information on personal exposure of people, as for some pollutants, this may vary substantially among micro-environments [53, 80], i.e. in urban, traffic-prone areas spatial variability is very high [109, 110, 118]. Several pollution sources have been addressed with success. However, persistent problems remain in urban areas, where traffic and domestic heating are important sources [58]. Next to the technical solutions (e.g. electrical mobility), people's personal perceptions, behaviour and choices play a major role in addressing these issues and facilitating change in a bottom-up manner.

Participatory sensing, involving citizens in environmental monitoring, can have multiple potential benefits. Firstly, it can increase coverage of monitored areas, both in time and space, due to the ability to distribute the monitoring activities to multiple individuals [70]. Secondly, the act of monitoring pollution by citizens could facilitate learning and increase their awareness of environmental issues [27]. A recent report on environmental citizen science concludes that few studies on public participation in science and environmental education have rigorously assessed changes in attitudes towards science and the environment, and environmental behaviors, and there appear to be relatively few examples of participatory citizen science having a tangible impact on decision making, although the potential is often noted [117]. An important issue concerns the technological versus social aspect of such projects. Most of the existing projects concentrate mainly on the technological side of participatory air quality sensing, i.e. how to build the sensing devices and map pollution. However, participant engagement, participatory patterns, learning and awareness are equally important aspects. By collecting subjective data as well, monitoring campaigns can enable not only air quality data collection, but also analysis of volunteer behaviour, strategies and a possible increase in awareness.

5.3.2 The AirProbe International Challenge

To address these issues, in the context of the EveryAware project, an international case study was organized: the AirProbe International Challenge (APIC) [22]. This was conducted in a period of 6 weeks at the end of 2013 simultaneously in four cities: Antwerp (Belgium), Kassel (Germany), London (UK) and Turin (Italy). In this test case a web-based game, air quality sensing devices and a competition-based incentive scheme were combined to collect both objective air quality data and data on perceived air quality, to analyse participation patterns and (changes in) perception and behaviour of the participants. The test case was organised as a competition between the cities, to enhance participation.

During this test case, volunteer participants were asked to get involved in two activity types. The first one consisted in using a sensing device (Sensor Box), to measure air pollution (black carbon (BC) concentrations) in their daily life, generating what we call *objective* data. The second activity was playing a web game (AirProbe), where volunteers were asked to estimate the pollution level in their cities, by placing flags (so called *AirPins*) on a map and tagging them with estimated black carbon (BC) concentrations on a scale from 0 to 10 $\mu g/m^3$, resulting in *subjective* data on air pollution (perception). Volunteers involved in the measuring activities were encouraged to play the game and bring other players as well (create a team).

The two data types allow for an analysis of user behavior and perception throughout the challenge. To enable this, the test case was composed of three phases. In phase 1, only the online game was available, so we could obtain an initial map of the perceived air pollution. In phase 2 the measurements started in a predefined area in each of the cities (corresponding also to the web game area), with the web game running in parallel. Phase 3 introduced a change in the game, so that players could purchase information about the real pollution in their cities in the form of sensor-box measurements averaged over small areas (so called *AirSquares*). At the same time, measurements were continued, this time without a restriction of the area to be mapped. Incentives in the form of small prizes were given at the end of each phase to best teams/players.

The data collected during the test case is used here to analyse participation patterns, in terms of activity and coverage, and any changes in perception. Our results indicate that better coverage is obtained when volunteers are assigned a specific mapping area, compared to when they are asked to select the time and location of their measurements. Additionally, when allowed to measure freely, they seem to be attracted to places with higher pollution levels. Furthermore, while at the beginning of the challenge the general perception was that pollution was higher than in reality, perceptions changed in time indicating increased knowledge of real pollution levels. The amount of data collected in the test case, together with the first insights we obtained from these, suggest that bottom-up participatory sensing approaches are effective in attracting participants with high levels of activity and also in enhancing citizen awareness of real pollution levels. The test case presented here has successfully involved 39 teams of volunteers in 4 european locations, gathering 6,615,409 valid geolocalised data points during the challenge (the measuring device collects one data point per second). Information on perception of pollution has been extracted from the online game. After the filtering procedure, the platform had 288 users in total, over six weeks. Their activity resulted in 70,758 AirPins at the end of the test case, that we will use to assess perceived pollution levels. As we said, the discussion of the device part is beyond the scope of this thesis, we will focus in the following on the web experiment.

5.3.3 The AirProbe Web-Game

In order to gather subjective opinions about the air pollution in the four cities we decided to follow the *game with a purpose* approach and accomplish the task using a web-game. We started designing the game taking inspiration from the peculiar kind of data we wanted. Our specific aim was not only to get a map of perceived air pollution but also to study how the perception is affected by objective data. Also, we want an integrated way to show these objective data. Specifically, we needed to monitor volunteer opinion before, during and after exposition to objective pollution data, obtained by the sensing device. This meant keeping the players engaged in the game for the longest time possible, in order to monitor the opinion shift of each player. Beside this, opinions about air pollution had to be geo-localised so the game had to take place on the maps of the four cities. In particular, for each city we defined a mapping area of approximately 3 km^2. The mapping areas are represented in Fig. 5.10.

Considering all this, the most suitable type of game appeared to be a management simulation, like the famous FarmVille or Harvest Moon. In this type of game the user has the task of managing a given territory. By improving their management performances, the users increases their income in the game. Thus they may access a wider set of features, for example they can expand their territory or buy more objects, trying to get a further improvement. The periodic rhythm of this vicious, or virtuous circle, is marked (in FarmVille-like games) by the ripening time of the income: in order to generate a revenue, each action is required at a given time, spanning from few seconds to several hours. This mechanism is an incentive to return to the game, in order to gather the results of ones effort.

The AirProbe web game is a simplified map management game. We reported in Fig. 5.11 the interface of the game. Players are called to fulfil their role of Air Guardians by annotating the map with AirPins: geo-localised flags tagged with an estimated or perceived pollution level (Black Carbon concentration in $\mu g/m^3$, on a scale from 0 to 10). The game area of each city

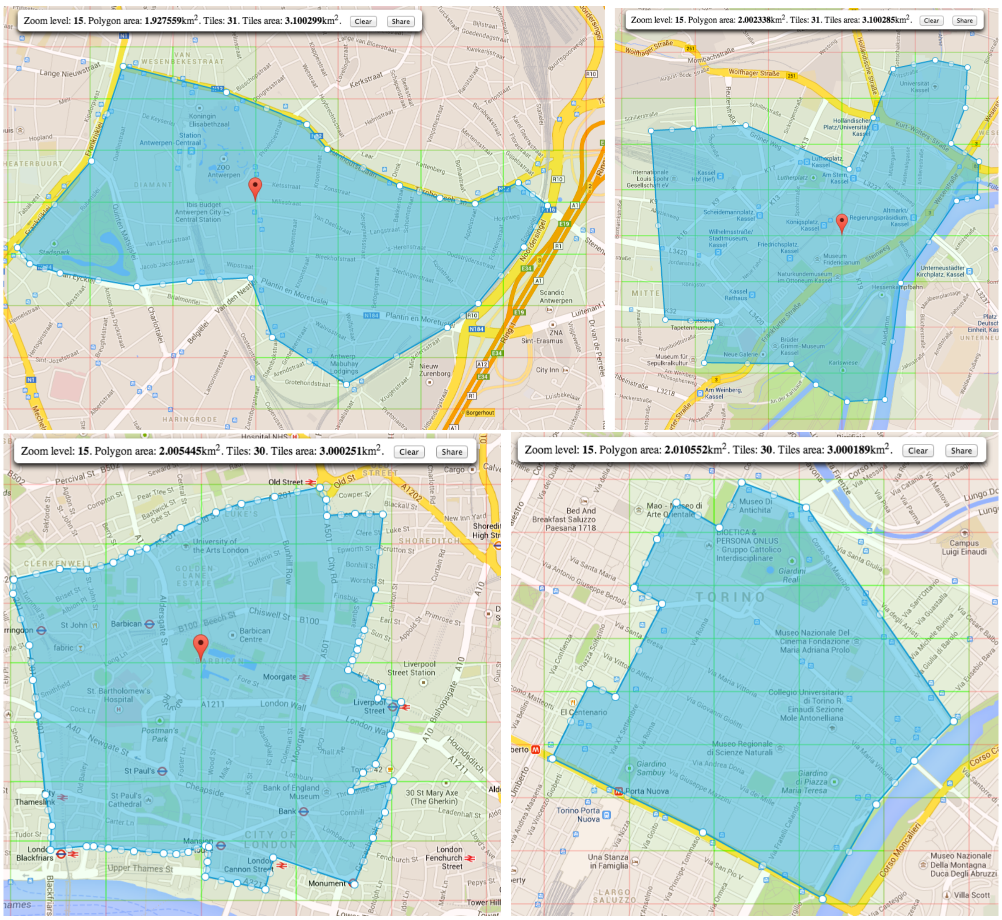


Figure 5.10: In green the game areas and in blue the measurements areas for the four cities. The grid represents the tiles division for the web game. From the top left to the bottom right: Antwerp, Kassel, London and Turin.

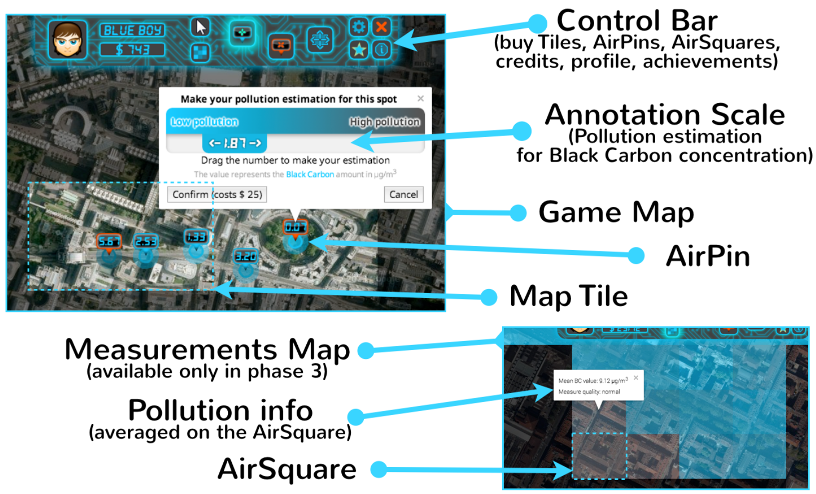


Figure 5.11: Screenshots of the game interface, with indication of the main entities and tools.

is divided into tiles as indicated in Fig. 5.10.

At the beginning of the game, users are asked to create a profile (by choosing an avatar and a name) and to choose a city and a team. Teams were linked to Air Ambassadors, and were an important part of the competition. Then the volunteer starts from a given Tile of the map of the chosen city. The user can interact by placing (or editing or removing) AirPins or by expanding their territory by buying more Tiles. Each day the AirPins placed generate a revenue based on the precision of the annotation. In order to collect the revenue generated every day by each AirPin, the user has to access the game daily, otherwise the revenue will be wasted. The revenue collected will be added to the user balance, and can be used to buy more AirPins and more Tiles. In order to improve motivation and fidelity, there are bonus for day-in-a-row accesses and a large set of achievements. These achievements consist of prizes at given milestones in the game story: a certain number of AirPins or Tiles, precision in the annotation, and so on.

In phase 3 of the case study we made available information about objective measurements gathered with the sensor box during phase 2. We avoided to give punctual information about measurements, otherwise it was likely that users would simply copy the values. So we decided to release aggregated information by introducing a new map partition named AirSquares. Each Tile contains 12 AirSquares, that can be purchased just like AirPins or Tiles. Once users bought an AirSquare, they can see the average pollution value in that area.

5.3.4 A change of perception

Let us now look at the perceived levels of pollution in the mapping area, the AirPin values. In particular, to inspect awareness improvement and the learning process, we are interested in the relation between these annotations and the 'true' pollution values available in the web-game during phase 3 in the form of AirSquares. Thus we define the APD (AirPin difference) as the difference between the AirPin value (perception of the volunteer) and the relative AirSquare value (real pollution level). In other words, the APD is the amount of 'error' in the annotation intended as distance from the real measurement. Figure 5.12 shows several distributions of the APD. In the left part we have APD distributions in each phase for Turin, Kassel and London. Antwerp did not reach the critical mass of data required for this analysis (the amount of web game volunteers was very restricted).

In phase 1, when no volunteer had been exposed to real measurements, we observe three different opinion structures in the three cities, representing the initial perception of volunteers. A systematic overestimation of pollution is present, with peaks around $\sim 4\mu g/m^3$. This is likely to be caused by a scale misunderstanding: players, which were not accustomed to the BC concentration scale, ignored almost completely which values were to be considered reasonable an thus used the middle of the scale (i.e. $5\mu g/m^3$) as 'normal' value. This results in the overestimation observed since the average measured BC concentration lies between 1 and 2 $\mu g/m^3$.

In phase 2 things began to change. Some volunteers (so called *Air Ambassadors*) were given the sensor boxes to start performing measurements. The web game players consisted of these volunteers plus a set of other players recruited by them (so called *Air Guardians*). No data, except for the direct feedback from the boxes (directly accessible only for Air Ambassadors), was shown to the volunteers. Even so, a change is visible in the distribution of APD reported in the left part of Figure 5.12. Volunteers saw that in general BC concentrations are lower than what they believed, and responded by changing the values of the AirPins or taking the information into account when placing new ones. Since the change is quite significant, we also believe that those volunteers equipped with sensor boxes spread the information about what they were measuring, so that all players changed their perception. This decrease in the pollution levels reported in the subjective data of phase two is a first strong indication of learning during this phase. The right side of Figure 5.12 shows APD distributions separately for AirAmbassadors (performing measurements) and AirGuardians (who had no direct exposure to measurements until phase 3). We analyzed just the Turin dataset because in the other cities there was no clear distinction due to Ambassadors sharing their sensor-boxes. The opinion shift in phase 2 is very strong for AirAmbassadors, but some change is also visible for AirGuardians, at least for part of the AirPins. This indicates that there was interaction among players, so that not only volunteers performing measurements, but some of their friends also, changed their perceptions.

Phase 3 brought an important change in the web-game. AirSquares were made available, so players could buy aggregated information (punctual information would have been easily copied) about the pollution detected with sensor-boxes. There is a corresponding radical change in the opinion emerging clearly in the left part of Figure 5.12. In all cities, there is a peak around zero in phase 3 in the APD distribution. This was in some way expected, since we are giving strong hints about pollution levels by means of AirSquares but there is something more happening. In London there is another bigger peak and also in the other cities the distributions show an asymmetry, pointing out that people are not trusting completely the hints because in that case the distribution would have been more similar to delta functions, narrow and symmetric.

In order to describe this phenomenon we defined a stochastic transformation to reproduce the APD distribution for phase 3 starting from the APD distribution of phase 1. This transformation should reproduce the effect of the hints received by our volunteers on the initial distribution of their errors. Based on the empiric observation, the transformation takes into account two main effects: the possibility of complete trust in the hint, so that the opinion is reset near the hint, and the possibility of incomplete trust, so that the opinion is just shifted closer to the hint. The mathematical definition can be found in the following.

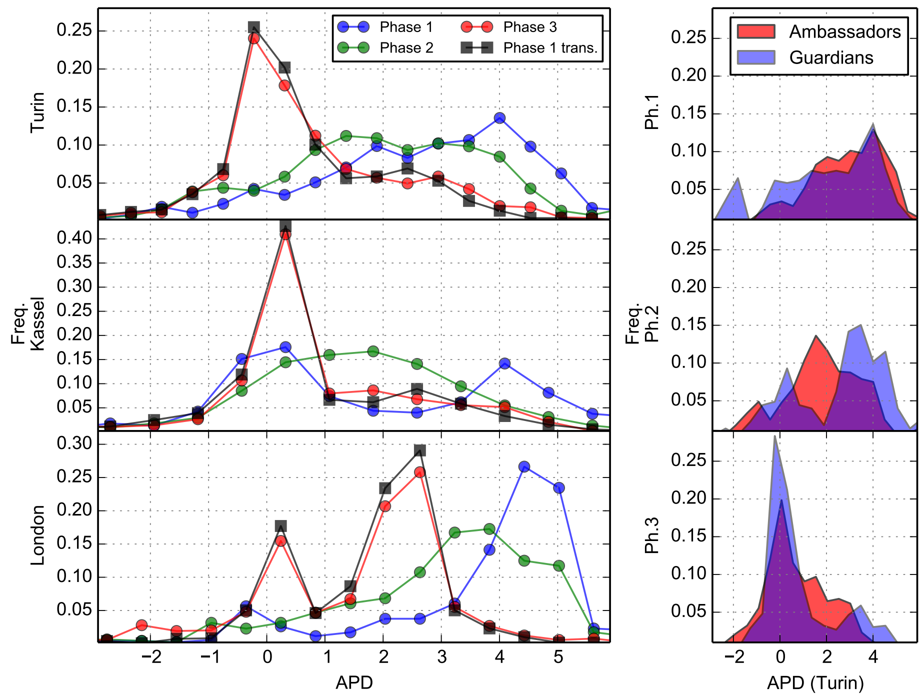


Figure 5.12: **Web-game data: APD distributions.** APD is the deviation between the air quality level annotated (the AirPin value) and the aggregated measurements from sensor-boxes (the AirSquare). The left part shows the distributions of the deviations in each phase for Turin, Kassel and London. An estimation of phase 3 distribution elaborated from phase 1 data with our model is also showed. Right part shows Turin distributions in each phase for AirAmbassadors (volunteers with sensor-box that played the web-game) and AirGuardians (only web-game players).

5.3.5 Opinion evolution model

Let us now look again at the APD histograms aggregated according to each phase in four dedicated log-plot reported in Fig. 5.13.

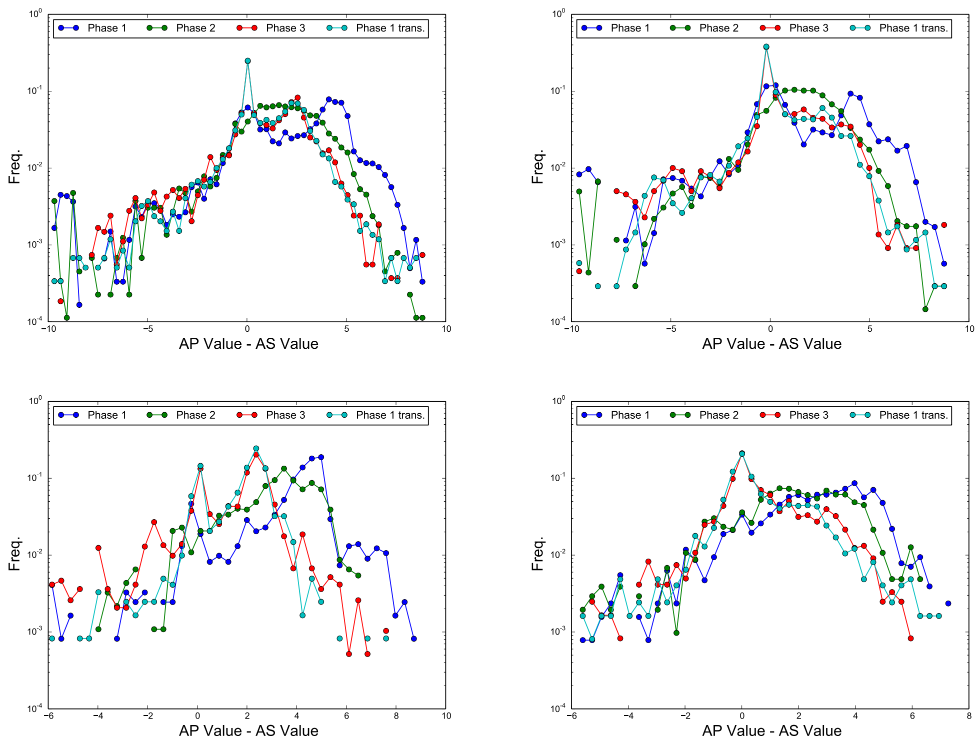


Figure 5.13: Clockwise, from the top left: the APD histogram for the overall, for Kassel, for Turin and for London in each phase of the challenge and with an estimation of phase 3 data obtained from phase 1 data through the transformation defined in Eq. 5.10.

If we look at phase 3 histograms two main features attract our attention: a narrow peak in 0 and a deeply asymmetric structure. The first feature was somehow expected since players are trusting the AS (AirSquare) values shown in the AS, and they are annotating accordingly. Fortunately, the peak at zero is not delta like, that would be expected if users were copying the AS value. Rather players still have their opinion on the environment and keep it despite the on field measurements. This may happen because they are really trying to follow the basic ideas of the game but also because copying it is not the best strategy, since the AS value is aggregated, i.e. it is the average of all sensor-box measures taken in the corresponding AS, while the real measurements used for revenue calculation were punctual values which could be substantially different. So the shape of the distribution around zero seems to be caused by users learning the most likely air quality value and trying to estimate fluctuations. But graphs in Fig. 5.13 show something more. There is a clear asymmetry for phase 3 distributions, since the great part of APD values fall in the positive range. This could be a consequence of the fact that AS values were around 3 $\mu g/m^3$ so there was a 30% probability to underestimate that value and 70% to over estimate, but if we look at the phase 1 distributions, this asymmetry effect seems better explained by a sort of memory effect or inertia of players in changing their opinions. This hypothesis seems realistic if we look at the London graph. The main peak around $4\mu g/m^3$ is still present in phase 3, although it is shifted. In order to measure this effects we defined a transformation that takes into the account both features just discussed: the accumulation around 0 and the shift. Let us consider a given set of opinions o_i about a certain number of topics provided by a certain number of subjects. At a given time those subjects are exposed to values h_i, which are perceived as hints of the true values. We are interested in what happens to the difference between opinions and hints before and after the exposition, to understand how this information will affect the opinion structure. To this aim, we define the set of differences d_i between the opinions and the relative hints and analyse the distribution of those difference before and after the exposition. Obviously, the variation of the differences is only due to the variation of the opinions. As we said, we want to reproduce the phenomenon of the accumulation around the hints (i.e., $d_{aft} \sim 0$) and the shift of the general opinion, that we will try to describe as a sort of rescaling (i.e., $d_{aft} \sim d_{bef}/r$ where r will be the rescaling constant). Which of the two phenomena will take place will be decided randomly: with a given probability p_0 the opinion will reset around 0, otherwise, with probability $1 - p_0$, the opinion will just be rescaled. Finally, around this two attractors we add a certain amount of noise. We decided for a Cauchy distribution $C(X)$ centered in 0 in one case and in d_{bef}/r in the other, i.e.

$$C(x; \mu, \gamma) = \frac{1}{\pi\gamma\left(1 + \left(\frac{x-\mu}{\gamma}\right)^2\right)} \tag{5.9}$$

where μ is the average (and the center of this symmetric distribution) and γ represents a scale factor. It is worth to note that the variance of this distribution is not defined, since the second momentum of the distribution does not converge. This choice seems reasonable because tails seem to be power law-like rather then gaussian-like, as the log plots in Fig. 5.13 show. Let us define our transformation and its effect on the difference d_{bef} between the opinion and the hint before the exposure. According to the rules we stated earlier, d_{aft} will be distributed according to this density function: 5.10

$$T(d_{aft}; d_{bef}, p_0, r, \gamma_0, \gamma_r) = \begin{cases} C(d_{aft}; 0, \gamma_0) & \text{with prob. } p_0 \\ C(d_{aft}; d_{bef}/r, \gamma_r) & \text{with prob. } 1 - p_0 \end{cases} \tag{5.10}$$

The transformation we just defined introduces four parameters:

- p_0, which is the probability that the old opinion is reset around $d = 0$; thus, with probability $1 - p_0$, the opinion shows a certain inertia; this resistance to change causes a shift toward the hint instead of a complete reset;
- r, the rescale factor quantifying the shift of resilient opinions;
- γ_0 and γ_r , the γ scale factors for the Cauchy distributions centered respectively in 0 and in d_{bef}/r introduced to add a realistic noise.

We used our data to infer the parameters of our model for Kassel, London, Turin and for the complete set of data. If we apply the transformation to phase 1 data, we get an estimation of phase 3 distances between opinions and hints. Then, to evaluate how good is the estimation, we use a two sample Kolmogorov-Smirnov two sided test. This kind of test gives as result the probability p_{val} that the hypothesis that the two samples are drawn from the same distribution cannot be rejected. Usually, a value below 5% means that the hypothesis has to be rejected otherwise the hypothesis is likely to be true. If the p_{val} is around 10% the two samples come from two distribution which are, in any case, very close. Above the 30% the samples can be considered with a good degree of confidence as coming from the same distribution. We explored the space of parameters with 10% steps and repeating the test 100 times to find the combinations with the highest p_{val} for Kassel, London, Turin and for the overall. These optimal combinations are reported in Table 5.1 with the relative results for the Kolmogorov-Smirnov test.

Table 5.1: Parameters combination with the highest p_{val} resulting from the Kolmogorov-Smirnov test. Parameter space has been explored with 10% steps and each configuration has been tested 100 times. The average p_{val} is reported. Some peaks in the tails for London compromised the test, causing as a result unsatisfying values for the parameters. We reduced the range in the most meaningful area, which is $(-1:4])$. We found the best parameters testing only this area, obtaining a remarkable result ($p_{val} = 27\%$). Then we made again the test reintroducing neglected data, obtaining a $p_{val} = 9\%$ which is still a satisfactory result.

dataset	p_0	r	γ_r	γ_0	$< p_{val} >$
Kassel	0.336	1.62	0.381	0.0138	0.192
London	0.147	1.90	0.100	0.030	0.267 (0.087)
Turin	0.583	1.56	0.304	0.300	0.417
Overall	0.204	1.767	0.28	0.015	0.262

From the table it appears that the reset of the opinion around the hint does not happen so often. In London, for example, it is almost a secondary effect. In the *best* case, Turin, the reset seems to be there slightly more than in half of the cases. We also reported in Fig. 5.13 an estimation of the APDs for phase 3 obtained by applying the transformation 5.10 with the *optimal* parameter combination to the data of phase 1. The similarity between estimation and phase 3 real data is pretty clear.

5.3.6 Our path on the air

The AirProbe web game is a simplified map management game and thanks to its design and to the recruitment activities of the EveryAware consortium partners, a community of hundreds of volunteers players have been kept engaged. Players fulfilled their role of Air Guardians by annotating the map with so called *AirPins*: geo-localised flags tagged with an estimated or perceived pollution level (black carbon concentration in $\mu g/m^3$, on a scale from 0 to 10).

Subjective data we gathered allowed for analysis of perceived pollution levels and learning mechanisms. We observed, by analyzing differences between perceived and measured pollution levels, that users are able to reduce the 'errors' in the annotations, by learning the 'true' values. However some inertia in changing the old opinion structure was also observed, since asymmetric tails and old peak shifts are present. We also looked at differences between AirAmbassadors (volunteers with sensor-boxes that played the web-game) and AirGuardians (only web-game players). We can argue that the personal experience of the Ambassadors produce a smoother transition (which begins in phase 2), while the in game information produce radical changes. But still both approaches shows the inertia we described earlier, even if in different forms.

In general, we can conclude that all our evidence show that involving volunteers in monitoring campaigns can result in large amounts of data collected. These data show that participation can help learning, to create a more accurate perception of air quality. Thanks to our case study, it has been also possible to outline some of the mechanisms behind the resistance of subjective opinions to objective results.

It is very likely that Eq. 5.10 is not the real transformation of the opinion due to the subjects exposure to hints. We made strong assumptions and we reduced our data set to focus on the interesting part. Also, we are analyzing and modeling the phenomenon on a very narrow timescale (weeks) without knowing almost anything about the others (for example, if we considered months the dynamics could be potentially extremely different). Despite this considerations, the results we showed seem to point out with sufficient reliability that the main ingredients are there. The model we referred to helped us to measure how our volunteers were influenced by the hints we gave them. We may now affirm with a certain degree of confidence that even when people do not trust completely the AS values, they still get influenced by them. Another way to see this is that, even if people do not reset their opinions, the space itself in which their opinions are arranged is deformed by the exposure to hints. Obviously these considerations are justified if the subjects consider the source of the hints as *objective*. In other cases, for example, if volunteers are told that opinions come from other volunteers, completely different dynamics are expected to come into play.

The AirProbe web-game is still available on the XTribe platform. In the hot topic of the Smart Cities it can be a strategic experiment and will probably be replicated soon (with some adjustments) in other urban context.

Chapter 6

Conclusion and perspective

In this thesis we accounted for the complex systems investigation methods evolution from data mining to web-experiment and, finally, to an experimental web platform hosting several experiments. We traced a methodological route amongst experimental experiences in the context of complex techno-social systems. Each step of this path entail both scientific and methodological achievements.

Our experience with the HBC dataset has been a typical example of data mining research. It helped us, in first place, to state the legitimacy of the approach. We took those "spilled" data and validated them against homologous data coming from classic experiments, finding that, with a reasonable filtering, data could be used. Then we used them for novel analysis, rebuilding the word association network and analyzing the duality of patterning in language. Through this experience we provided evidence that not only the approach, with the adoption of suitable procedure, can lead to novel findings but also that this methodology opens brand new ways and scales of interaction with the subject of study, with a considerably minor cost in term of both economic and human resources.

Then, in our following experience, we override the passive data mining. We outlined how, in the context of the EveryAware european project, the WideNoise web experiment allowed us to gather otherwise unachievable data about noise pollution and its perception. (By the way, the very existence and successful course of the EveryAware european project, and of a lot of other similar publicly founded projects, is an indirect proof of the strategic value of these researches which is becoming more and more important also in public research fundings programs.) The WideNoise dataset would have been probably impossible to be obtained in other ways: a similar data gathering campaign would require prohibitive efforts to reach the same spatial-temporal scale and the same extensiveness. Thanks to this dataset, we could address the central question of the EveryAware project, i.e. the measurement of the environment awareness, more specifically referring to noise pollution, in this case. We managed to quantify elusive entities awareness and opinions and observe their dynamics in subject exposed to objective measurements. We also studied the involvement, which is a crucial point in this kind of experiment. The whole experience pointed out how the web experiment methodology seems to be, if not the unique, the most suitable and convenient way to address this kind of subject.

The following step was the introduction of the Experimental Tribe platform. Once we have identified the web experiment as a powerful tool of investigation, the natural evolution is the consolidation and the systematization of this approach. The proposed platform aims exactly in that directions. The main basic idea is to host a growing set of instruments useful to implement, in a simplified way, web experiments. The one part of this instruments and services are already offered by the platform. Another part will soon be released by researchers themselves, that will eventually sharing codes of their finished experiments, for others to verify their results or to use that code as a starting point for new experiments. This is made intrinsically easy because the platform suggests some intuitive development choices (e.g., the division of the interface from the game manager) that make sharing and reusing code much more simple. In addition, the platform itself provides code sharing tools. Researchers are also expected to form a community, sharing best experimental practice, beside codes. Players are expected to form a community too, making recruiting easier for all researchers.

The platform has proven its usefulness with several games already implemented by different researchers. The already existing games refer to studies in language and opinion dynamics, where the human component plays a crucial role, and are designed as web based social experiments. In the final part of this thesis we presented some typical experiments already occurred on the platform. We introduced CityRace, a web experiment to measure urban route selection strategies. This aims at analyzing human behavior in the presence of traffic data, and study to which extent it affects player decisions. We observed what happens with different kind of motivation, namely money and entertainment. Up to our observations, when a significative cognitive effort is required, investing in the entertaining and fun part may take to a better quality in data gathering than paying directly the volunteers. We reported also about the "laPENSOcosí" experiment, which introduced a generalized notion of vote, intending to override the limits imposed by its standard definition. Through this experience we explicitly investigated opinions on political entities (parties, coalitions, individual candidates) of the Italian political scene before the political election of February 2013. The results pointed out that apparently political perception follows a Weber-Fechner-like law, meaning also that the people ability of distinguish between political entities decrease with the appreciation rank. We also presented the AirProbe web-experiment, performed as part of the final case study of the EveryAware european project. Implemented as an international contest between four cities (Antwerp, Kassel, London and Turin), the application gathered in a ludic way opinions about the air quality and its evolution when players were exposed to objective measurements. Analyzing this dataset, we managed to study the learning abilities of players but, more importantly, we measured also the resistance to opinion shift and we proposed a simple functional form for the process. In particular, beside the formulas, we were able to state that, when exposed to objective information, people may reset their opinion to sticking to the information received or they can resist the change. Still, even if they resist, their opinion is anyway subject to a slighter shift, meaning that information perceived as objective seems to have non-trivial effects.

All these three experiments are surely non-exhaustive, and more data will be needed to confirm (or falsify) derived findings, or to deepen the research. And since they are web-experiment implemented on the XTribe web platform, and their codes will be soon shared through the platform, reboot them for new data gathering campaign will comport a minimum effort. Also for a matter of reproducibility, systematization and standardization of web-experiments seems a forced step in the scientific and methodological path of the complex systems investigation methods. Thus the potential value of a platform like the one proposed in this thesis is clear. It is also equally clear that to unleash this potential is a very complex and multi-sided task. The platform development requires not only attention to the scientific needs: that is just the top of the pyramid. There are hardware solution to be find and software choices to be done, all in a perspective of a platform which have to endure years and have to be open and accessible for researchers and developer with different cultural and coding background. And this can be really difficult in a continuously evolving technologic background like ours. Beside this, also proper communication strategies have to be designed and applied in order to gather and keep involved the researchers and the players community. Going viral it is not an easy task and the success is never granted, not even with strong investments. And, since we are dealing with citizens data, legal issues have to be faced as well. Privacy laws have to be respected and this deeply influences the structure of the platform, and its further developments design. The complexity of the task may sound overwhelming. But also building the greatest experimental facility in history, CERN Large Hadron Collider, required a lot of interdisciplinary efforts, and a set of very heterogeneous expertise ranging from architecture, engineering and science to legal, fund rising and project management. This titanic and coordinated efforts, crowned with success, sure are an encouraging and inspirational example. Also, an effort in refining our investigation techniques of the complex systems, and in particular of techno-social systems, seems not only important but necessary, in a certain perspective. There still so little we know about dynamics of the involved processes but surely there is a huge interest around them, not only in science. An easy example may be opinion dynamics. Even if its functioning it is still largely not understood how it works, there are several efforts with the aim of influencing it. In dictatorial states attempts for opinion control are common practices. And in almost every country of the world, influencing opinions about commercial products is another manifestation of these influencing attempts. In both cases, influencing strategies are evolving to adapt to the new context, with unpredictable outcome. It is clear that a better understanding of the opinion dynamic will help keep people opinions safe from external malicious influences. In more general terms, to reach a better comprehension of complex systems is a urgent need in our societies, in order to face the several challenges mankind is facing in its actual unprecedented condition. A huge amount of questions lies still unanswered. Finding answers about real world is exactly the historic mission of science. And making experiments is the main, if not the only, way in which science advance.

Bibliography

[1] Broder, M, et al., *Graph structure in the Web*, *Computer Networks*, **33**, 1- 6, 309 - 320, (2000).

[2] Newman, M. E. J., Forrest, S., Balthrop, J.,Email networks and the spread of computer viruses, *Physical Review E* **66** (2002) 035101.

[3] Vazquez, A., *Statistics of citation networks*, 05/2001, arXiv:cond-mat/0105031.

[4] Heaps, H. S., *Information Retrieval: Computational and Theoretical Aspects* (Academic Press, Inc., Orlando, FL, USA, 1978).

[5] D. C. van Leijenhorst and Th. P. van der Weide, *A formal derivation of Heaps' Law* (Information Sciences, 170, p. 263, 2005).

[6] Pastor-Satorras, R. and Vespignani, A., *Evolution and Structure of the Internet: A Statistical Physics Approach* (Cambridge University Press, New York, NY, USA, 2004).

[7] Barrat, A., Barthélemy, M., and Vespignani, A., *Dynamical processes on complex networks* (Cambridge University Press, 2008).

[8] Newman, M. E. J., Assortative mixing in networks, *Physical Review Letters* **89** (2002) 208701.

[9] http://cds.cern.ch/record/1165534/files/cern-brochure-2009-003-eng.pdf.

[10] http://en.wikipedia.org/wiki/arab_spring.

[11] http://en.wikipedia.org/wiki/great_recession.

[12] http://home.web.cern.ch/about/updates/2013/03/new-results-indicate-new-particle-higgs-boson.

[13] http://ntl.bts.gov/lib/jpodocs/repts_te/13953_files/13953.pdf.

[14] http://press.web.cern.ch/press-releases/2009/11/lhc-sets-new-world-record.

[15] http://press.web.cern.ch/press-releases/2011/04/lhc-sets-world-record-beam-intensity.

[16] http://store.steampowered.com/stats/content/.

[17] https://www.facebook.com/notes/facebook-engineering/under-the-hood-hadoop-distributed-filesystem-reliability-with-namenode-and-avata/10150888759153920.

[18] http://www.autobahn.nrw.de/.

[19] http://www.datacenterknowledge.com/archives/2013/01/18/facebook-builds-new-data-centers-for-cold-storage/.

[20] *Bottom-up Environmental Justice in the UK: A fairer, Greener London*, 2007.

[21] G. et al Aad. Observation of a new χ_b state in radiative transitions to $\upsilon(1s)$ and $\upsilon(2s)$ at atlas. *Phys. Rev. Lett.*, 108:152001, Apr 2012.

[22] AirProbe International Challenge. www.everyaware.eu/category/apic/, 2013.

[23] A Alexander. Smartphone usage statistics 2012, 2012.

[24] P Aoki, RJ Honicky, A Mainwaring, C Myers, E Paulos, S Subramanian, and A Woodruff. A vehicle for research: Using street sweepers to explore the landscape of environmental community action. *proc. CHI 2009 ACM Conference on Human Factors in Computing Systems*, pages 375–384, 2009.

[25] PM Aoki, RJ Honicky, A Mainwaring, C Myers, E Paulos, S Subramanian, and A Woodruff. Common sense: Mobile environmental sensing platforms to support community action and citizen science. *Adjunct Proceedings Ubicomp*, pages 59–60, 2008.

[26] S Arnstein. A ladder of citizen participation. *JAIP*, 35(4):216–224, 1969.

[27] UN General Assembly. Rio declaration on environment and development. *Agenda*, 21, 1992.

[28] Wolfgang Babisch, Bernd Beule, Marianne Schust, Norbert Kersten, and Hartmut Ising. Traffic noise and risk of myocardial infarction. *Epidemiology*, 16(1):33–40, 2005.

[29] Philip Ball. Crowd researchers make pilgrimage safer. *news@nature*, jan 2007.

[30] Andrea Baronchelli, Tao Gong, Andrea Puglisi, and Vittorio Loreto. Modeling the emergence of universality in color naming patterns. *Proceedings of the National Academy of Sciences*, 107(6):2403–2407, February 2010.

[31] Martin Becker, Saverio Caminiti, Donato Fiorella, Louise Francis, Pietro Gravino, Mordechai (Muki) Haklay, Andreas Hotho, Vittorio Loreto, Juergen Mueller, Ferdinando Ricchiuti, Vito D. P. Servedio, Alina Sîrbu, and Francesca Tria. Awareness and learning in participatory noise sensing. *PLoS ONE*, 8(12):e81638, 12 2013.

[32] C Beckx, LI Panis, T A, D Janssens, R Torfs, S Broekx, and G Wets. A dynamic activity-based population modelling approach to evaluate exposure to air pollution: Methods and application to a dutch urban area. *Environmental Impact Assessment Review*, 29:179–185, April 2009.

[33] Y. Benkler. Coase's penguin, or, linux and the nature of the firm. *Yale Law Journal*, 112:369–446, 2002.

[34] Bernard Berelson, Paul Lazarsfeld, and William McPhee. *Voting: A study of opinion formation in a presidential campaign*. University of Chicago Press, 1954.

[35] Geoff Brumfiel. High-energy physics: Down the petabyte highway. *Nature*, 469(7330):282–283, jan 2011.

[36] J Brunet, V Parra Garciaa, A Paulya, C Varennea, and B Laurona. An optimised gas sensor microsystem for accurate and real-time measurement of nitrogen dioxide at ppb level. *Sensors and Actuators B: Chemical*, 134:632–639, September 2008.

[37] Mark Buchanan. How strategists design the perfect candidate. *Science*, 306(5697):799–800, 2004.

[38] G. Caldarelli. *Scale-Free Networks*. Oxford University Press, Oxford, UK, 2007.

[39] Maria Cristina Carotta, Michele Benetti, Elena Ferrari, Alessio Giberti, Cesare Malagu, Marco Nagliati, Beatrice Vendemiati, and Giuliano Martinelli. Basic interpretation of thick film gas sensors for atmospheric application. *Sensors and Actuators B: Chemical*, 126(2):672 – 677, 2007.

[40] C Castellano, S Fortunato, and V Loreto. Statistical physics of social dynamics. *Reviews of Modern Physics*, 81(2):591–646, Apr-Jun 2009.

[41] Matthew Chalmers. Large hadron collider: The big reboot. *Nature*, 514(7521):158–160, oct 2014.

[42] Arnab Chatterjee, Marija Mitrovi/'c, and Santo Fortunato. Universality in voting behavior: an empirical analysis. *Scientific Reports*, 3, 2013.

[43] Lydia B. Chilton and et al. Seaweed: a web application for designing economic games. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, HCOMP '09, pages 34–35, New York, NY, USA, 2009. ACM.

[44] K. Church and P. Hanks. Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1):22–29, 1990.

[45] Vittoria Colizza, Alain Barrat, Marc Barthélemy, and Alessandro Vespignani. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proceedings of the National Academy of Sciences of the United States of America*, 103(7):2015–2020, 2006.

[46] European Commission et al. Directive 2002/49/ec of the european parliament and of the council of 25 june 2002 relating to the assessment and management of environmental noise. *Official Journal of the European Communities L189 of*, 18(2002):12–25, 2002.

[47] Seth Cooper and et al. Predicting protein structures with a multiplayer online game. *Nature*, 466(7307):756–760, August 2010.

[48] Riley Crane and Didier Sornette. Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of the National Academy of Sciences*, 105(41):15649–15653, 2008.

[49] Dana Cuff, Mark Hansen, and Jerry Kang. Urban sensing: out of the woods. *Communications of the ACM*, 51(3):24–33, 2008.

[50] S. Darby. Why, what, when, how, where and who? developing uk policy on metering, billing and energy display devices. In *Proceedings of ACEEE Summer Study on Energy Efficiency in Buildings*, 2008.

[51] Jeffrey Dean and Sanjay Ghemawat. Mapreduce: Simplified data processing on large clusters. *Commun. ACM*, 51(1):107–113, January 2008.

[52] Stanislas Dehaene. The neural basis of the weber–fechner law: a logarithmic mental number line. *Trends in cognitive sciences*, 7(4):145–147, 2003.

[53] Evi Dons, Luc Int Panis, Martine Van Poppel, Jan Theunis, and Geert Wets. Personal exposure to Black Carbon in transport microenvironments. *Atmospheric Environment*, 55(0):392–398, August 2012.

[54] Sebastian Egger, Peter Reichl, Tobias Hosfeld, and Raimund Schatz. "time is bandwidth"? narrowing the gap between subjective time perception and quality of experience. In *Communications (ICC), 2012 IEEE International Conference on*, pages 1325–1330. IEEE, 2012.

[55] S. B. Eisenman, E. Miluzzo, N. D. Lane, R. A. Peterson, G-S. Ahn, and A. T. Campbell. The bikenet mobile sensing system for cyclist experience mapping. In *SenSys '07: Proceedings of the 5th international conference on Embedded networked sensor systems*, pages 87–101, New York, NY, USA, 2007. ACM.

[56] I. Elmi, S. Zampolli, E. Cozzani, F. Mancarella, and G.C. Cardinali. Development of ultra-low-power consumption mox sensors with ppb-level voc detection capabilities for emerging applications. *Sensors and Actuators B: Chemical*, 135:342–351, December 2008.

[57] European Environment Agency. NoiseWatch [Mobile application software], 2011.

[58] European Environment Agency. Air quality in Europe - 2013 report. Technical report, 2013.

[59] D Fischer and et al. Planet Hunters: The First Two Planet Candidates Identified by the Public using the Kepler Public Archive Data. *Monthly Notices of the Royal Astronomical Society*, 419:2900–2911, 2012.

[60] A. Flanagin and M. Metzger. The credibility of volunteered geographic information. *GeoJournal*, 72:137–148, 2008.

[61] Lin Fritschi, Lex Brown, Rokho Kim, Dietrich Schwela, and Stelios Kephalopolous. *Burden of disease from environmental noise: Quantification of healthy life years lost in Europe.* http://www. who. int/quantifying_ehimpacts/publications/e94888/en/, 2011.

[62] Kyle Gabler. Kyle gabler's web page. `http://kylegabler.com/`.

[63] B Galantucci, C. Kroos, and T. Rhodes. The effects of rapidity of fading on communication systems. *Interaction Studies*, 11(1):100–111, 2010.

[64] Antonios Garas, David Garcia, Marcin Skowron, and Frank Schweitzer. Emotional persistence in online chatting communities. *Nature Scientific Reports*, 402:1–34, 2012.

[65] M. F. Goodchild. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69:211–221, 2007.

[66] M. F. Goodchild. Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0. *International Journal of Spatial Data Infrastructures Research*, 2:24–32, 2007.

[67] PIETRO GRAVINO, VITO D. P. SERVEDIO, ALAIN BARRAT, and VITTORIO LORETO. COMPLEX STRUCTURES AND SEMANTICS IN FREE WORD ASSOCIATION. *Advs. Complex Syst.*, 15(03n04):1250054, may 2012.

[68] Mary M Haines, Stephen A Stansfeld, RF Soames Job, Birgitta Berglund, and Jenny Head. A follow-up study of effects of chronic aircraft noise exposure on child stress responses and cognition. *International Journal of Epidemiology*, 30(4):839–845, 2001.

[69] M Haklay, L Francis, , and C Whitaker. Noise Mapping Helps Citizens Take Action, 2008.

[70] David Hasenfratz, Olga Saukh, Silvan Sturzenegger, and Lothar Thiele. Participatory air pollution monitoring using smartphones. *Mobile Sensing*, 2012.

[71] Selig Hecht. The visual discrimination of intensity and the weber-fechner law. *The Journal of General Physiology*, 7(2):235–267, 1924.

[72] M. Hilbert and P. Lopez. The world's technological capacity to store, communicate, and compute information. *Science*, 332(6025):60–65, feb 2011.

[73] Charles F. Hockett. The origin of speech. *Scientific American*, 203:88–96, 1960.

[74] R Honicky, EA Brewer, E Paulos, and R White. N-smarts: networked suite of mobile atmospheric real-time sensors. In *NSDR '08: Proceedings of the second ACM SIGCOMM workshop on Networked systems for developing regions*, pages 25–30, New York, NY, USA, 2008. ACM.

[75] J. Howe. The rise of crowdsourcing. *Wired*, 14(06), 2006.

[76] A. Hudson-Smith, M. Batty, A. Crooks, and R. Milton. Mapping for the masses: Accessing web 2.0 through crowdsourcing. *Social Science Computer Review*, pages 1–15, 2009.

[77] Bret Hull, Vladimir Bychkovsky, Yang Zhang, Kevin Chen, Michel Goraczko, Allen Miu, Eugene Shih, Hari Balakrishnan, and Samuel Madden. Cartel: a distributed mobile sensor computing system. In

SenSys '06: Proceedings of the 4th international conference on Embedded networked sensor systems, pages 125–138, New York, NY, USA, 2006. ACM.

[78] R. Willis J. Collins, G. Thomas and J. Wilsdon. Carrots, sticks and sermons: Influencing public behaviour for environmental goals. Technical report, Demos/ Green Alliance for the UK Department for Environment, Food and Rural Affairs, 2003.

[79] T. Jackson. Motivating sustainable consumption - a review of models of consumer behaviour and behavioural change - a report to the sustainable development research network. Technical report, Policy Studies Institute, 2005.

[80] S Kaur, M J Nieuwenhuijsen, and R N Colvile. Fine particulate matter and carbon monoxide exposure concentrations in urban street transport microenvironments. *Atmospheric Environment*, 41(23):4781–4810, 2007.

[81] S. Kaur, M.J. Nieuwenhuijsen, and R.N. Colvile. Fine particulate matter and carbon monoxide exposure concentrations in urban street transport microenvironments. *Atmospheric Environment*, 41:4781–4810, July 2007.

[82] F. Khatib and et. al. Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nat Struct Mol Biol*, 18:1175–1177, 2011.

[83] Nihal Kularatna and B. H. Sudantha. An environmental air pollution monitoring system based on the ieee 1451 standard for low cost requirements. *IEEE Sensors journal*, 8(4):415–422, April 2008.

[84] R. Lambiotte and M. Ausloos. Coexistence of opposite opinions in a network with communities. *Journal of Statistical Mechanics*, page P08026, 2007.

[85] Lester B Lave and Eugene P Seskin. *Air pollution and human health*, volume 6. Routledge, 2013.

[86] A. Lawrence. The first cuckoo in winter: Phenology, recording, credibility and meaning in britain. *Global Environmental Change*, 19:173–179, 2009.

[87] David Lazer, Alex Pentland, Lada Adamic, Sinan Aral, Albert-Laszlo Barabasi, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall Van Alstyne. Social science: Computational social science. *Science*, 323(5915):721–723, February 2009.

[88] Jing Li, Yijiang Lu, Qi Ye, Martin Cinke, Jie Han, and M. Meyyappan. Carbon nanotube sensors for gas and organic vapor detection. *Nano Letters*, 3:929–933, 2003.

[89] L Francis M Haklay and C Whitaker. Mapping noise pollution. *GIS Professional*, 2008.

[90] Yajie Ma, Mark Richards, Moustafa Ghanem, Yike Guo, and John Hassard. Air pollution monitoring and mining based on sensor grids in london. *Sensors*, pages 3601–3623, 2008.

[91] N Maisonneuve and M Stevens. Measure and map noise pollution with your mobile phone, 2008.

[92] N Maisonneuve, M Stevens, M Niessen, P Hanappe, and L Steels. Citizen noise pollution monitoring, 2009.

[93] Nicolas Maisonneuve, Matthias Stevens, and Bartek Ochab. Participatory noise pollution monitoring using mobile phones. *Information Polity*, 15(1):51–71, 2010.

[94] Winter Mason and Duncan J. Watts. Financial incentives and the "performance of crowds". KDD-HCOMP '09, Paris, France, June 28 2009.

[95] Alexander Mehler. Large text networks as an object of corpus linguistic studies. In Anke Lüdeling and Merja Kytö, editors, *Corpus Linguistics. An International Handbook of the Science of Language and Society*, pages 328–382. De Gruyter, Berlin/New York, 2008.

[96] G. A. Miller. Wordnet - about us. `http://wordnet.princeton.edu`.

[97] G. A. Miller and C. Fellbaum. *WordNet: An electronic lexical database.* MIT Press, Cambridge, MA, 1998.

[98] R. Milton and A. Steed. Mapping carbon monoxide using gps tracked sensors. *Environmental Monitoring and Assessment*, pages 1–19, 2007.

[99] Kai Nagel and Michael Schreckenberg. A cellular automaton model for freeway traffic. *Journal de Physique I*, 2(12):2221–2229, dec 1992.

[100] Douglas Nelson, Cathy McEvoy, and Thomas Schreiber. The university of south florida free association, rhyme, and word fragment norms. *Behavior Research Methods*, 36:402–407, 2004. 10.3758/BF03195588.

[101] Andreas Nieder and Earl K Miller. Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron*, 37(1):149–157, 2003.

[102] Noise Nuisance Association. Noise Nuisance [Mobile application software], 2012.

[103] Brian A. Nosek, Mahzarin R. Banaji, and Anthony G. Greenwald. E-research: Ethics, security, design, and control in psychological research on the internet. *Journal of Social Issues*, 58:161, 2002.

[104] Oded Nov. What motivates wikipedians? *Commun. ACM*, 50(11):60–64, 2007.

[105] G. Paolacci, J. Chandler, and P. Ipeirotis. Running Experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5(5):411–419, 2010.

[106] Willy Passchier-Vermeer and Wim F Passchier. Noise exposure and public health. *Environmental health perspectives*, 108(Suppl 1):123, 2000.

[107] E Paulos, R Honicky, and E Goodman. Sensing atmosphere workshop - position paper for the sensing on everyday mobile phones in support of participatory research. *ACM SenSys*, 2007.

[108] E Paulos, RJ Honicky, and B Hooker. Citizen science - enabling participatory urbanism. In M. Foth, editor, *Handbook of Research on Urban*

Informatics: The Practice and Promise of the Real-Time City, pages 414–433. IGI Global, 2009.

[109] J Peters, J Theunis, M Van Poppel, and P Berghmans. Monitoring PM10 and ultrafine particles in urban environments using mobile measurements. *Aerosol and Air Quality Research*, 13:509–522, 2013.

[110] J Peters, J Van Den Bossche, M Reggente, M Van Poppel, B De Baets, and J Theunis. Cyclist exposure to UFP and BC on urban routes in Antwerp, Belgium. *Atmospheric Environment*, 92:31–43, 2014.

[111] Economist Technology Quarterly. The digital geographers, 2009.

[112] Peter Reichl, Sebastian Egger, Raimund Schatz, and Alessandro D'Alconzo. The logarithmic nature of qoe and the role of the weber-fechner law in qoe assessment. In *Communications (ICC), 2010 IEEE International Conference on*, pages 1–5. IEEE, 2010.

[113] Stefan Reimann. Challet, d., marsili, m., and zhang, y.: Minority games: Interacting agents in financial markets. *Journal of Economics - Zeitschrift für Nationalökonomie*, 87(2):204–207, feb 2006.

[114] Matthew J. Salganik and Duncan J. Watts. Web-Based Experiments for the Study of Collective Social Dynamics in Cultural Markets. *Topics in Cognitive Science*, 1(3):439–468, 2009.

[115] Yukie Sano, Kenta Yamada, Hayafumi Watanabe, Hideki Takayasu, and Misako Takayasu. Empirical analysis of collective human behavior for extraordinary events in the blogosphere. *Phys. Rev. E*, 87:012805, Jan 2013.

[116] Kevin Schawinski and et al. The Sudden Death of the Nearest Quasar. *The Astrophysical Journal Letters*, 724(1):L30, 2010.

[117] Science Communication Unit - University of the West of England - Bristol. Science for Environment Policy In-Depth report : Environmental Citizen Science. Technical report, 2013.

[118] E Setton, J D Marshall, M Brauer, K R Lundquist, P Hystad, P Keller, and D Cloutier-Fisher. The impact of daily mobility on exposure to traffic-related air pollution and health effect estimates. *Journal of Exposure Science and Environmental Epidemiology*, 21(1):42–48, 2011.

[119] M. Srivastava, M. Hansen, J. Burke, A. Parker, S. Reddy, G. Saurabh, M. Allman, V. Paxson, and D. Estrin. Wireless urban sensing systems - cens technical report number 65. Technical report, Centre for Embedded Networked Sensing Systems, UCLA, 2006.

[120] Australia State Government of Victoria. Theories and models of behaviour change, 2006.

[121] Matthias Stevens. *Community memories for sustainable societies: The case of environmental noise*. PhD thesis, Vrije Universiteit Brussel, 2012.

[122] M. Steyvers and J. B. Tenenbaum. The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, 29:41–78, 2005.

[123] Siddharth Suri and Duncan J. Watts. Cooperation and contagion in web-based, networked public goods experiments. *PLoS ONE*, 6(3):e16836, 2011.

[124] K. Sznajd-Weron and J. Sznajd. Opinion Evolution in Closed Community. *International Journal of Modern Physics C*, 11:1157–1165, 2000.

[125] TAGora. Theoretical tools for modeling and analyzing collaborative social tagging systems. Deliverable D4.1, TAGora project, 2007.

[126] Alexander Todorov, Anesu N Mandisodza, Amir Goren, and Crystal C Hall. Inferences of competence from faces predict election outcomes. *Science*, 308(5728):1623–1626, 2005.

[127] Francesca Tria, Bruno Galantucci, and Vittorio Loreto. Naming a structured world: A cultural route to duality of patterning. *PLoS ONE*, 7(6):e37744, jun 2012.

[128] Wataru Tsujitaa, Akihito Yoshinoa, Hiroshi Ishidab, and Toyosaka Moriizumia. Gas sensor network for air-pollution monitoring. *Sensors and Actuators B: Chemical*, 110(2):304–311, October 2005.

[129] J.P. Viricellea, A. Paulyb, L. Mazetb, J. Brunetb, M. Bouvetc, C. Varenneb, and C. Pijolata. Selectivity improvement of semiconducting gas sensors by selective filter for atmospheric pollutants

detection. *Materials Science and Engineering: C*, 26:186–195, March 2006.

[130] Péter Völgyesi, András Nádas, Xenofon Koutsoukos, and Ákos Lédeczi. Air quality monitoring with sensormap. In *IPSN '08: Proceedings of the 7th international conference on Information processing in sensor networks*, pages 529–530, Washington, DC, USA, 2008. IEEE Computer Society.

[131] Luis von Ahn. Games with a purpose. *Computer*, 39(6):92–94, 2006.

[132] Luis von Ahn and Laura Dabbish. Labeling images with a computer game. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 319–326, New York, NY, USA, 2004. ACM.

[133] Fahui Wang and Yanqing Xu. Estimating o–d travel time matrix by google maps api: implementation, advantages, and implications. *Annals of GIS*, 17(4):199–209, 2011.

[134] E. H. (Ernst Heinrich) Weber, David J Murray, Helen E. (Helen Elizabeth) Ross, 1795-1878. Tactu Weber, E. H. (Ernst Heinrich), 1795-1878. Tastsinn Weber, E. H. (Ernst Heinrich), and Experimental Psychology Society. *E.H. Weber on the tactile senses.* Hove : Erlbaum (UK) Taylor and Francis, 2nd ed edition, 1996. '2nd edition' - t.p. - Originally published: London : Academic for the Experimental Psychology Society, 1978.

[135] J. Gaines Wilson, Simon Kingham, Jamie Pearce, and Andrew P. Sturman. A review of intraurban variations in particulate air pollution: Implications for epidemiological research. *Atmospheric Environment*, 39:6444–6462, November 2005.

[136] World Health Organization, The WHO Centre for Health Development, Kobe, and United Nations Human Settlements Programme. Hidden cities: unmasking and overcoming health inequities in urban settings. 2010.