Online Identities, Profiling and Cyber Bullying

Version 2.0 – 14 March 2015

Author(s)		Maria Christodoulou	COIN
		Onno Hansen	EZZEV Foundation
		Christos Anthis	CrystalClearSoft
		Mattheos Kakaris	CrystalClearSoft
		Nektarios Konstantinou	CrystalClearSoft
Reviewer(s)		Mattheos Kakaris	ChrystalClearSoft

Lifelong Learning Programme

This project has been funded with support from the European Commission.

Revision History

Version	Date	Author	Description	Action	Pages
0.1	13 Feb	MCH	Creation of the document	C	18
0.2	14 Feb	MKA	Insert additional paragraphs	I	19
0.3	15 Feb	OHN	Insert additional paragraphs	I	20
0.4	18 Feb	CAN	Insert additional paragraphs	I	21
0.5	18 Feb	NKO	Insert additional paragraphs	I	22
1.0	19 Feb	MCH	Comments on text	U	20
1.1	25 Feb	MKA	Insert additional paragraphs	I	21
1.2	27 Feb	OHN	Insert additional paragraphs	I	22
1.3	27 Feb	NKO	Finalise document for review	U	22
2.0	14 Mar	MCH	Finalise document	U	22

(*) Action: C = Creation, I = Insert, U = Update, R = Replace, D = Delete

Executive Summary

In this analysis we present the goals and delineation of our research, based on a theoretical framework, in which we've combined (1) core concepts of Goffman's theory of expressing and creating identities, and (2) Clarke's distinction between a projected (presented) persona and an imposed persona in online contexts. We use this theoretical framework to shed light on the relevance and potentially harmful consequences of profiling by companies for children's identity expressions in online contexts.

TABLE OF CONTENTS

LIST OF FIGURES

1. Introduction

Using the internet, whether to look up information on Google, to connect with friends via social network sites, to buy products in online stores, or to share videos or other materials in web 2.0 environments, has become second nature to most of us living in 'digital societies'. This is all the more true for children and adolescents, who have never known a world without internet and all the connectedness it brings. A German court recently ruled that being online is as much of a necessity in daily life as is having a car or a refrigerator.[1] While the truth and the reach of this claim can, of course, be debated, it does reveal that in the few decades of its existence the internet has grown to be a very central part of the infrastructure of our everyday lives.

The internet offers children a wealth of opportunities to share and connect with others, to play, to learn about new topics and discover new knowledge or perspectives, and to construct and express their identities. Its global reach, its networked yet non-bordered nature, and its horizontal organisation all contribute to an open, free zone for playing, experimentation and exploration. At the same time, however, the internet also harbours a number of hazards and dangers, some quite innocent or rare, yet others quite serious and harmful. It is not surprising, therefore, that parents and teachers guard the online activities of their children with great care. In order to ensure that children grow into wise, savvy internet users, parents and teachers alike must help them in learning to understand the opportunities and pitfalls of the internet.

1.1. Profiling internet users

One of the internet 'harms' that has recently been discussed quite regularly in the media is that of *profiling*.

Profiling refers to the **use of "*sophisticated pattern recognition*"** (Hildebrandt, 2006) by **governments and businesses**, which employ this technique to **distil meaningful information from massive amounts of data about individuals or groups of people**, for example for the purpose of **targeted advertising** and personalised services in the case of businesses, or **policing**, crime prevention and detection, combating terrorism and **surveillance** in the case of governments. Profiling revolves around the idea that **large sets of randomly collected data about individuals** and groups of people **can generate interesting, surprising and meaningful correlations** that machines, with their vast powers of calculation can detect, while we as humans cannot.

Businesses can use such correlations to improve their services to customers, or provide better product suggestions and hence increase sales and customer satisfaction levels. Governments can use such correlations to detect undesired behaviours, and criminal or terrorist acts, even as they are in the making.

As the name suggests, profiling may lead to the creation of extensive *profiles*, in which information about individuals or groups of individuals are accumulated, stored, and used for the purposes cited above. Such profiles may build on the *explicit*, *intentional* actions of users on the internet, and, as we will discuss more extensively below, on the *implicit*, or *unintentional* traces they leave behind as they surf the web, for instance by monitoring their behaviour observed via their clickstream (i.e. when they click on links to navigate the web).

[1] See http://www.reuters.com/article/2013/01/24/us-germany-internet-idUSBRE90N15H20130124

The high potential of profiling based on usage data, more technically referred to as '*web usage mining*' (Mobasher et al., 2000), has been well studied for over 15 years. Generally, these data are collected with the best intentions, as observed user behaviour is known to be crucial to optimize website design (for example, link and menu structure) and a key ingredient to realize high quality search engine results.

1.2. The risks of profiling

Widespread profiling practices are however not without risk. Given the complexity of the data itself, combined with the complexity of human behaviour patterns, one of the concerns relating to profiling is the occurrence of 'false positives' (Rubinstein *et al.*, 2008): the software finds correlations in the data that are deemed meaningful, when in fact the correlation is accidental and random. When false positives form the basis of decisions in the real world, this is a worrying phenomenon indeed. The occurrence of false positives may lead to relatively harmless mistakes, for instance to product recommendations that do not meet the users' wishes and desires, but also to very serious ones, for example to false and unwarranted accusations of terrorist acts or other criminal conduct.

Moreover, one of the most serious concerns surrounding profiling is the *opaqueness* that surrounds it. It is often unclear to internet users when, where and for which purposes they are profiled.[2] It is also unclear to users in which cases they are presented with decisions that build on profiling processes, or even *that* this may be the case. As said above, businesses and governments may use profiling techniques to create profiles of consumers or citizens, respectively, which may help them predict preferences, choices, desires and potential future behaviours of their respective target groups. These profiles may thus be used to target individuals – read: both grown-ups and children! – with commercial offers, without these individuals knowing that this is the case, or which profiles or digital traces these recommendations are based on. Especially in relation to children, this is a serious issue. Many online games for children, for example, abound with subtle (or not so subtle) product recommendations made on the basis of children's actions within the game or even outside, for instance when they've linked their profile in the game to their profile on social media platforms such as Facebook. Since such recommendations may be personalised, based on profiling, the seduction to buy the products offered may be much greater for these children. This may draw children, sometimes from very early ages onwards, into commercialised worlds in which the goal is to sell as many products as possible, while the children themselves are oblivious to this fact.

Profiling may also provide a basis to make decisions that have a negative effect upon individuals' real world, for example by limiting their freedom to travel by airplane, or to acquire certain products or services. For example, profiles may be used to exclude 'high risk' individuals from health insurance packages or mortgages – read: individuals who have been profiled as more likely than average to run the risk of getting a life threatening disease. Note that, while the decision to turn down high risk individuals may be made by humans (most likely on the basis of evidence collected and/or analysed using computer support), in some cases the decision may even be left completely to the machine, building on the outcomes of profiling. What's more, the individuals who suffer the consequences of such a decision may never know that the decision was based on profiling, nor which of his/her behaviours formed the basis for the profile to emerge, nor whether or not human beings were involved in making the decision.

[2] As a matter of fact, we suspect that the majority of internet users know, at least to some degree, that they are being profiled when going online; but not in which specific instances, or for what reasons. Furthermore, we hypothesise that internet users have limited knowledge (if any) of how profiling technologies work, or what the effects of the use of such technologies, both for their online and their offline lives, can be. These are topics to be verified in the survey among children/teenagers, which will be presented in D2.1.

Clearly, these issues are equally relevant to children as they are to adults. They, too, may suffer potentially negative consequences from actions conducted online, either intentionally or accidentally/outside their awareness. Think for instance of attracting surveillance by police officers after using certain keywords that 'raised red flags' in a conversation that the youngster him/herself considered entirely innocent on a social network site, or even being denied further access to such a site as a result of profiling practices. And, since the internet 'never forgets', think of the potential consequences, years down the line, of accidental goofs or intentional foolishness by children and teenagers, for example resulting in lesser/different job opportunities or other life choices being barred or altered.

What this discussion reveals is that profiling is opaque because, first, the systems used remain opaque – we mostly do not know in which situations our actions are profiled, and to which extent (we return to this point later on, when diving deeper into the technical details) –, and, second, because the algorithms and rules used in profiling remain invisible (Hildebrandt, 2008). Finally, it is opaque because we do not know in which situations decisions with which we are confronted are the result of online profiling, nor do we know the reasoning behind these decisions (Brown and Korff, 2009; Hildebrandt, 2008). Combined, these factors make it very difficult for individuals to combat profiling practices and the consequences they may have.

Because of its opaqueness, some argue that profiling children raises an important, fundamental and ethical question, viz. whether profiling children is fair, or perhaps should be considered inherently unfair. One could claim that by using ever more subtle and sophisticated profiling and marketing practices, businesses aim at exploiting children's naivety or their incapacity to see through marketing messages when playing in digital playgrounds or socializing on social network sites. This applies not only to younger children, but also to teenagers, when the boundaries between marketing and entertainment blur and behaviour is influenced on more subconscious levels. Perhaps, therefore, it is fundamentally unfair to use profiling techniques for these target groups.

Targeted advertising and personalised services are becoming more and more common in commercial environments on the internet. And with the threat of terrorism and the pressures on many Western governments to combat crime effectively there is a presumed significant increase in government profiling as well, although numbers are hard to come by, since much of the government's profiling practices fall under the umbrella of the various national intelligence agencies, and hence it is hard to know the extent to which government profiling actually takes place (Brown and Korff, 2009).

In this project we will focus on *profiling by commercial parties*. Aside from the fact that it is more difficult to research government profiling, due to the secrecy of its practices as described above, the main reason for this choice is that children are much more likely to encounter profiling by commercial (rather than government) parties in their online activities – ranging from in-app product recommendations to targeted advertisements based on their web surfing behaviours and to subtle ways of evoking materialistic or other commercially attractive desires. It is important that children learn to recognise the targeted messages they encounter online, or, the ways in which businesses attempt to inconspicuously prompt needs and desires in children to maximise their profits, all of which may be part of profiling practices. They should learn to consider strategies to avoid or reduce profiling (and thereby reduce the risks discussed above), if they so desire. Specifically, in this project we want raise children's awareness of profiling *in relation to their own identities*.

Before we turn to a more in-depth discussion of how profiling and identity are related, we will first shed some more light on the current reality of commercial profiling practices in relation to children as these exist on the internet today, and present some facts and figures regarding the extent of this phenomenon.

1.3. Profiling children in practice: Depicting the current reality

Of course, many uses of profiling technologies are perfectly valid, desirable even: serving to enhance online experiences of adults and children alike. The recently completed and EU funded PuppyIR project, for instance, has demonstrated how simple techniques can be used to identify the web pages most suited for young internet users, and, using this information, also select a search engine's queries that are most likely issued by these children and adolescents. Consider for example the ODP, a large scale web taxonomy, where the Kids & Teens category annotates that part of the web that is most suitable for the two age groups. Propagating that age information along the links that connect these pages together (using a variant of Google's pagerank algorithm that has been aptly named 'agerank'), captures a significant large part of the web that is most useful for children and teenagers (Gyllstrom and Moens, 2010). Simple variants of counting how many of the result pages returned upon a query issued to a search engine (or, the percentage of search result pages actually clicked upon) are labelled as suitable for children, the search engine provider may even estimate the user's age (Duarte and Weber, 2011).

Another study demonstrates how the level of school children's capability to search the web effectively (their 'search literacy') can be inferred from monitoring their actual search engine use during class-wide assignments, which could in turn be deployed to direct the teachers' attention to those who needing help the most (Eickhoff et al., 2012). PuppyIR results include a wide variety of demonstrators that show positive applications of such technologies developed, including tools to adapt search engines or blog readers to consider child-suitability, to improve query assistance, to ease access to medical information, and even to mitigate bias in results when children search for contentious topics.

Applying these profiling techniques requires access to the users' online behaviour, which is in principle limited to the entity that controls the (web) server accessed. Thus, the opportunities to use (and misuse) web and search engine usage data (legally) may seem limited to the largest online players; and they have a strong brand reputation to maintain, so one the basis of these facts one could claim that the actual risks of profiling could be considered rather limited.

In practice however, websites may attempt to track users also after they have left their servers, across their own various services, and, by sharing this information across different entities. A well-known example is the wide-spread use of so-called *third party cookies*, a phenomenon that led to a rather constraining law in The Netherlands – a law that may have as a negative side-effect that online services may not always achieve the best possible online experience for their users. Unfortunately, we could not find scientific literature that details to what extent commercial parties have targeted children to build profile information. In the next section, we will however discuss findings from the Wall Street Journal.

Alternatively, profile information may also be collected by crawling social media sites. A recent study by the Polytechnic Institute of New York University demonstrates how easy it is to build up detailed knowledge about minors by analysing their social media usage – despite the fact that they are supposed to be legally protected by laws such as the US Children's Online Privacy Protection Act (COPPA); a law specifically designed to protect the privacy of children. In their technical report (Dey et al., 2012), the researchers describe how they collected (a large proportion of) the Facebook profiles of pupils from three different high schools in New York. They discovered the current high school, graduation year, inferred birth year and a list of school friends of most of the students. All of this information, not usually accessible on the profiles of minors, was collected without the need to establish any friend links with students.

Ironically, the COPPA law indirectly facilitated this 'attack'. In order to bypass restrictions put in place due to the COPPA law, some children will lie about their ages when registering, and as a result their profile information is protected much less than it would be should they give their

true age as a minor. This unfortunately not only increases the exposure for themselves, but also for their non-lying friends.

1.4. Facts and figures about profiling children and teenagers

Perhaps the best online resource to give an actual impression of tracking and profiling targeting children and teenagers has been made available by the Wall Street Journal, through a special section of their *What They Know* blog.[3] For one of their newspaper articles, looking specifically into tracking minors, they analysed 50 of the most-visited U.S. websites for children and teens (as ranked by the comScore Media Metrix report from April 2010). They analysed each of these websites using special-purpose software, and summarized their findings in 'Tracker Scorecards'. Results indicate that the use of tracking and profiling varies a lot among websites; the worst case example they identified uses an incredible number of 248 different ways to track their users, out of which more than one hundred may keep the gathered information indefinitely, and 28 do not provide any means for users to opt-out. Viacom's Nickelodeon TV network accounted for eight of the 50 sites in the survey. On average, the eight installed 81 tracking tools, close to the 82 average for all 50 sites. The journalists conclude from their investigation that children face intensive tracking on the web.

After this discussion on the issues surrounding online profiling practices, especially when these are applied to children, we will now turn to the topic of this project: raising awareness on the relationship between profiling and identity in children.

[3] http://blogs.wsj.com/wtk-kids/ (last accessed on 26-02-2013).

2. Identity

As we've argued in the Introduction children and teenagers may use the internet for a host of different reasons. One of them is to express, and potentially experiment with, their identity (Turkle, 1995, 1996, 2011). Using social network sites such as Facebook, online role playing games such as World of Warcraft, and social media such as Twitter, children and teenagers can connect with others, interact with them, share ideas, images and movie clips, and engage in a variety of versions of 'digital flea picking'[4]. Developing, expressing and experimenting with identities is a central element of growing from childhood into maturity, and therefore it is worthwhile to investigate how the internet affords and inhibits children's abilities to engage in online self-exploration. But before going into this in more detail, we will first present some characteristics of the process of creating, expressing and experiencing identity in our *offline* lives, as put forth by the Canadian 20th century sociologist Erving Goffman (Goffman, 1959; Lemert and Branaman, 1997; Van den Berg, 2008). Goffman's perspective on identity remains very influential to this day, and as we will show below can be applied to online environments to reveal a deeper understanding of identity in online worlds as well.

2.1. Constructing offline identities: 'giving' versus 'giving off'

Erving Goffman's key work on the expression and construction of identities is called *The presentation of self in everyday life* and was published in 1959. As a micro-sociologist Goffman was interested in the small, everyday, concrete situations that make up our daily lives, and especially in the interaction rituals that took place between human beings in these situations (also see Goffman, 1951, 1961, 1963, 1971, 1982). He argued that each and everyone of us conducts a number of different *roles* each day, depending on the social situation that we find ourselves in (Meyrowitz, 1985, 2005; Van den Berg, 2010). We display a professional part of ourselves when at work, a private part when at home, and a 'customer part' when visiting a supermarket. Depending on the location and the presence of other people in that location we decide to show some sides of ourselves and not others.[5] Self-presentation, thus, is both situated and contextual (Van den Berg, 2010).

How do self-presentations relate to identities? For Goffman, identity literally comes about in and through the self-presentations that we engage in when we interact with others. Identity is the '*dramatic effect*' of such interactions (Branaman, 1997; Goffman, 1959; Van den Berg, 2008). In the eyes of Goffman, identity is simply the sum of all the roles we play in our lives. Thus, identity is not some innate quality, nor an essence in itself. Instead, identity is the socially constructed result of all our engagements with others (Van den Berg, 2008).

When playing roles in each context, individuals hope to present a favorable image of themselves. This is why, for Goffman, *impression management* is a key element of self-presentations. Individuals aim to present self-images that are consistent, coherent and convincing (Goffman, 1959). It is important to them to 'maintain face', and hence to minimize the risk of undermining their presentation before each audience. For one thing, it is important that individuals guard their image in each situation, so that it will not be contaminated by information from other roles performed in other situations before other audiences, especially when information from other roles might discredit a convincing performance in the current situation (Goffman, 1959: 137). For example, a person whose professional role consists of displaying authority, such as a political leader, may try to shield not being in charge at all

[4] See http://www.vn.nl/Standaard-media-pagina/DigitaalVlooien.htm (in Dutch, last accessed on 2013-02-18).

[5] Note that both location and the presence (or absence) of others is relevant for our role choices: we display different sides of ourselves when at home with our spouses than when we receive our colleagues into that same home, and we play different roles and display different sides ourselves when we're in the office after hours alone than when all of our colleagues are present as well.

when at home. Shielding this fact from those encountered in professional life helps him to maintain his professional authority.

But roles can also be discredited or undermined by information that emerges *within* the same situation and role. To explain how this works, Goffman uses a distinction between '*giving*' and '*giving off*' information.

'Giving' information refers to all information an individual actively and intentionally shares to bring across a certain image of himself before his audience, including verbal cues, intentionally used gestures and deliberate facial expressions. Goffman describes this as '*communication in the traditional and narrow sense*' (Goffman, 1959: 2). In contrast, **'giving off' information refers to all the cues that an individual shares *un*intentionally, accidentally or inadvertently at the same time.**[6]

Sometimes the cues that an individual gives off, for example through his or her posture, tone of voice, facial expressions or body language may support and strengthen the image (s)he is deliberately attempting to portray (give) before an audience. However, the cues an individual gives off may also, at times, undermine or contradict the image (s)he is aiming to get across. For example, when an individual has to present a piece of work before an audience, she may act self-assured and composed, yet her trembling hands and shaky voice may unintentionally reveal to the audience that she is nervous about the task at hand.

The conceptual framework we've described here can also be applied fruitfully to online contexts in which individuals present their virtual selves. Evidence shows that when posting information (images, text, music, tags etc.) on a social network site, users tend to attempt to create as favourable an image of themselves as possible (Gosling *et al.*, 2007). They may post their prettiest pictures of themselves, or even use Photoshop to brush them up a little, and often will do their best to show their audience how interesting, fun and happy their lives are. Social network sites facilitate this tendency through a host of mechanisms, ranging from displaying the number of friends a user has ('look at how popular I am') to being able to 'check in' ('look at all the fun stuff I'm doing'), to tagging pictures and messages ('look at me and my friends spending time together'), and so on and so forth. All of these forms of communication fall under the heading of 'giving' information: actively, intentionally providing information about oneself to get across a certain image of oneself before the (online) audience.

As a matter of fact, one line of research on Computer Mediated Communication (CMC), called the Hyperpersonal Model (Walther, 1996), states that online communication enables individuals to manipulate the impressions that others have of them to a higher degree than those generated in face-to-face communication. This is so, because of the following four characteristics of CMC (Walther, 2007: 2541):

(1) Computer mediated communication can be edited. Users can spend much more time weighing their words and expressions.

(2) Users have much more time to edit and construct messages and don't have to respond on the spot.

[6] Another way to phrase the difference between 'giving' and 'giving off' information is to distinguish between *identity claims* and *behavioral residue*. Gosling *et al* define these two concepts as follows: "*Identity claims* are the symbolic declarations that individuals make to themselves or others in an attempt to convey how they would like to be seen. Examples of identity claims range from subtle clues found in an individual's clothing choice to more direct claims, like bumper stickers or explicit verbal statements made about beliefs. *Behavioral residue* refers to the inadvertent clues left by one's behavior." Samuel D. Gosling, Sam Gaddis and SImine Vazire, "Personality impressions based on Facebook profiles," Paper presented at ICWSM'07, at Boulder, Colorado (USA): page 1

(3) When users compose messages they are not in physical proximity to the receiver(s), and hence "senders do not exude their natural physical features and non-deliberate actions into the receiver's realm of perception. There is much less ''leakage'' in CMC since there is no unwanted nonverbal indication of undesirable affect or attitude" (Walther, 2007: 2541). In other words, they don't 'give off' the same types of information as they would in offline communications. Having said that, as we'll discuss extensively below, users *do*, in fact, give off quite a bit of information in their online communications. It's just not the same information they would give off in offline contexts.

(4) Users can use all of their cognitive resources to focus on the content and form of the communication, rather than having to take into account a variety of other cues from the environment as well.

Empirical research reveals that users communicating with others via computers are more self-aware, potentially because of these four features of computer-mediated communication, than those individuals who interact face-to-face (Okdie *et al.*, 2011).

While the internet offers us unprecedented facilities to manage the impressions we leave behind in some respects, at the same time, we can also see that unintentional information may easily seep through in our online self-presentations. If one's 'friend-counter' remains stuck on a very low number ('this person has no friends') or, alternatively, skyrockets to a huge number ('this person is a 'Facebook-slut' because (s)he befriends anyone who asks'), the unintended message is that a user has issues finding or maintaining (real) friendships. But this is not the only way in which identity information is 'given off' in online contexts. As we will see below, profiling also makes extensive use of information that is 'given off', which in turn may have significant consequences for a user's online identity. To see how this works, we will first need to know a little more about creating, expressing and maintaining identities on the internet.

2.2. Constructing online identities: 'presented self' vs. 'imposed self'

When going online we leave behind our bodily, physical selves, but on the internet we also have identities. This is so in a very basic sense – whenever we surf the internet, there is *identifying information*, in the form of the IP address of the computer we use, to establish who we are, at least in a very minimal sense. We also actively share identifying information, for example when we provide an online store with our name, address and credit card information to buy a product and receive it in the mail. Aside from sharing identifying information, we also *present* our *identities* in more elaborate senses, for example when creating a profile on a social network site to engage in contact with friends and sharing stories about who we (think we) are. On all of these levels, but especially on the second and the third, we share personal information about ourselves, our identities, with others via the internet. As computer scientist Andy Clarke explains, this entails that our activities on the internet give rise to a certain 'online image' of each and everyone of us, which he calls our *digital persona* (Clarke, 1994). He defines this digital persona as '*a model of an individual's public personality based on data and maintained by transactions, and intended for use as a proxy for the individual*' (Clarke, 1994). The digital persona, then, functions as a proxy for the offline self of an individual.

Clarke points out that we do not have full control over the persona that we present online – others also contribute to the image that exists about each of us online, for example by posting messages or pictures in which we figure online. This is why he makes a distinction between what he calls a person's *projected persona* and his or her *imposed persona* (Clarke, 1994). The former refers to the image that *a person him/herself* attempts to leave behind online, while the latter refers to the image(s) created *by others* about the same individual. Since we

find the chosen term 'projected persona' somewhat confusing[7], we will use the concepts of a *presented persona* and an *imposed persona* instead.

The *presented persona* refers to the **person*'s own self-presentation online***, whereas the ***imposed persona*** refers to ***third parties' images which are attributed to that person***.

As Clarke rightly points out, individuals may choose to have multiple presented personae, i.e. they may create different representations for themselves for different online settings. This aligns nicely with Goffman's perspective on role playing, which we encountered above. Dividing one's online identity into various different personae not only recreates artificially on the internet what 'naturally' occurs in real-life situations (viz. that we play different roles in different places and for different audiences), but it is also an important key in protecting users' identities: when users have multiple presented personae, it becomes more difficult to track, trace and to profile them, or at least to build up a 'complete' picture of the individual behind these personae. Creating multiple presented personae, then, is not only an adequate representation of identity expression in the offline world, but also a safety measure in online environments, a way of insulating and protecting one's online (and, by implication, offline) identity. Hence, this solution will also be a part of the learning goals for the children and teenagers participating in this project.

One of the issues Clarke mentions in his discussion of performed versus imposed personae is the idea that while the individual has a reasonable amount of control over creating and maintaining *performed* personae, much less control can be exhibited over *imposed* personae. Note that this is much more of an issue in online environments than it is in offline contexts. In offline contexts, as we have seen in our discussion of Goffman's work, individuals can exert considerable control over the images they attempt to leave behind in their self-presentations. They never have full control, as we've seen above[8], because of the cues they may inadvertently give off while giving information about themselves. Yet they still have a marked degree of control over their self-presentations, and this is also due to the physical characteristics of the world in which contextual self-presentations occur: such presentations are bounded, both spatially and temporally (Meyrowitz, 1985; Van den Berg, 2010), and will be witnessed only by limited audiences, of whose presence and makeup we are usually (very) aware (Van den Berg and Leenes, 2011). In online environments, however, audiences may be much larger than we know or realise, and since the information we leave behind on the internet is not spatially nor temporally bounded (boyd, 2008a, 2008b; Bryce and Klang, 2009; Tufekci, 2008; Van den Berg *et al.*, 2011; Young and Quan-Haase, 2009), it may come to haunt us in very different contexts at very different times.

What's more, in our offline lives, for most of us, most of the time, the distinction between a performed and an imposed persona will be quite small. Our audiences will, more often than not, take our performances at face value, and interpret our identities as, by and large, what we express them to be. And even if the audience fails to accept a performance as is, most

[7] 'To project', in our understanding of the verb, can mean 'to convey', but also to 'attribute'. In the former meaning an individual would actively shape their projected persona by attempting to convey a certain message or image, but in the latter is would rather be at the receiving end of attributions made by others. To make a clearer conceptual distinction here, we will replace Clarke's concept of a 'projected persona' with that of a 'presented persona'.

[8] For example, the cues individuals accidentally give off may contradict the ones they are attempting to give, or at times discrediting information may enter the performance from elsewhere, for example seeping in from past performances before different audiences. Moreover, Goffman also spends a considerable amount of *The presentation of self in everyday life* on describing situations in which others may affect our performances, both positively (when teaming up with us, acting as accomplices, colleagues or mediators), or negatively (when acting as informers, competitors, hecklers etc.) See Erving Goffman, *The presentation of self in everyday life*. (Garden City, NY, USA: Doubleday, 1959): Chapter 4.

individuals will be socially skilled enough to pick up on this and adjust the performance on the fly so that a more believable picture will emerge. Only in some cases will a person's presented self and their imposed self deviate to a considerable degree. Think for example of a politician who hopes to display drive and strength (presented persona) by cutting deals with other parties so that political issues are resolved quickly, yet comes to be known as hasty, weak and lacking ideology (imposed persona) in the media precisely because of this course of action.

In online self-presentations the difference between a presented and an imposed persona can be much larger for all of us (though of course this need not always be the case (Evans *et al.*, 2008)). This is partly due to the fact that others have extensive abilities to respond to our self-presentations, and to do so with a lasting impact. When we share information about ourselves, for example in a social network site, others have a wide array of tools at their display to respond to this presented persona, and may sometimes, in response, impose ideas about us that contrast with, expand, or alter this presented persona into an imposed one. What's more, our audience can do more than merely respond. Others can actively (help) shape our personae in online environments by posting information about us, sometimes even without our knowledge or consent. Thus, friends may post images or messages about us in online spaces and hence actively and intentionally contribute to shaping our online self-presentation. Finally, and this is vital for the current project, an imposed persona may be the result of tacit, salient traces, and of information given off (rather than given), and constructed on the basis of profiling. In the next section we will present a model of how this works.

2.3. Profiling and identity: the model

We have seen that identities emerge, are constructed and experienced in and through, social interactions with others. When we present ourselves to others, this is a mixture of information that is deliberately, actively and intentionally *given*, and cues that are unintentionally, accidentally and implicitly *given off*. We have also seen that in online contexts our identities, on the one hand, consist of self-presentations, which we've called *presented personae*, in which we actively, consciously and deliberately display (ideas about) ourselves, and that others, on the other hand, may project their ideas and images about us onto our selves, leading to *imposed personae*.

What has remained unclear, so far, however, is how *businesses profiling practices* are related to our identities. In order to shed some more light on this process, we will discuss an example of an online company that makes extensive use of profiling: Amazon.com. When a user visits the Amazon store, their purchases and search history are stored in a personal profile. Based on this information, the user receives personalized ads via email and shopping suggestions whenever (s)he visits the store again, both with the aim of helping users find more products to their liking and sell more goods. However, Amazon does not merely use the *personal* shopping and search history of each individual. It combines this history with that of *other visitors* to the store. The purchases and search histories of customers who have looked at, or bought, products similar to the individual will be aggregated collectively to help the latter find even more products that might meet their desires, even if (s)he is not explicitly looking for these specific products. Whenever a customer buys a product, (s)he will be shown a list of other goods that other customers have bought, using the principle of '*planned serendipity*', i.e. '*the assumption that if a group of people bought book A and also bought book B, others who buy A might also be interested in B*.' (Weinberger, 2007: 59) Thus, the collective shopping behaviours of customers are combined with the personal search and shopping history of individuals, and lead to an elaborate and refined profile that is used to make personalised product suggestions.

When looking at the outcome of the combination of collective and personal search and purchasing behaviours in a store such as Amazon's it is immediately apparent that a presented and an imposed persona can differ, simply because the aggregation of these two

sets of data may lead to product suggestions that do not match our personal preferences at all. While Amazon uses very large data sets to generate their product suggestions, we may still, at times, feel a sense of surprise or estrangement at the product suggestions offered to us, since as unique individuals we obviously do not always align with what others have shown interest in.

In the Amazon example above we've described how the personal search and purchasing history of individuals is combined with that of the collective. But oftentimes profiling by businesses builds on much more tacit and implicit recordings as well, for example on the clickstream behaviour of individuals, i.e. on a registration of all of the links they've clicked as they navigate a website, for example an online store. The path that users choose to navigate information says something about what they do and do not find relevant and interesting, and hence recording this information (in great detail), and analysing it for meaningful patterns can lead to more insight in what a user is looking for, and potentially to selling more products or delivering a more personally relevant service.

Note that in this latter case, the information that is used is not left behind deliberately by the user, but rather is distilled from their behaviour patterns. To phrase it in Goffman's terms, this is information that is *given off*, rather than given. Note also that this form of profiling, more than any that we've discussed before, can lead to a significant difference between the persona an individual *presents* online, and the persona that is *imposed* on him or her by a profiling entity. First, this is so, because the profiler uses both the information a person intentionally *gives*, but also information that (s)he inadvertently *gives off* when surfing the web, as we've seen in our discussion of recording clickstream behaviour. Second, this is so, because the profiler *combines* the image of the individual that result from these two sources of information with the information that *others* give and give off. The resulting picture is projected back onto the individual as an *imposed persona*, as is represented in Figure 1 below.

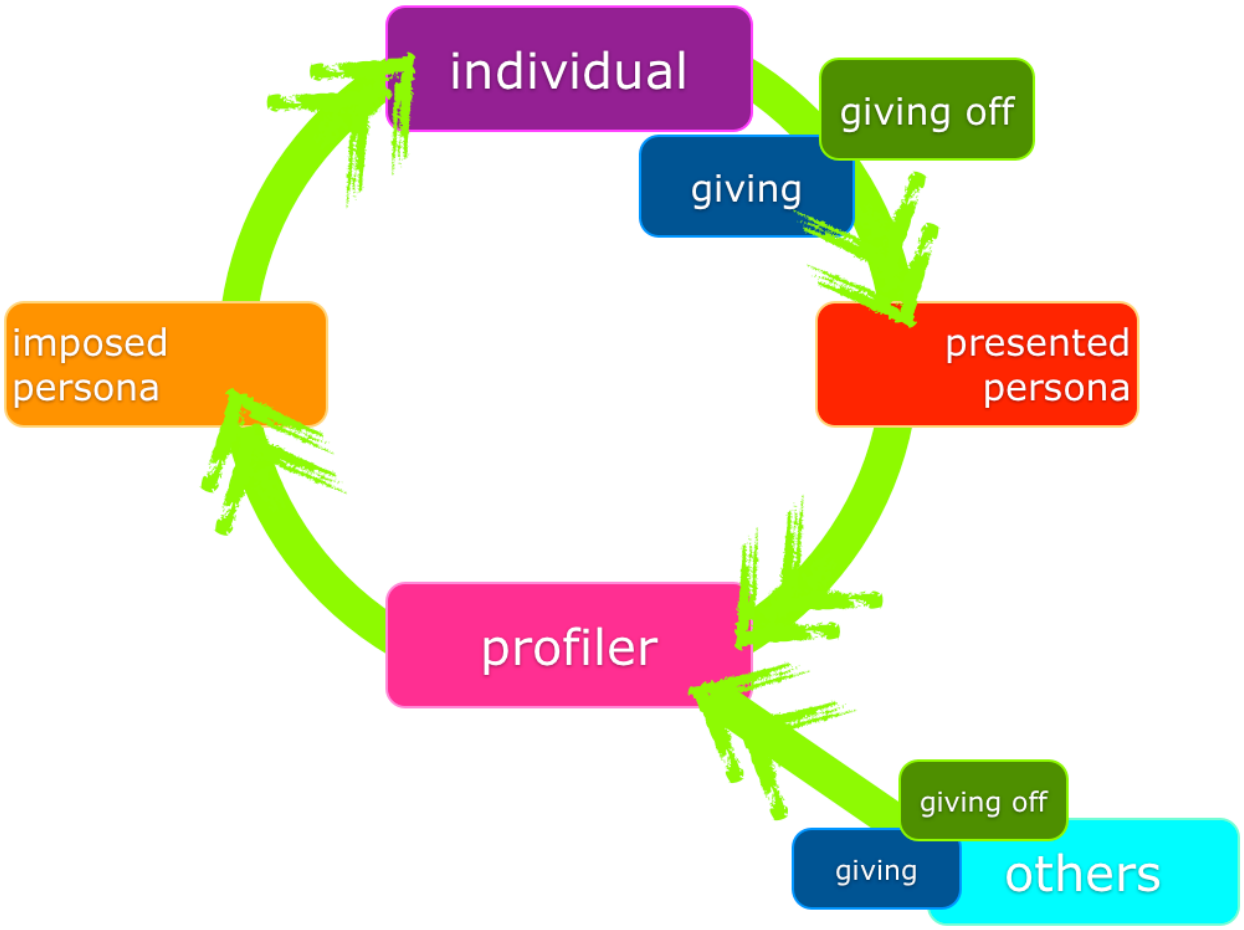

Figure 1: Profiling and identity - the model

How does this **model apply to children**? In short, this figure describes the following process:

- ✓ A child goes online to experiment with or **express his/her identity**, for example in an online role-playing game or on a social network site.
- ✓ (S)he presents an **image of herself/himself**, containing both **explicitly given identity information** (stories about himself/herself, pictures etc.), and information that is **implicitly given off**.
- ✓ The totality of these two we've called the **presented persona**.
- ✓ **Profilers** (e.g. businesses) **harvest the presented persona** of this child and **many others like him or her**, and **create profiles** of, and about, children and their preferences on the basis of that information.
- ✓ They **use these profiles** to seduce children to buy products or sign up for services by presenting them with **targeted advertising** and other personalised messages.
- ✓ The content of such messages adds up to what we've called the **imposed persona** of the same child: a picture of the child's identity and its preferences based on the profiles that businesses generate.

What is most **important** to note in this model is the **potential for a discrepancy between the presented personae of children and the imposed personae** they may be confronted with on the basis of businesses' profiling practices.

As we've argued above, children may not (always) be (sufficiently) aware that they are exposed to marketing messages by companies when they play online games, connect with friends using social media, or engage in identity exploration activities. They may be even less aware that profiling techniques are used by businesses as they experiment with their identities in online role-playing games and social network sites. This is why the goal of this project is to (a) investigate children's levels of awareness surrounding these practices by companies, and (b) improve their skills to recognise targeted advertising and profiling, and to use strategies to avoid or reduce profiling, if they wish.

We will discuss the delineation and goals for this project in more detail below.

3. Lessons to be learnt: or applying the model

In this deliverable we've seen that children use the internet to play, to find information, to interact with friends, and to explore and express their identities. We've argued that while the internet offers a wealth of opportunities for children there are risks and pitfalls as well. In this project we focus on one of these hazards: *profiling*. Businesses and governments use profiling to distil meaningful information about individuals or groups of individuals from large sets of (random) data. We've focused on the use of profiling techniques by businesses in this project. As we've seen, profiling is used by businesses to target individuals with commercial messages and personalised services. This also applies to children – they, too, are profiled when surfing the web, interacting on/through social media, and experimenting with their identities. We've discussed a number of concerns concerning that practice, among them the issue of fairness: since children may not be (sufficiently) aware that they are being profiled by businesses, for example when playing an online game, and that they are insufficiently equipped to value such message correctly and resist the temptations they offer, one could argue that they are exposed to the seductions of a commercialised world at too young an age.

We've also shown that children, like grownups, use the internet to explore, experiment with and express their identities. When going online to do so, individuals self-consciously present images of themselves to their publics – their presented personae. At the same time, however, implicit and less conscious messages about their identities seep through in their self-presentations. They do not merely 'give' information about themselves, but they also 'give off' information. Others can respond to their 'readings' of individuals' self-expressions (both the information that is given and that which is given off), and in the process individuals can find themselves confronted with third-parties' perceptions of themselves. These we've called 'imposed personae'. Most of the time, a person's presented and imposed persona are likely to align reasonably well, but this need not necessarily be the case. The internet, with its wide variety of channels and options for communication greatly increases the risk of the emergence of a discrepancy between the two. This is a serious issue, since an imposed persona may clash with a presented one, e.g. factually or normatively – for example, imagine a child who is claimed to be a bully online, when in fact (s)he isn't. Matters are made worse by the fact that the internet 'doesn't forget', so images that appear online have a tendency to stay online for a long (if not indefinite) period of time.

Profiling techniques by businesses, we've shown, may contribute to the emergence of a discrepancy between children's presented and imposed persona(e), since they use children's complete set of identity expressions (i.e. both the deliberately shared and the more implicit information), and mix these with the identity expressions of other children to create profiles, which they in turn project back onto these children. Aside from the general concerns regarding profiling and children that we discussed above, this raise another issue as well: children may not recognize themselves in the imposed images about them that circulate the web. This is why it is important to increase children's awareness of profiling practices and help them develop skills to protect themselves from the potential downsides of these practices.

3.1. Delineation and goals

Workshops intend to empower children's understanding on how online identities are related to online risks and hazards. The final goal for this project is to increase children's security and sense of control when engaging in online experiences, especially related to identity expression and experimentation. This will be brought about by enhancing children's understanding of internet safety in general, and the workings and applications of profiling in particular.

Specifically, based on the theoretical framework and the model presented above, this project aims to raise awareness of, and generate preventative tools for, the effects of profiling on the online identity expressions and experimentations of children and teenagers. The focus in this project will be on the results of profiling by commercial parties.

The goals for this project are:

- ✓ **Assess children's level of awareness with respect to profiling practices by businesses**
  How much do they know about profiling? Do they understand the technical (im)possibilities of profiling technniques? So they understand the practical implications of profiling practices? Do they understand the benefits and risks of profiling by businesses?
  This assessment will be made on the basis of the project's survey.

- ✓ **Discuss children's ethical stance towards profiling practices by businesses** What are children's opinions about it? How do they evaluate (value) profiling by businesses? Do they mind being profiled as they go about their online activities? Why (not)
  Discussing their ideas on profiling will enable children to develop a critical stance and an informed opinion about profiling by businesses. Facilitating this discussion is the first goal of the educational package to be developed.

- ✓ **Let children experience the benefits and risks of profiling by businesses**
  Let children experiment with the results of profiling by businesses, e.g. by generating multiple controlled presented personae online and watching the effects of using these personae when using Google, or Facebook, or Hotmail (the search results and targeted advertising will be different each time).
  Providing children with hands-on experience of the consequences of profiling will enable them to assess in practice what these consequences are, and to experience what the impact of profiling can be. This will raise their awareness with regard to the frequency and reach of the use of profiling techniques by businesses. Moreover, it will give them a first opportunity to think about strategies to respond to (manipulate, minimise, alter, affect) profiling by businesses if they so desire. Facilitating this experience is the second goal of the educational package to be developed.

- ✓ **Provide children with skills to gain more control over their presented and imposed personae**
  Work with children to develop skills to recognize profiling practices and targeted advertising, to respond wisely to such practices (e.g. not respond to commercial allurements without parents' involvement, or move out of online environments that use these techniques too much, in the eyes of the child him/herself), and, if the child wants to, to know how to minimise, alter or affect profiling by businesses, for example through the use of separate presented personae for different online environments.
  This final goal of the educational package will empower children by increasing their online safety as they surf the web, connect with others and experiment with their identities.

4. References

boyd, danah (2008a). Facebook's privacy trainwreck. *Convergence: The International Journal of Research into New Media Technologies* Vol. 14 (1): 13-20.

——— (2008b). *Taken out of context: American teen sociality in networked publics*. PhD Thesis, University of California, Berkeley, California, USA.

Branaman, Ann (1997). Goffman's social theory. In *The Goffman reader*, edited by C. C. Lemert and A. Branaman. Cambridge (MA): Blackwell Publishers: xlv-lxxxii.

Brown, Ian, and Douwe Korff (2009). Terrorism and the proportionality of Internet surveillance. *European Journal of Criminology* Vol. 6 (2): 119-134.

Bryce, Jo, and Matthias Klang (2009). Young people, disclosure of personal information and online privacy: Control, choice and consequences. *Information Security Technical Report* Vol.: 1-7.

Clarke, Roger (1994). The digital and its application to data surveillance. *Information Society* Vol. 10 (2): 77-92.

Dey, Ratan, Ding, Yan, and Ross, Keith W. (2012). The High-School Profiling Attack: How Online Privacy Laws Can Actually Increase Minors' Risk. Technical Report, Polytechnic Institute of New York University, November 16. (URL http://research.poly.edu/~ross/HighSchool.pdf, last accessed Feb 26th, 2013.)

Duarte Torres, Sergio R. and Weber, Ingmar (2011). *What and how children search on the web.* In: Proceedings of the 20th ACM international conference on Information and knowledge management (CIKM).

Eickhoff, Carsten, Dekker, Pieter, and De Vries, Arjen P. (2012). *Supporting Children's Web Search in School Environments.* In *Proceedings of the 4th Conference on Information Interaction in Context (IIiX)*, Nijmegen, The Netherlands

Evans, David C., Samuel D. Gosling, and Anthony Carroll (2008). *What elements of an online social networking profile predict target-rater agreement in personality impressions?* Paper presented at ICWSM'08, at Seattle, Washington (USA).

Goffman, Erving (1951). Symbols of class status. *The British Journal of Sociology* Vol. 2 (4): 294-304.

——— (1959). *The presentation of self in everyday life*. Garden City, NY, USA: Doubleday.

——— (1961). *Encounters: Two studies in the sociology of interaction*. Indianapolis (IN): Bobbs-Merrill.

——— (1963). *Behavior in public places: Notes on the social organization of gatherings*. New York (NY): Free Press of Glencoe.

——— (1971). *Relations in public: Microstudies of the public order*. New York (NY): Harper Colophon Books, Harper & Row Publishers.

——— (1982). *Interaction ritual: Essays on face-to-face behavior*. 1st Pantheon Books ed. New York (NY): Pantheon Books.

Gosling, Samuel D., Sam Gaddis, and SImine Vazire (2007). *Personality impressions based on Facebook profiles*. Paper presented at ICWSM'07, at Boulder, Colorado (USA).

Gyllstrom, Karl, and Moens, Marie-Francine L., (2010). *Wisdom of the Ages: Toward Delivering the Children's Web with the Link-based AgeRank Algorithm*, Proceedings of the International Conference in Information and Knowledge Management (CIKM).

Hildebrandt, Mireille (2006). From data to knowledge: The challenges of a crucial technology. *Datenschutz und Datensicherheit* Vol. 30: 548-552.

——— (2008). Profiling and the rule of law. *Identity in Information Society (IDIS)* Vol. 1: 55-70.

Lemert, Charles C., and Ann Branaman, eds. (1997). *The Goffman reader*. Cambridge (MA): Blackwell Publishers.

Meyrowitz, Joshua (1985). *No sense of place: The impact of electronic media on social behavior*. New York, NY, USA: Oxford University Press.

——— (2005). The rise of glocality: New senses of place and identity in the global village. In *The global and the local in mobile communication*, edited by K. Nyíri. Vienna (Austria): Passagen Verlag: 21-30.

Mobasher, Bamshad, Cooley, Robert and Srivastava, Jaideep (2000). Automatic Personalization Based On Web Usage Mining. Communication of ACM, Volume 43, Issue 8, August, 2000.

Okdie, Bradley M., Rosanna E. Guadagno, Frank J. Bernieri, Andrew L. Geers, and Amber R. Mclarney-Vesotski (2011). Getting to know you: Face-to-face versus online interactions. *Computers in Human Behavior* Vol. 27: 153-159.

Rubinstein, Ira S., Ronald D. Lee, and Paul M. Schwartz (2008). Data mining and Internet profiling: Emerging regulatory and technological approaches. *University of Chicago Law Review* Vol. 75: 261-285.

Tufekci, Zeynep (2008). Can you see me now? Audience and disclosure regulation in online social network sites. *Bulletin of Science, Technology and Society* Vol. 28 (1): 20-36.

Turkle, Sherry (1995). *Life on the screen: Identity in the age of the Internet*. New York (NY): Simon & Schuster.

——— (1996). Parallel lives: Working on identity in virtual space. In *Constructing the self in a mediated world: Inquiries in social construction*, edited by D. Grodin and T. R. Lindlof. Thousand Oaks (CA): Sage Publications: 156-176.

——— (2011). *Alone together: Why we expect more from technology and less from each other*. New York: Basic Books.

Van den Berg, Bibi (2008). Self, script, and situation: Identity in a world of ICTs. In *The future of identity in the information society: Proceedings of the third IFIP WG 9.2, 9.6/11.6, 11.7/FIDIS International Summer School on the Future of Identity in the Information Society*, edited by S. Fischer-Hübner, P. Duquenoy, A. Zuccato and L. Martucci. New York, NY, USA: Springer: 63-77.

——— (2010). *The situated self: Identity in a world of Ambient Intelligence*. Nijmegen: Wolf Legal Publishers.

Van den Berg, Bibi, and Ronald Leenes (2011). Masking in social network sites: Translating a real-world social practice to the online domain. *IT - Information Technology* Vol. 2011 (1): 26-34.

Van den Berg, Bibi, Stefanie Pötzsch, Ronald Leenes, Katrin Borcea-Pfitzmann, and Filipe Beato (2011). Privacy in Social Software. In *Privacy and Identity Management for Life*, edited by J. Camenish, S. Fischer-Hübner and K. Rannenberg. Dordrecht, Heidelberg, London: Springer: 33-61.

Walther, Joseph B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication Research* Vol. 23 (1): 3-43.

——— (2007). Selective self-presentation in computer-mediated communication: Hyperpersonal dimensions of technology, language, and cognition. *Computers in Human Behavior* Vol. 23: 2538–2557.

Weinberger, David (2007). *Everything is miscellaneous: The power of the new digital disorder*. 1st ed. New York: Times Books.

Young, Alyson L., and Anabel Quan-Haase (2009). Information revelation and internet privacy concerns on social network sites: A case study of Facebook. In *C&T '09*. University Park, Pennsylvania, USA: ACM.