Persuasive technologies and self-awareness: a philosophical-anthropological enquiry

Author: Lorenzo Olivieri

Supervisor: Professor Ciano Aydin

Second examiner: Dr. Michael Nagenborg

External subject expert: Margoth González Woge

University of Twente, Enschede, the Netherlands

Academic Year 2018-19

Faculty of Behavioural, Management and Social Sciences

MSc Philosophy of Science, Technology and Society - PSTS

Abstract

Persuasive technologies are interactive systems designed to change and shape users' behaviours towards specific goals. By discussing a study-case, persuasive ambient mirrors, this work focuses on the new modalities of experience and perception of the self which emerge in the interaction with these devices. The aim is thus to enquire how persuasive technologies shape and transform self-awareness. By addressing this question, this thesis conceptualises persuasive technologies as cognitive artefacts which, through their properties, reconfigure and transform users' cognition and thus their interaction with the world. To do this, I rely on two different theoretical frameworks, "Material Engagement Theory" and "Niche construction theory", which illustrate the role played by objects in the emergence of self-awareness as well as in supporting and extending human cognition. I will argue that while persuasive mirrors might lead to a highly personalized and highly scaffolded form of self-awareness, one which integrates multiple modalities of self-recognition and self-projection, they might also tend to configure a conditioned, standardized self. Finally, I will consider persuasive technologies not only in relation to single users, but also in the light of their possible function in the ecological niches that characterize contemporary world. In particular, I will suggest that (some) persuasive technologies can be conceived as means for coping with the fast technological transformations that occurred in human societies in the last decades.

Table of contents

Introduction

Persuasive technologies are interactive systems designed to change and shape users' behaviours towards specific goals. The variety of cognitive and psychological stimuli and feedback enabled by computing technologies, as well as their potential ubiquity, offers new means for steering and influencing human choices and action, and promises to improve or radically transform previous modalities of persuasion. Unlike traditional persuasive media, like books or newspapers, persuasive technologies are interactive and can be highly personalized; unlike persuasive social actors – teachers, doctors, parents – persuasive technologies can be ubiquitous and rely on larger amount of data. Persuasive systems thus invite users to behave in certain ways, suggesting what is better for them, but they also perform autonomous actions and adopt emotionally or intellectually engaging persuasive strategies in order to constrain and nudge users' actions. For these reasons, persuasive technologies have received, over the last years, considerable attention from a moral perspective, as they seem to challenge the standard understanding of freedom, autonomy and responsibility. However, the impact of persuasive systems on human life is not limited, I suggest, to their possible unethical consequences or to the ways they might modify, and eventually inhibit, the ability for moral reasoning, but it includes also new modalities of experience and perception of the self. The goal of this work is then to enquire how persuasive technologies shape and transform self-awareness.

By addressing the research question, this thesis aims at conceptualising persuasive technologies as *cognitive artefacts* which, through their properties, reconfigure and transform users' cognitive processes and their interaction with the world. The notion of self-awareness is a useful entry for framing my discussion because it allows to focus on two important modalities of self-experience: minimal and narrative self-awareness. Whereas the former refers to a pre-reflective point of origin of action and it is related to the sense of self-ownership and self-agency, narrative self-awareness refers to the sense of being a unique individual extended in time and it includes the ability of remembering past experience (self-recollection) and of thinking, imaging and planning about the future (self-projection). Both these notions of self-awareness are at stake in the interaction with persuasive systems. On the one hand, persuasive technologies aim at modifying the sense of agency, through cues and feedback designed to stimulate users' actions and intentions towards certain directions; on the other hand, they aim at making users' more aware of their past and current actions and at shaping a "future" self who is liberated from its harmful behaviours. The focus on self-awareness will thus allow to analyse how the use of persuasive systems transform users' action in the world and how it changes users' conception of themselves. Moreover, without being ethically charged, the notion of self-awareness might represent a useful conceptual tool for further ethical analysis of persuasive systems, as it offers an anthropological and philosophical starting point for analysing the effects of these technologies on human cognitive system. This work, then, might be considered as a preliminary attempt to connect the philosophical debate about

the boundaries of human cognition with the ethical issues raised by persuasive technologies. While I will limit myself to expose the ethical concerns surrounding persuasive technologies, this thesis might nonetheless prepare the ground for understanding whether, and to what extent, persuasive technologies might modify moral reasoning and moral behaviour.

To enquire the relation between self-awareness and persuasive technologies, I will rely on two theoretical frameworks, Material Engagement Theory and Niche Construction Theory, which will illustrate, from different perspectives, the role of artefacts in integrating human cognitive processes and bringing forth new modalities of self-experience. Rooted in cognitive archaeology, Material Engagement Theory promises to offer a nuanced and diachronic understanding of the role played by objects in the emergence of new modalities of self-recognition and self-objectification. Importantly, this process of material anchoring of consciousness is not limited to human phylogenesis, but it is reproduced, generation after generation, in the individual life span. A second reason for which Material Engagement Theory seems particularly well suited for discussing persuasive technologies and human cognition is its emphasis on material agency. More than other ecological approaches to cognition, Material Engagement Theory in fact conceives of thought and action as the result of the dynamic relation between human and things and hence allows to better understand the peculiarly active nature of persuasive technologies. Niche Construction Theory, on the other hand, stresses human attitude of engineering the environment in order to support intelligent action. In this way, persuasive technologies can be conceptualised as cognitive artifacts explicitly designed to improve decision making and problem-solving. As I will try to show, Material Engagement Theory and Niche Construction Theory offer a complementary view on the relation between persuasive technologies and human cognition: the former by focusing on the extended self and hybrid form of agency emerging in the relation with objects, the latter by treating the construction and use of epistemic tools as a way for increasing human fitness in the world by modifying the informational character of the environment.

To answer my research question, this thesis is structured in three chapters. The first two chapters will break up my research question into narrower sub-questions, while the third chapter will be dedicated to elaborate my answer. The first chapter will seek to answer *what are persuasive technologies and why do they allow radically new, and ethically challenging, modalities of persuasion*. By following the conceptual framework developed by Fogg (2003), I will begin by illustrating the main features of these systems. In particular, Fogg's 'functional triad' will show how persuasive technologies are able to simultaneously play three different roles: they can be tool, by making activities easier or more efficient to do; they can perform the role of social actors, by creating bonds and relations with the users; they can be conceived as media providing symbolic or sensorial information, in the form of texts and graphics or through audio and video supports. Persuasive technologies have thus been designed for a wide range of applications, from health and education to marketing and e-commerce, and they can be implemented in different material supports. Due to the vastness of the field, in this work I will discuss a specific subset of persuasive systems, namely self-imposed persuasive technologies designed for helping and

motivating people to adopt beneficial behaviours and avoid harmful ones. My choice then cuts out persuasive systems imposed by third parties without the explicit consent of the users, as in the case of social networks or websites persuading users to spend more time online and to buy more products. Also, it cuts out persuasive systems imposed by third parties (governments, institutions, companies) to steer the behaviours of citizens, students or employees toward supposedly good behaviours. Self-imposed persuasive technologies are, I suggest, less problematic from an ethical perspective, at least *prima facie*, because, by presupposing the deliberate decision of the users to deploy them, they seem to better preserve freedom of choice and action. In this respect, those who willingly decide to use a persuasive system are also in the condition of being aware of the ongoing persuasive process to which they are subjected and thus might represent a better starting point for discussing the effects of persuasive technologies on self-awareness. To enquire the design and functioning of self-imposed persuasive technologies, I will thus present a study case, persuasive ambient mirrors. The main intuition behind these devices is to provide users with personalized and emotionally engaging feedbacks which visually reflect their alignment and their progress toward a specific goal. To provide accurate feedback, persuasive mirrors track, store and process data about users' actions and present the relevant information back to the users in persuasive modalities. Moreover, they are an interesting case because different design solutions and feedback mechanisms have been proposed, ranging from highly personalized to more metaphorical representations of users' actions. By offering an overview of the incentive strategies used by persuasive technologies, such as operant conditioning and mutual surveillance, persuasive mirrors will introduce the ethical issues at stake with persuasive technologies. Isaiah Berlin's concepts of positive and negative freedom will help to frame the peculiar moral agency of persuasive systems. On the one hand, they limit positive freedom by acting in paternalistic ways and steering the users toward predefined directions; on the other hand they challenge negative freedom by actively interfering with users' behaviour and with their status as autonomous being. However, while this way to look at persuasive technologies is undoubtedly useful from an ethical perspective, it fails to properly recognize the role played by material culture in shaping not only behaviours, but also cognition and selfhood.

In the second chapter I will thus attempt to conceptualise *how prolonged and persistent relations with artefacts transform human experience of the self and its cognitive skills and abilities*. As anticipated above "Material Engagement Theory" and "Niche construction theory" will provide the theoretical frameworks for discussing the effect of technologies on human action and cognition. In Material Engagement Theory, the boundaries between biological and cultural components, between brain, body and things are progressively blurred: human cognition and action cannot be detached from their external surroundings, which, on the contrary, become part of an extended self. In particular, the notion of *tectonoetic awareness* will allow to stress the process of material anchoring of consciousness, namely the idea that a self, or a person, cannot emerge, phylogenetically and ontogenetically, aside from a process of material engagement. By describing the process through which traditional, low-tech objects, such as beads or rings, contributed to the emergence of human self-awareness, it will become clear how

the interaction with artefact have deep effect on the way human beings perceive themselves as agents in the world. On the other hand, Niche Construction Theory will illustrate human ability to physically transform the environment in order to support adaptive behaviour and increase cognitive capacities. Niche Construction Theory is a relatively new approach in evolutionary biology which, unlike standard evolutionary theories, highlights organisms' capabilities to modify the environment in order to enhance their adaptive fitness. While this phenomenon is common to diverse forms of life, it is particularly relevant in the case of humans being. Most of human niche construction activity takes in fact the form of epistemic action, which refers to the intentional modification of the informational character of the environment in order to solve problems and improve decision making. In particular, Sterelny's model of *scaffolded mind* will describe how human cognitive capacities depend on and have been transformed by cognitive artefacts and will delineate a taxonomy of external cognitive resources. It is worth to note, however, that niche construction theory does not consider tools as the only source of cognitive development, but also social learning and the intergenerational transmission of ecological and technical expertise.

In the last chapter I will seek to answer my research question by combining niche construction theory and Material Engagement Theory with the examples of persuasive technologies discussed in the first part. I will suggest that the use of persuasive mirrors can be thought as an example of epistemic action aiming at supporting rational choices and behaviour by transforming complex cognitive problems into easier, perceptual ones. By tracking users' activities, providing customized feedback and cues to action, persuasive ambient mirrors enable new modalities for self-recollection and self-projection, making possible the emergence of a highly scaffolded form of self-awareness. On the other hand, some of the persuasive strategies deployed by these persuasive systems operate upon the psychological and emotional aspect of the users, inviting them to perform specific patterns of behaviours, and thus they might then configure not only a scaffolded and extended self, but also a "conditioned" one, whose actions and decisions are predictable and standardized. To evaluate these alternatives, it might be useful to consider the function of persuasive technologies not only in relation to single users, but also in the light of their possible role in the ecological niches that characterize contemporary world. In particular, I will suggest that some persuasive technologies can be conceived as means for coping with the fast technological transformations that occurred in human societies in the last decades. Examples of this transformation include the nutrition transition which has led to the spread of obesity, the rapid increase of CO2 emissions that deeply alters the ecological balance of the Earth, or the smartphone addiction that threatens attention and concentration. Due to the short time frame of these changes, it might be the case that previous forms of education and social learning have become obsolete, as the members of the previous generations did not have face to face similar challenges. Persuasive technologies could then represent a way for filling this gap.

Chapter 1

1.1 An introduction to persuasive technologies

B.J. Fogg, pioneering figure in the field of persuasive interactive systems, coined the term "captology" to define the study of computer and computing systems as persuasive technologies. Captology, he states, covers the area where computing technologies and persuasion overlap and it focuses on the design, research, and analysis of interactive computing products created for the purpose of changing people's attitudes or behaviours (Fogg 2002, p. 5). "Computing technologies" include applications, websites, videogames, smart and virtual environments, while "persuasion" refers to a set of human behavioural and cognitive phenomena such as behavioural and attitude conversion, motivation, changes in world views and compliance. Two essential features of persuasive technologies are hence interactivity, which marks the main advantage over other traditional media designed for persuasion, such as books, newspapers or advertisements, and interdisciplinary, as they combine human-computer interaction, psychology, motivation and communication science, social marketing and behavioural theories (Fogg et al. 2002). Captology is neutral for what concerns the target behaviours of persuasive systems and indeed Fogg's seminal book *Persuasive technology: using computers to change what we think and do* (2002) contains examples taken from highly different domains, from health care, education and environmental sustainability to e-commerce and social networks. Moreover, since the publication of Fogg's book the idea of persuasive computing systems has become increasingly popular and it has been alimented by other approaches, such as decision supporting systems, gamification (Deterding, Dixon, Khaled, & Nacke, 2011), positive computing (Lee et al., 2019), reflective technologies (Munson, 2012) and mindless computing (Adams et al. 2015) which have widened the field of captology offering new tools and solutions for persuading users. The aim of this chapter, however, is not to provide a comprehensive review of the numerous examples of self-imposed persuasive technologies developed in the last decades, but rather to delineate the theoretical framework which underlies their design, to expose a specific case of persuasive device, namely persuasive ambient mirrors, and, finally, to discuss the ethical challenges related to captology.

Before outlining Fogg's framework, I will briefly linger on the definition of persuasion provided by Miller in his essay *On Being Persuaded: Some Basic Distinction* (1974). Miller's text is particularly useful because, in contrast to Fogg's rather hurried definition of persuasion as «an attempt to change attitudes or behaviors or both (without using coercion or deception)» (Fogg 2002, p. 15), it offers a more detailed account of the process of persuasion. Published in 1974, Miller's text does not explicitly discuss the persuasive role of technology, nonetheless it introduces some of the ethical and political issues which still characterize the debate surrounding persuasive technologies. Miller's interest lies in the definition of the phrase "being persuaded", which he defines as:

> «situations where behaviour has been modified by symbolic transactions (messages) that are sometimes, but not always, linked with coercive force (indirectly coercive) and that appeal to the reason and emotions of the person(s) being persuaded» (Miller 1974, p. 5).

Two relevant characteristics emerge from this definition. First, unlike Fogg, Miller treats the persuasive process as "indirectly coercive". This indirectly coercive force does not depend on the exercise of physical or economic punishment, like with weapons or economic sanctions, but on the credibility of threats and promises expressed in verbal and nonverbal symbols by the persuader (Miller 1974, p. 2). More specifically, the effectiveness of the strategies used to gain compliance depends on the persuader's ability to distribute and inflict rewards and punishments. This coercive dynamic applies, according to Miller, at the individual level, in the case of promises, treats or aversive stimulation, but also at the collective level, where phenomena of social approval and disapproval (altruism, moral appeal, esteem position and esteem negative) hold a strong coercive potential (Miller 1974, p. 3). Second relevant element of Miller's definition is the "appeal to the reason and emotions of the person(s) being persuaded." As Miller notes, the dichotomy between reason and emotions evokes the distinction between convincing and persuading, between rationality and irrationality. Following this distinction, persuasion is said to rely primarily on symbolic strategies that trigger the emotions of intended persuadees, whereas conviction is accomplished by using strategies rooted in logical proof and appealing to individuals' reason and intellect. This dichotomy, Miller argues, is doubtful because ordinary language is already laden with emotional connotations and the very appeal to "be logical" can have a normative force (Miller 1974, p. 4). Persuasive discourse should thus be treated as the combination of rational and emotional elements, where certain messages may differ for the relative amount of each component. Moreover, this duality seems to underlie an ethical one, according to which the influence resulting from logical arguments is ethically preferable to influence appealing to emotions. In regard to this, determining the boundaries between persuasion and manipulation represents indeed one of the most debated ethical issues of persuasive technologies (Spahn 2012, Knowles, Coulton, Lochrie, & Wohl, 2014).

The phase "being persuaded" refers to instances of behavioural conversion, a process which, according to Miller, can take three different forms. First, persuasive process is mainly associated to a *response-changing process*. In these cases, "being persuaded" is equal to "being changed" in the reaction to a certain issue. *Response-shaping process* occurs when individuals do not have established patterns of responses to specific environmental stimuli. This is the cases of subjects who have limited prior learning experiences or with subjects who need to be persuaded in order to learn how to react to radically novel stimuli. Interestingly, Miller notes that a relevant part of children education takes the form of response-shaping process (Miller 1974, p. 6). The process of socialization, for instance, consists of persuading the child to respond consistently to the stimuli received from the social environment. In this cases the persuasive function is performed by parents, teachers or peers, making the distinction between persuasion and learning blurred. Finally, *response-reinforcing processes* aims at strengthening currently

held convictions and at making them more resistant to change. Sunday sermons, Miller argues, have this reinforcing function, which stresses how "being persuaded" is never a one-message proposition, as people are constantly "in the process" of being persuaded (Miller 1974,).

Persuasion, then, is a process of behavioural conversion relying on the manipulation of verbal and non-verbal symbols, which appeals both to the emotions and to the reason of the person persuaded and that might have indirectly coercive force. Captology adds three new elements to this picture: the distinction, from a design perspective, between macrosuasion and microsuasion, the so-called functional triad and the credibility of computing devices as persuasive actors.

According to Fogg, the process of persuasion involves two levels, macrosuasion and microsuasion (Fogg 2002, 17-20). He uses the former to refer to the general objective of the technology, the reason for which the product has been designed. Microsuasion, on the other hand, concerns the elements introduced in the system for shaping and guiding subject's behaviour in order to achieve the overall goal and to incentive the user to continue to use the technology. Microsuasive elements shape the interaction patterns between the persuasive system and the user through the use of dialogue boxes, visual or auditory feedbacks, points and levels.

The notion of functional triad refers to the three different functions – tools, media and social actors - that persuasive technologies can perform simultaneously. A tool, Fogg argues, can be persuasive by making target behaviour easier to do, by leading people through a process and by performing calculation and measurements that motivate. This function is accomplished through the use and combination of different "persuasive technology tools": reduction (making a complex task easier, by reducing the number of steps), tunnelling (users are guided through a predetermined sequence of actions and events, step by steps), tailoring, suggestions, self-monitoring, surveillance, operant conditioning. (Fogg 2002). Likes media, persuasive technologies can provide users with first-hand or vicarious experiences by simulating cause and effect relationships or thanks to virtual environments. Users have thus the possibility to explore and understand the consequences of their actions in a compressed span of time and to perceive in a more tangible form what is perceived in abstract terms. Finally, computing devices can play the role of persuasive social actors by providing physical, psychological and social cues in order to develop a social bond with the users. The popularity of Tamaghotci, according to Fogg, eloquently shows that even a poor stimulus coming from a computer is able to trigger individuals' emotions and to create a social bond. As social actors able to emotionally engage the users, computers can hence be used to reward people with positive feedback, to provide social support or to perform surveillance functions.

Last conceptual pillar of captology is the credibility of computers. The persuasive potential of technologies depends, according to Fogg, on their capacity to be considered as reliable sources of information, instructions, measurements. Credibility, Fogg argues, is a perceived quality emerging from two components: trustworthiness and expertise. The former refers to the perceived goodness or morality of the source, as it happens with professions like judges, physicians or referees. The perception of

expertise, one the other hand, results from the perceived knowledge, skills, and experience of the source, as it occurs with doctors or professors. In the case of persuasive technologies, what is at stake is the credibility of the different actors involved in their development: designers, behavioural scientists and software programmers, but also researchers for health and wellness, education experts or environmental scientists. The novelty of persuasive devices is thus to combine and enhance both the traditional tools of persuasion and the persuasive role historically performed by human actors. As Fogg highlights in the first pages of his book, computers offers then six advantages over human persuaders: they are ubiquitous and more persistent, they offer anonymity, they can rely on a larger volume of data, they can use many modalities of influence, they can scale easily and can go where humans cannot go or are not welcomed (Fogg 2002, 7-10).

Before turning to discuss a specific case of persuasive technologies, it might be useful to briefly outline two relatively new branches of persuasive computing, *mindless computing* (Adams et. al 2015) and *reflective computing* (Munson, 2012), as they represent two opposite direction to persuasive technologies. While these two approaches are still relatively new, they deserve attention because, as I will show in the last chapter, they transform users' agency as well as their perception of themselves in a radically different way. The idea of mindless computing is to design systems that do not rely on users' motivation and ability but rather aim at influencing users' behaviour in subliminal, subconscious ways. As the authors state: «we define a technology as a Mindless Computing technology if it is a mobile or ubiquitous, persuasive technology designed to subtly influence the behaviour of the user without requiring their conscious awareness» (Adams et al 2015, p. 1). More specifically, mindless computing is grounded on dual process theories which postulate a distinction between two cognitive systems said to control human behaviour: *System 1* and *System 2*. System 1, also known as automatic mind, operates quickly and automatically, it effortlessly originates impression and feelings and it does not require voluntarily control. System 2, also called reflective mind, is slow and serial, it operates in a controlled fashion and it is usually associated with the subjective experience of agency and choice (Evans, 2014, Kahneman 2011). The authors highlight that Fogg's persuasive strategies depend on conscious awareness, and implies a strong reliance on motivation and capacity of self-control. Mindless computing, on the contrary, aims at operating below the threshold of conscious awareness, in order to automatically trigger the desired behaviours while relieving the users from the burden of motivation and reflection.

If Mindless computing stands "at the right" of captology, reflective computing seats at its left. The purpose of reflective, or mindful, technology is in fact «to enable users to better know their own behavior, to support reflection and/or self-regulation in pursuit of goals that the users have chosen for themselves» (Munson 2012, p. 2). In other words, the aim is to provide users with relevant data about their own behaviors in order to help them to reflect about it, but without having the system prescribing what to do, inviting to action or setting specific goals that must be achieved. The design of these systems thus does not include microsuasive elements, but rather it merely reveals data to the user. According to

Munson, reflective computing might be a more flexible framework for changing people's behaviours, as it is better suited for handling exceptions or for understanding how different personalities react to the same persuasive strategies.

1.2 A case of persuasive technologies: persuasive ambient mirrors

To better understand how the principles of captology are translated into the design of a persuasive system, this section will discuss persuasive ambient mirrors. The goal of persuasive mirrors is to support behaviour change by providing users with personalized visual feedback that reflects the progress toward a beneficial lifestyle. Whereas a traditional mirror reflects a person's physical appearance and offers a new perspective for perceiving the self, a persuasive mirror reflects back to the person her alignment to desired target behaviours in order to increase awareness about her choices and actions. Persuasive mirrors can alter the image reflected either through augmented reality or by displaying virtual scenarios, and can extract and process diverse sources of users' data. In this respect, persuasive mirrors are a case of ambient intelligent technology, an approach which combines ubiquitous computing and Intelligent User Interface to design "sentient environments" able to recognize users (profiling) and adapt to them (context awareness) in a silent and calm modality (Brey 2005, Brey & Soraker 2007). Persuasive ambient mirrors are thus usually equipped with sensors and optical devices able to recognize user's presence and actions, to collect and analyse user's data over time and to present relevant or persuasive information back to the user in different visual modalities. To this regard, the choice of a mirror - or, more broadly, of an ambient display - as object for persuading people is justified, according to designers, by the amount of time that people usually spent in front of it, often involved in other activities. In addition to this, the purely visual nature of mirrors make them particularly adapt to persuasion, since people are more receptive to visual stimuli than to auditory or tactile ones. (Del Valle and Opelach 2005). The metaphor of the mirror, then, is based on the persuasive role performed by an interactive visual feedback.

Depending on the design of the visual feedback, two approaches to persuasive ambient mirrors can be individuated. One, more faithful to mirrors' traditional function, conceives of persuasive mirrors as enhanced object which transforms user's face reflection or which displays relevant or persuasive information for the user. Early examples have been the *I-mirror* (Ushida, Tanaka, Naemura, & Harashima, 2002) which had a look young-older function and a memory function for remembering and reflecting past scenes in the room, and the *Aware Mirror*, designed to recognize the user and to provide useful information like transportation troubles or weather forecasts. More recently, drawing from similar systems like *Smart Mirror* (Hossain et al. 2007) and the *Medical Mirror* (Poh et al. 2011), the *Wize Mirror* has been designed to be an "index of health status" of the user. The *Wize Mirror* monitors facial signs over time and translates them into cardio-metabolic risks. It also includes a personalized guidance system which provides customised and personalised suggestions in the form of textual or visual clues. The most paradigmatic case is, perhaps, the *Persuasive Mirror* designed by Del valle and Opelach

(2005). In this case the visual feedback is the result of user's lifestyle and compliance to target behaviour: the mirror can deform user's face reflection as the consequence of a harmful behaviour, or match a healthy behaviour with a pleasant colour appearing in the background of the display. As the authors note, in order for the mirror to perform its persuasive role, the visual feedback given has to be grounded on the psychological concepts of positive reinforcement, punishment and reward, likes and dislikes.

The second approach to persuasive ambient mirror has a much more metaphorical attitude since the visual feedback does not reflect user's face or his surroundings, but rather a digital image which changes accordingly to user's behaviour. In particular, the ideas formulated by Tatsuo Nakajima and his group in a number of papers offer interesting examples of this type of persuasive ambient mirrors (Kimura & Tatsuo Nakajima, 2010; Nakajima & Lehdonvirta, 2013; Nakajima, Lehdonvirta, Tokunaga, & Kimura, 2008; Shiraishi et al., 2009). *Persuasive Art* uses digital paintings in order to motivate users' to walk 8,000 steps everyday. The numbers of step is automatically monitored by an electric pedometer and is then translated in the painting, which works as a visual feedback of user's current status of exercise. Different typologies of painting can be used. In the case of a landscape painting, the growth of a tree depends on user's behaviour, so that if the user maintains healthy walking habits, the tree will grow, whereas, if the user stops, the tree will get sick and eventually die. Following a similar logic, *Virtual Acquarium* is designed to improve users' dental hygiene by associating correct toothbrushing practices with the cleaness of a virtual acquarium displayed in the bathroom. When a user begins to brush her teeth, a scrub inside the aquarium starts cleaning algae off the aquarium wall while the fishes start moving in a playful manner. Both the activity of the fish and the movement of the scrub are designed in such a way as to give the user hints regarding the correct method of toothbrushing. To provide accumulated feedback, fishes will start laying eggs in the aquarium if the user maintains the correct behaviour for a certain period of time. Finally, *Ecoisland* is a game-like persuasive application designed to encourage families to reduce CO2 emissions. The application aims at reflecting the current ecological situation: every family is represented by a virtual island and the objective of the game is to save it from the loss of land caused by higher sea levels. The amount of greenhouse emissions is tracked by self-reported data and the sea level rise or decreases according to the activities reported. The sea level metaphor works as symbolic feedback: it evokes the relevance of ecological behaviour and it provides a positive reward when it decreases. To motivate users, *EcoIsland* relies on mutual surveillance, both among the family members and between different families, and on a trading system for selling emissions. Despite different target behaviours and incentive strategies, these examples of persuasive ambient mirrors follow the same design architecture, where central is the concept of feedback loops. A lifestyle tracking component obtains information on user's behaviour though a "sentient artefact". A feedback logic determines how the feedback loop is configured (positive and negative stimuli, accumulated or immediate feedback) and what are the persuasive incentives used (physical, psychological, social, ideological, economic). Feedback information is then presented to the user in

order to emotionally engage him in unobtrusive ways, coherently with the principles of ambient technology. As the authors note, this framework is different from Fogg's more abstract one and, they argue, simplifies the design process (Nakajima & Lehdonvirta 2013, 17). However, besides the differences and analogies with Fogg, what is interesting in persuasive ambient mirror is the idea of an emotionally engaging visual feedback reflecting (the progress toward) a specific behaviour. This kind of persuasive solution is currently deployed by numerous persuasive applications available in different daily objects. For instance, some cars' dashboards currently deploy a similar metaphor of Persuasive Art for making drivers aware of the ecological consequences of their driving habits. The owners of Nissan Leaf can see a pine grows on the display of their cars if they drive in environmentally friendly manner. Similarly, Honda Insight's dashboard has been implemented with a little display field where 'leave icons' increase or decrease in relation to the sustainable or less sustainable driving habits of the drivers. Applying the same arboreal metaphor, *ForestApp* is a smartphone application which help users to "put down the smartphone" and to remain focus on their activities. Users can set a focus time and if they manage to not touch the phone, a tree will grow on the screen. A growing forest represents the accumulated feedback, it shows the healthy trees as well the deaths ones and it can be compared with other friends' forest.

1.3 The ethics of persuasive technologies

Even though the actual efficacy of persuasive technologies is still debated (see Fritz 2014 for a study on long-term impact of health and fitness persuasive technologies), their potential impact on human behaviour has received considerable attention by moral philosophy. An interactive object as the persuasive mirror aims in fact at transforming how people make decisions and perceive themselves, by recommending and prescribing what to do and when, by providing constant feedback about the consequence of current and past actions and, eventually, by reminding us how we will be in the future if we persevere with the same attitudes. The notion of self-awareness, that I will explore in the next chapter, will attempt to capture this peculiar dynamic that persuasive technologies establish not only with present experience, but also with the past and future self.

However, in envisioning a future where persuasive ambient mirrors are blended into the environments of our daily life, Nakajima and Liedhonvirta consider whether it is a better lifestyle for the future, as these pervasive ambient lifestyle feedback systems could take control of our attitudes, causing serious ethical problems (Nakajima & Lehdonvirta, 2013). These problems can be summed up in three questions. First, how do persuasive technologies interfere with human freedom and moral responsibility (Brey, 2005; Guthrie, 2013; Nagenborg, 2014; Rughiniş, Rughiniş, & Matei, 2015; Spahn, 2012; Verbeek, 2009); second, what are the ethical responsibilities of the designers of persuasive technologies (Berdichevsky & Neuenschwander, 1999; P.-P. Verbeek, 2006); finally, how to protect users' privacy in the light of the amount of personal data collected and processed by persuasive technologies (Brey, 2005; Est, Timmer, Kool, Nijsingh, & Rerimassie, 2016). For the purpose of this

work I will discuss the first of this problem, as the notion of self-awareness can be thought as a condition of possibility to freedom and moral responsibility. How persuasive devices transform users' perception of themselves and to what extent they transform self-awareness, are in fact questions from which depend the exercise of freedom and moral reasoning.

However, few lines about designers' responsibility will serve as a methodological note to my research. The scope of this study is to enquire how individuals' self-awareness is transformed and shaped by the use of a persuasive computing object, but persuasive technologies are not produced in a vacuum. Beyond the users, who experience the object in more or less specific modalities, and the artefact that provides feedback and suggestions, there are designers, who make the choices and develop the solutions that give final shape to the object. Behaviour emerges then as the result of these three different agencies, each one also representing a critical moral point: the persuasive methods embedded in the persuasive object, the intended or unintended outcomes of persuasion on the user, the motives of the persuaders (Verbeek 2006). But the motives of the persuaders, designers and designers' employers, are not always easy to establish. Private companies, which develop and design the vast majority of persuasive systems and applications, might have interest not only in improving users' life but also in increasing the number of purchases, downloading or online views of their products. For commercial purpose the design of these systems might take advantage of human physiological and psychological features to steer the users to keep using them or make more difficult to opt out of them. The study of the motivations of the multiple actors involved in the design of persuasive technologies would fall outside the objective of this research, but this, I hope, does not mean to underplay or neglect the role that designers' inevitably have in determining how users experience the persuasive system.

Having said this, the challenge to freedom and moral responsibility posed by persuasive technologies depends on their peculiar interference with human actions. As Verbeek notes, those who adapt their lifestyle because a Persuasive Mirror has repeatedly confronted them with the potential consequences of continuing on the same harmful behaviours are not taking a fully autonomic decision but are allowing themselves to be educated by technology (Verbeek 2009). In other words, a portion of our freedom of choice is delegated to a persuasive system, but this "act of delegation" has a twofold nature: persuasive systems actively help people to behave in certain ways, but, in doing so, they might deploy indirectly coercive mechanisms that help to maintain the correct behaviour or to start a new one. This ambivalent nature of persuasive systems is usually framed through Berlin's famous distinction of positive freedom and negative freedom (Berlin 1969). Negative freedom consists in the absence of any form of deliberate interference of other humans; it is a 'space' in which the subject should be left free to act or make choices without any impediments. When this condition of "liberty from" is not respected and actions are limited by external factors, then the subject is being coerced. The notion of positive freedom, on the other hand, refers to the actual source of someone's actions, choices and thoughts and it is linked by Berlin to the concepts of self-mastery and self-government. We wish, Berlin argues, to be moved by reasons and conscious purposes that belong to us and not by uncontrollable and external

causes that may affect our actions (Berlin 1969, 131). Irrational or overwhelming impulses and desires, as well as the seek for immediate pleasure, are thus the obstacle to a "dominant self" able to rationally guide individual choices and behaviours.

Discussing Ambient Intelligence Technology, Brey (2005) notes that «paradoxically control is supposed to be gained through a delegation of control to machines. In other words, control is to be gained by giving it away. But – Brey asks - is more control gained than lost in the process?». Brey concludes that even though these systems might enhance our freedom by making certain goals easier to reach, they nonetheless have the potentiality to negatively affect both notions of Berlin's freedom: they limit negative freedom by confronting humans with object performing autonomous actions which constrain the liberty of choice; they limit positive freedom by prescribing needs, telling what to believe and what is good. These conclusions apply to Ambient Intelligent and are valid also for persuasive technologies imposed by third parties, but they might be more questionable in the case of a self-imposed persuasive system. As Berlin notes, the notion of positive freedom suggests a "divided man" in which the empirical self, with his contingent passions and feelings, has to be subjected to the will and control of a real self, identified with reason and driven by ideal purposes (Berlin 1969, p. 134). In this respect, the decision to purchase and use a persuasive ambient mirror might be interpreted as a deliberate act of positive freedom that, ironically, points exactly to reconcile the "empirical self" unable to reach certain goals with the "rational self" aiming at a healthier lifestyle. In other words, starting to use a persuasive technology may be seen as a free, voluntary act stemming from a person's awareness that she would not be able, by herself, to follow the desired lifestyle and thus she willingly chooses to receive the support of external aid.

One might counters this claim by pointing to the quality of agency which emerges from the use of these technologies and that appears, overall, substantially decreased. Delegating control to these devices, in fact, may be seen as the acritical acceptance of external suggestions, and, as a consequence, as something that progressively erodes the need of self-reflection about one's own actions. In this sense, self-imposed persuasive technologies might also encourage moral laziness, providing a form of instant morality which could eventually lead to the commodification of morality (Verbeek 2009, 236). Along similar lines, Guthrie (2013) describes the use of persuasive systems in terms of "outsourcing of conscientiousness". But, according to Guthrie, the moral issue is not so much that persuasive devices diminish the capacity of moral reflection, because, he argues, smart technologies or nudge applications already make many contextual judgments for us and promise to do so more and more. Rather, what is more ethically relevant are the more subtle ways through which persuasive technologies address our unconscious self and shape our characters (Guthrie 2013, 328).

On the other hand, there might be good reasons for defending the paternalistic function performed by persuasive systems. Libertarian paternalists, for instance, start from the assumption that human rationality is not something pure. Findings from behavioural science have in fact shown the bias inherent to human reasoning, which suffers from cognitive illusions that are virtually common to everyone and

that make people not simply irrational, but predictably irrational (Ariely 2008). Moreover, the context in which a choice is made often influence decision processes that are not always characterized by full rationality, perfect information processing and complete self-control (Binder and Lades 2015). Drawing from these scientific results, the nudging program (Sunstein and Thaler 2008) call for the development of governmental, institutional and commercial solutions for overcoming people's cognitive biases and steer them toward beneficial choices. The core idea is that since every choice architecture, like default rules or the framing of a question, affects and influences people's decisions, then institutions or corporations cannot avoid a kind of nudging and should develop contexts of choice which promote people's welfare. The focus of Thaler and Sunstein is broader than computing technology, but it can be fruitfully applied to it. The Mindless Computing approach described above, for instance, applies nudging principles by altering user's environment to exploit human cognitive features and bias. According to its advocate, the strength of libertarian paternalism is that, despite his paternalist nudging function, it leaves the possibility to opt out or stop using the system and thus preserves subjects' freedom (Sunstein 2007).

However, to what extent persuasive technologies interfere with negative freedom is a debated issue which largely depends on the ethical interpretation of the (indirectly) coercive strategies included in the system. Fogg mentions deception and coercion as "always unethical" persuasive methods, but he also identifies an ethical grey-area including the use of emotions, surveillance and operant conditioning. These persuasive methods, he argues (Fogg 2002, pp. 221-226), raise a red flag and might be ethically acceptable or inacceptable according to the types of strategies applied. As we have seen, persuasive ambient mirrors are, in this regards, less morally scrupulous, since they combine the principles of operant conditioning (positive and negative reinforcement and positive and negative punishment) to encourage or discourage user's behaviors (Nakajima & Lehdonvirta, 2013, 14). In *Ecoisland*, mutual surveillance (being monitored by others, watching others, comparing with others) works as a negative reinforcement when someone feels bad for his negative contribution. As the authors recognize, this kind of feedback is unavoidable in persuasive applications for collectivist societies because users have negative feelings when they feel not like the others even if the application offers positive feedback (Kimura & Tatsuo Nakajima 2010, p. 704). But, according to Nagenborg (2014), it is the very inclusion of these microsuasion mechanisms exploiting psychological human traits, such as the "follow the herd" effect, that makes the system unethical. These microsuasion elements are introduced to promote compliance but, simultaneously, they also lead users to continue using the system, or, on the contrary, make less likely for them to opt out. Persuasive technologies would then limit negative freedom not because they perform autonomous actions, but because they threaten the status of users as autonomous beings. (Nagenborg 2014, p. 46).

As Nagenborg acknowledges, framing the debate by emphasizing the notions of autonomy might reflect a modernist approach to human-technology relations, one which postulates a "pure" freedom of choice and action which precedes any interaction with the mundane world. On the contrary, according to Verbeek (2009), persuasive technologies are a radicalization of an influence that objects have always

had on human behaviour and freedom. From a post-phenomenological perspective, in fact, any relation with an object determines a form of technologically mediated intentionality, a hybrid resulting from people's intentions and the material intentionality of the objects. Technology, Verbeek argues, is not at odds with freedom, rather it «contributes to the constitution of freedom by forming the material environment in which human existence is enacted and takes shape» (Verbeek 2009, 238).

Verbeek's insight suggests that, to understand whether and to what extent persuasive technologies might revolutionize moral life and reasoning, an anthropological analysis of the relations between humans and technologies is needed. The next chapters of this work will be dedicated to this purpose, but rather than relying on post-phenomenology, I will turn to two other frameworks, Niche Construction Theory and Material Engagement Theory. More specifically, the notion of tectonoetic awareness will allow stressing the process of material anchoring of consciousness, namely the idea that a self, or a person, cannot emerge, phylogenetically and ontogenetically, aside from a process of material engagement (Malafouris 2008). Persuasive devices could then be thought as a highly personalized form of tectonoetic awareness, one which enhances the possibilities for self-recognition and self-projection. From the perspective of niche construction, the use of persuasive technologies can be considered as an epistemic action aiming at supporting choices and behaviour by willingly alter environmental features.

Chapter 2

2.1 Self-awareness and material engagement: an archaeological perspective

The previous chapter offered an overview of the main features of captology and presented a study case, persuasive ambient mirrors, in order to have a closer look at the design and functioning of persuasive systems. In persuasive mirrors users' behaviour is first monitored and processed and then presented in the form of a visual feedback reflecting users' progress towards specific goals, such as decreasing CO2 household emissions or leading a healthy lifestyle. From a moral perspective, the novelty of persuasive technologies lies in their deliberate, explicit interference with human experience: persuasive systems prescribe to users patterns of action and choice and, in doing so, they might rely on indirectly coercive mechanisms which exploit human psychological or physiological features, such as emotional engagement, operant conditioning or forms of soft surveillance. Due to this ambivalent nature, Berlin's notions of negative and positive freedom are undoubtedly useful for framing the ethical debate about persuasive technologies, but, when applied to human-technology relations, they lack a preliminary philosophical and anthropological analysis of the role played by artefacts and objects in shaping or transforming human beings. To put it metaphorically, they need to be immersed in the mundane, material world inhabited by humans. Berlin himself seems to be aware of this when he writes: «[t]he conception of freedom directly derives from the view that is taken of what constitutes a self, a person, a man. Enough manipulation with the definition of man, and freedom can be made to mean whatever the manipulator wishes» (Berlin 1969, p. 134). This chapter will then be dedicated to enquire «what constitutes a self» by enquiring the role of material culture in shaping human self-awareness. What is the relation between self-awareness and artefacts? How do objects transform and extend the perception of the self? To answer these questions, I will turn to Material Engagement Theory (Malafouris, 2004, 2008a, 2008b, 2008c, 2013), a relatively new approach in cognitive archaeology aiming at understanding how material culture changes and re-organizes cognition during human phyologenesis and ontogenesis. Stemming from archaeology, the goal of MET is primarily to offer a radically new perspective on the evolution of human mind as well as on the interpretation of archaeological findings, nonetheless its theoretical framework can be fruitfully applied to contemporary technologies (Aydin, González Woge, & Verbeek, 2018; Clowes, 2018; Poulsgaard, 2017).

Malafouris's critical targets are archaeological and evolutionary interpretations which study the development of human cognitive evolution starting from the assumption that cognitive skills and abilities precede the production of material culture. According to this view, minds *enable* and *come before* culture and behaviour, so that, for instance, the production of tools by early hominins is considered as the result of a mental template which enabled the realization of the tool (Jeffares 2010). As a consequence, changes and transformation in material culture are treated as epiphenomenal reflections of, or proxies for, pre-defined aspects of human thought (Malafouris 2013, p. 11). This

conception, according to Malafouris, is flawed by a modernist, computational and neurocentric view of cognition which identifies the mind with the brain and that conceives of cognitive processes as disembodied and abstract operations. On the contrary, Malafouris argues for what he describes as "the constitutive intertwining of cognition with material culture". His main methodological move is in fact to re-define the supposed boundaries of cognition by choosing a different "unit of analysis" (Hutchins 2010), one which includes not only the brain, but also the body and, above all, the material agency of objects. In this way, rather than being situated internally, within the biological boundaries of the skin and the skull, human agency and cognitive processes are conceived as the result of a dynamic interaction between matter, body and brain. In putting forward his theory, Malafouris combines archaeological records with neuroscientific evidence and philosophy of mind. Neuroscientific research has in fact shed light on the highly plastic nature of the human brain throughout the course of human life, showing that the use of tools and artefacts leads to a re-modelling of the functional, structural and anatomical functioning of the brain (Malafouris 2013, pp. 45-50). From the perspective of MET, this neural plasticity enables to view the interaction between brain and objects as a dynamic relation, a co-evolution in which the brain, far from being a static structure, is constantly shaped and transformed by the interactions with its ecological and material surrounding. For what concerns philosophy of mind and cognitive science, MET is inspired by theoretical approaches which stress the embodied (Maturana & Varela 1987, Varela et al. 1991), ecological (Hutchins 2010, Bateson 1970) and extended (Clark & Chalmers 1998) dimension of cognition. Drawing from these frameworks, Malafouris develops the notion of *extended self*, namely «a self that is located neither inside nor outside the brain/body, but is instead constantly enacted in-between brains, bodies and things and thus irreducible to any of these three elements taken in isolation» (Malafouris, 2008b, p. 1997). Importantly, then, this *extended self* should not be thought as made up of different overlapping layers, where the social, cultural or technological component are progressively added to a biological, inner core. Neither, Malafouris continues, the extended self is merely an embodied self, but rather «a self enacted through the act of embodying».

As cognitive archaeologist, Malafouris is especially interested in the emergence of human sense of self-awareness. To clarify this concept, Malafouris relies on a distinction between two notions of awareness, *noetic awaren*ess and *autonoetic awareness*, which echo other philosophical conceptions of the self (Gallagher 2000, Tulving 1983). Malafouris uses *noetic awareness*, or noetic consciousness, to refer to the basic sense of oneself as acting in and on the environment at a time, according to one's first-person perspective. This level of noetic awareness corresponds to what Gallagher defines as *minimal self*, namely the consciousness of oneself as an immediate subject of experience unextended in time (Gallagher, 2000, p. 15). The minimal self is a pre-reflective point of origin for action, experience and thought in which access to first-person experience is immediate and non-observational. To the concept of minimal self Gallagher associates two other elements: the sense of ownership and the sense of agency. The former refers to the sense that is my body which is undergoing an experience, to the feeling of mineness experienced towards our body parts, feelings or thoughts. The sense of agency refers to the

sense that I am the initiator or source of action. As Gallagher (2010) states, the sense of agency is complex, as it includes multiple components, from sensory-motor processing to higher-order cognitive components involving intention formation. The notion of *autonoetic awareness*, on the other hand, refers to the ability of reflecting on oneself from a third person perspective and it introduces a temporal dimension in the perception of the self, namely the process of self-recollection (the mental reinstatement of past events and experiences) and self-projection (the ability of thinking, imagining and planning about the future). Autonoetic awareness has strong affinities with Gallagher's *narrative self*, a more or less coherent self (or self-image) that is constituted with a past and a future in the various stories that we and others tell about ourselves (Gallagher 2000, pp. 18-19). This narrative, continuous self is extended in time to include memories of the past and intentions toward the future and, for these reasons, the proper functioning of episodic memory and time-sense are particularly relevant for the formation of narrative self (Gallagher 2000, p. 20).

In concluding the paper, Gallagher poses a number of "outstanding questions": What relationship exists between the minimal self and the narrative self? Is one generated from the other? Do they operate independently of each other? And, from an evolutionary point of view, why Great Apes seem to have a core self-system as well as the notion of an autobiographical self, but lack a conceptual, reflective selfhood? The answer to these question lies, according to Malafouris in the production, use and proliferation of artefacts which enabled human beings to extend into their material surroundings and simultaneously to detach themselves from the temporal a spatial contingencies of ordinary, subjective experience. To capture this process, Malafouris forges the notion of *tectonoetic awareness*. As he writes:

> «The basic assumption behind tectonoetic awareness is simple: a self or a person cannot emerge (ontogenetically or phylogenetically) aside from a process of material engagement. Tectonoetic consciousness should not be understood as a distinct separate stage between the two - although this can be argued to be the case from an ontogenetic viewpoint - but as a scaffolding process of ongoing structural coupling that grounds in action and integrates the noetic and autonoetic aspects of selfhood» (Malafouris 2008, p. 1998)

The use of the prefix "tecto-", Malafouris explains, aims precisely at overcoming the limits of terminology in the cognitive sciences, which fails to capture the 'act of embodying' as a continuous and interactive coordination between neural and extra-neural physical resources (Malafouris 2008, p. 1997). Integrated in human cognitive processes, artefacts provide then the necessary scaffolding, the material support for the emergence of a new experience of the self which would be impossible for the "naked" body.

Two examples well illustrate this process of material anchoring of consciousness: early body decoration (Malafouris, 2008a) and a Mycenean golden ring (Malafouris, 2008b). Starting with the former, the forty-ones Nassarius Kraussianus shells found at the Blombos Cave in South Africa proved the use of personal ornaments about 75 kya. While there are no doubts that these shells were intentionally collected and transformed into beads, the question regarding their possible symbolic or representational function remains open. Malafouris suggests to reverse the standard archeological interpretations which

assign symbolic value to body-decoration, because, he argues, the presence of reflective self should not be taken for granted. Rather, it is precisely through the material engagement implied by the use of those beads that a new type of self-awareness and self-knowledge became possible. In this regards, Malafouris describes the beads as a dual entity. As attached and visible prosthesis, beads modify the perception of the body and reorganizes self's cognitive system. Particularly, Malafouris refers to neuroscientific research (Berti & Frassinetti 2000) which shows how the use of objects have strong plastic effects in the cognitive topography of peripersonal space, namely the behavioural space that immediately surrounds the body within the hand-reaching distance. On the other hand, pointing outward, beads address others hominins and thus also function as a second, social skin. However, Malafouris carefully underlines that this social skin should not be understood symbolically in terms of social identity, such as gender or group identity, but as the process *through which* reflexive awareness and the self-other distinction emerge. What is crucial, then, is that due to their materiality, to their enduring presence, beads decouple the self from the here and now of ordinary experience, from the feeling of being in the world, and allow the possibility of being reflectively conscious of oneself.

A second powerful example deployed by Malafouris is a Mycenean gold signet ring depicting a battle scene found in a grave in Mycenae. What, Malafouris asks, can such a ring tell us about the Mycenean self? What was its causal efficacy on the cognitive and self-system of Mycenean warriors? Malafouris' discussion starts with an analysis of the Homeric self, since, according to a classical interpretation (Snell 1960), Homeric heroes lack the full self-consciousness of a unitary, integrated self that acts as an autonomous being and that makes decisions of which he is aware. Agamennon and the other warriors act as if decisions are made *for* them rather than *by* them, leading interpreters to treat Homeric heroes as lacking a proper, fully-developed notion of selfhood. This interpretation, according to Malafouris, is flawed because it projects back to the Homeric world the western, Cartesian conception of the self as an isolated human agent acting in an inanimate environment. On the contrary, as with the Blombos beads, central is, for Malafouris, the act of embodying: like the beads, the ring touches the body and is experienced as part of it, blurring the distinction between a supposed biological and cultural self, between personal and peripersonal space. But besides transforming the cognitive system of the Mycenean self, the ring also holds a strong mnemonic potential. Most obviously, the iconic scene of war depicted on the ring's surface had, for the Mycenaen warrior, a mnemonic function. Moreover, the ring, as a material object, can take on a separate life on its own. Drawing from Rowlands (1993), Malafouris notes that objects are individually and culturally invested with memories and events associated to their use and ownership, coming to possess their own personal trajectories. Things, Malafouris writes, «effect temporal anchoring and binding. They help us to move across the scales of time and to construct bridges between temporal phenomena that operate at different experiential level» (Malafouris 2013, p. 247). The ring thus enables the Mycenean warriors to move along multiple temporalities, and in this sense, it embodies a "dynamic cognitive biography" which redefines the boundaries of working memory, overcoming the biological ones. It is through this «complex associative

enchainment between the 'internal' and 'external' elements of remembering» that the ring is able to bring forth a new kind of autonoetic consciousness, making possible explicit self-recognition through objectification (Malafouris 2008b, p. 1999). It is important to note, moreover, that Mycenaean self-objectification did not depend on one specific object, but it was realized through a variety of material media. So, according to Malafouris (2008c), the sword of a Myceanean warrior did not represent just a weapon, but rather it extended and transformed his very self by embodying a novel concrete situational perspective of being-in-the-Mycenaean world.

In principle, then, any modality of interaction with tools and objects might enable the transition to autonoetic awareness. Tools use and manufacture, for instance, played, before body decoration, a decisive role in the development of new levels of awareness (Jeffares, 2010). Manufacturing tools and taking care of them provided in fact hominins with a way for focusing their mind and keep them on tasks. The very act of carrying a stone tool, according to Jeffares, facilitated, through its physical presence, to stay on task and to not get distracted. Stone tools then, had a decisive function in the emergence of human intentionality, but what is crucial for Malafouris is that specific modes of tectonoetic awareness depend on the qualities inscribed in the objects. Beads, being permanently attached to the body, integrates the self with "epistemic qualities" which cannot be provided by tool use and manufacture. The same is true for the gold signet ring, which, by depicting an iconic scene of war, constantly reminds to its owner his place in the world, becoming a means of self objectification.

2.2 Niche construction and epistemic action

While the previous section showed, through the notion of tectonoetic awareness, how artefacts transform the experience of the self by offering multiple modalities of self-objectification and self-recognition, this section will explore how humans intervene and modify their environment in order to increase their adaptive fitness and to support thinking and reasoning. To discuss this "constructionist" attitude of living beings, and particularly humans, I will turn to a family of *niche theories* – niche construction theory (Laland & Kendal, 2015; Laland, Matthews, & Feldman 2016), cognitive niche theories (Sterelny 2004, 2006, 2010, Clark, 2005b, 2006) – which, by combining evolutionary biology, anthropology, philosophy and cognitive science, open new perspectives for understanding the processes which have shaped and that still characterize human cognition. In emphasizing the co-evolutionary and developmental dynamics between organisms and their environment, these theories have strong affinities with MET. However, as Malafouris notes (2013, p. 246), while his approach aims at showing how things impose their own dynamics, consciousness, and temporality on human bio-cultural evolutionary path, niche theories focus on cumulative ecological and epistemic engineering.

A niche is defined as the structural, temporal, and social context in which a species exists and it includes space, structure, climate, nutrients, and other physical and social factors (Fuentes 2015, p. 304). The main intuition behind niche construction theory is that organisms do not passively adapt to their niches, but rather they actively modify and change them in order to increase their possibilities of

survival. Niche construction theory is built upon four conceptual pillars: (1) organisms modify environmental states in nonrandom ways, thereby imposing a systematic bias on the selection they generate, (2) ecological inheritance strongly affects evolutionary dynamics and contributes to parents-offspring similarity, (3) acquired characters and by-products become evolutionarily significant by affecting selective environments in systematic ways, and (4) the complementarity of organisms and their environments (traditionally described as 'adaptation') can be achieved through evolution by niche construction (Laland, Matthews, & Feldman 2016, p. 192). In the light of this ecological and constructionist approach, niche construction theory marks an important discontinuity with the standard evolutionary model, according to which the evolution of a species results primarily from the interaction between genetic variations and the selective pressure of the external environment, and which leads to conceive of organisms mostly as "bearers" of genetic traits adapted to specific environmental conditions. Even though the seeds of niche construction theory can be found in Lewontin's critique to adaptionism (Lewontin 1983) as well as in earlier scholars of evolution[1], it was through the work of Odling Smee, who first coined the term, that niche construction and ecological inheritance started to be properly recognized as evolutionary processes on their own, along with natural selection and genetic inheritance (Laland et al., 2016).

Crucial consequence of niche construction activities is therefore the production of new feedback cycles which modify the selective pressure in the ecological landscape inhabited by a species. Importantly, these feedback cycles occur both at the phylogenetic and ontogenetic level: on the one hand, activities of niche construction transform the selective environment and thus permanently change the biological evolutionary path of a species, one the other hand this process has a direct analogue within lifetime learning, where the feedback cycles alter and transform processes of individual and cultural learning (Clark, 2005, p. 256). Classical examples of niche construction within the animal kingdom are birds' nest, spider webs or beavers' dams. Dams, for instance, are constructed by beavers in order to defend the colony against predators and to increase the availability of food. The physical presence of the dam modifies the selection pressure on the beavers who build it, but also on their offspring, who, in turn, will grow up in an environment where the river flow has been altered by the previous generation and, eventually, will inherit the dam.

By stressing the proactive attitude of organisms, niche construction theory has obvious implications for the study and interpretation of human culture and behaviour[2]. Particularly, compared to

[1] See Pocheville (2015) for an historical review of the concept of ecological niche

[2] See Wallach (2016) for a critique of Niche Construction Theory as explanatory framework for human phenomena. According to Wallach, Niche Construction Theory is unable to provide explanation for episodes of human past, as it does not specify the relations between causal factors and empirical facts. More specifically, Niche Construction Theory appears to be flawed by overly-generality and non-specifity, hence questioning its usefulness for understanding the evolution of a system, and, ultimately, its validity as a rigorous scientific theory. However, as he notes, authors like Sterelny tend to use the notion of Niche Construction Theory in descriptive ways, in order to illustrate the feedback loops between human activities and capacities and their societal and/or ecological environment (Wallach 2016, p. 2598)

evolutionary psychology, niche construction theory emphasizes the flexibility of human behaviour in adapting to environmental and social challenges. This difference is well captured by the interpretation of the so-called *adaptive lag*, namely the mismatch between past selective factors and the current environmental conditions. While a certain level of adaptive lag can be experienced by all species as consequences of disruptive ecological changes, in the case of human evolution the adaptive lag would be particularly significant due to the massive technological development which have considerably transformed human habitats in the last centuries. In other words, according to the perspective of evolutionary psychologists, human mind and its psychological mechanisms are adapted to the ancestral environments of the Pleistocene and thus unable to cope with some of the environmental, social and technological conditions that shape the twentieth century industrialized world. Against this view, advocates of niche constriction theory argue that humans usually construct their world in functional and adaptive ways, at least from a short-term perspective (Laland & Brown 2006, Laland & Kendal 2007). However, when human niche construction activities do modify the selective pressure of the environment, two alternative pathways can follow: genetic evolution and cultural niche construction. (Laland et. al 2007). The first case is well exemplified by the strong correlation existing across cultures between the presence of the genes for lactose absorption and a history of dairy farming. Strong theoretical evidence supports the hypothesis that dairy farming spread before the genes for lactose absorption and not the other way around. (Laland & Kendal, 2015, p. 55). Alternatively, humans can respond to the effects caused by prior niche construction activities through further niche construction. Human activities, for instance, might lead to pollute the environment, but this polluted environment may stimulate the invention and spread of a new technology to cope with the contamination, alleviating the problem. Similarly, a by-product of the transition into large sedimentary settlements like towns and cities was the production of domestic and industrial waste, which, in turn, led to the spread of diseases, such as measles, smallpox and typhoid. However, after experiencing new self-induced selection pressures, human population managed to find specific solutions, such as the construction of sewerage or drain systems (Laland & Brown, p. 100).

Niche construction activities then contribute to the adaptive fitness of a species and increase its possibilities of survival, but they do not necessarily have an epistemic purpose. Farming or the construction of sewerage systems certainly required high cognitive abilities, but did not aim at supporting or enhancing human cognitive capacities. However, according to the philosopher of biology Kim Sterelny, most of human niche construction activities is a form of epistemic engineering, which he defines as the production of cognitive tools or the organization of the physical environment in order to support and enhance intelligent behaviour (Sterelny 2004, 2006, 2010). Human life, Sterelny argues, is a long decision tree where solving problems, creating chances and making the right decisions require access to and use of different amount of information. The responses to these "high cognitive load" problems, as Sterelny defines them, depend on three different evolutionary mechanisms. First, Sterelny recognizes the role of modules, innate domain-specific cognitive specialisations, such as language or

naïve physics. However, innate modules suit predictability and environmental stability, and thus can only play a limited role in alleviating high load problems. Second, more than any other proponents of cognitive niche theories, Sterelny emphasizes the role of direct and indirect social learning in developing cognitive (and emotional) skills and behaviours. His "apprentice learning model" (Sterelny 2012) describes learning and the acquisition of expertise as resulting from a hybrid process in which trial-and-error learning is combined with observational learning and explicit instruction[3]. Moreover, in the case of humans, the intergenerational transmission of knowledge assume an even more prominent role, since human ontogenetic development is characterized, compared to other species, by a prolonged childhood (neoteny), which, extending the period of neural and behavioural plasticity, optimizes social learning. For this reasons infancy, childhood and adolescence are not merely biological stages, but rather developmental stages of a biocultural niche which supports intersubjective, emotional, communicative and cognitive engagement (Sinha 2015).

However, it is above all by acting as *epistemic agents* and *epistemic tool makers* that human are able to face informational challenges. The construction and proliferation of epistemic artefacts, Sterelny claims, expand human cognitive capacities and mark a qualitative leap with other primates. The transformation of the informational character of the environment to support behaviour and cognition occurs in different ways. Most obviously, humans alter their environment to ease memory burdens, for instance by trail marking for remembering the route or through environmental storage, like diaries or calendars. Also, humans transform difficult cognitive problems into easier perceptual problems, as it occurs when quantitative information is represented as pictorial patterns, pie-charts, graphs. Finally, difficult perceptual problems are transformed into easier ones, like pedestrian crossings painted in bright contrasting colours to stand out against the asphalt, or students highlighting text to help focus their attention and to aid search for later study. Two more considerations apply, according to Sterelny, to epistemic technologies. First, unlike the other evolutionary responses described above, epistemic objects are independent from the pace of environmental changes, so that, even in a fast changing world, they enhance the power of individual learning, and they enable solutions to be spread and improved horizontally (Sterelny 2006, p. 21). Second, the use of epistemic technology is itself informationally demanding. Public representations, models and templates require interpretations and such interpretations need to be learnt. Also, the ability to understand and manipulate public symbol systems is informationally demanding. To make a contemporary example, Clowes (2018) shows how the use of Google Maps for finding a place requires interpretative effort, a series of micro-interactions involving

[3] Interestingly, Sterelny mentions that reliable intergenerational transmission of skills and knowledge can begin as a side effect of adult activity, with no need of formal educational institutions or explicit teaching. The role of intergenerational transmission of knowledge seems particularly relevant in the light of Miller's comment about the subtle boundaries between teaching and persuasion (see p. 3 above). According to Miller, a large part of children education consist of response-shaping processes in which parents or teachers "show" to the children how to consistently respond to stimuli coming from the external environment. But, if Sterelny is right, it might be the case the part of this persuasive process occurs in more indirect ways, without parents and children being aware of it.

careful attention to the app's directions and speculative walking about in order to reach the desired destination.

Sterelny's understanding of cognitive scaffolding and the production of epistemic artefacts as part of niche construction activities well resonates with other frameworks which stress the ecological dimension of human cognition (Clark & Chalmers 1998, Clark, 2005b, Clark 2006, Bertolotti & Magnani, 2017; Magnani & Bardone, 2008). Particularly interesting is the comparison between Sterelny's model of scaffolded cognition (Sterelny 2010), and Clark and Chalmers' extended mind thesis (1998). The extended mind thesis is in fact grounded upon what has been defined as the parity principle, the idea that if an external resources play the same *functional role* performed by a cognitive process taking place in the brain, then the external resource becomes part of the cognitive system of the agent. Central idea of extended mind is thus that the human cognitive system includes those external components that have replaced internal resources. In contrast, Sterelny's scaffolded mind hypothesis is that human cognitive capacities *depend on* and have been *transformed by* external resources. Despite the clear affinities, Sterelny argues that extended mind hypothesis is a sub-set of broader niche construction activities and thus it is less useful from a heuristic perspective, as it tends to obscure rather than highlight the differences among external resources and their contribution to cognitive competences. He thus develops a taxonomy of "environmental fuels for cognition" built upon three different dimensions, each of them corresponding to a functional relationship between an epistemic resource and an agent: 1) trust, 2) interchangeability, individualisation and entrenchment and 3) individual and collective resources (Sterelny 2010, pp. 473-479)[4].

Trust is understood by Sterelny in terms of reliability, which refers to agent's assessment of the reliability of the access to a resource as well as the reliability of the resource itself. Crucial, according to Sterelny, is the fact that external resources are functionally different from internally stored information and hence they are not trusted, or should not be trusted, to the same extent that internal resources are. Paradigmatic, in this respect, is the famous example of Otto's notebook which, in Clark and Chalmers paper, is supposed to play the same functional role of Inga's biological memory. But, according to Sterelny, this is not the case, as Otto's notebook is less reliable after dark or when he forgets his glasses and, moreover, Otto's exomemories are vulnerable to confiscation, corruption and deception by other agents. For these reasons, the access to externally stored information is neither as reliable nor as uncontaminated as access to internally stored information. The second dimension identified by Sterelny concerns the temporal development of the relationship between agents and external resources. Cognitive artefacts can in fact be classified along a continuum which goes from standardised and

[4] It is worth to underline that Sterelny's taxonomy is not the only attempt to understand the different modalities of cognitive integration between agents and external objects. Heersmink (2015), for instance, proposes a detailed multidimensional framework consisting of eight different dimension of integration. Like for Sterelny, also Heersmink conceives of these dimensions as a matters of degree which, taken together, delineate the space in which human cognition is coupled to external objects. However, I choose to rely on Sterelny's taxonomy due to a principle of coherency, as a significant part of my work is grounded upon his theoretical framework.

interchangeable resources to individualised and entrenched ones. The process of individualisation requires time, but, one stabilised, external cognitive resources are persistently used. As a consequence, according to Sterelny, individualised and entrenched cognitive resources «have value added by being both adapted by the agent and by the agent adapting to them» (Sterelny 2010, p. 476). Finally, while Clark tends to conceive of cognition and the use of cognitive tools mainly as a solitary activity, according to Sterelny, human problem-solving activity is usually social and it occurs through the use of communal, shared resources. The most relevant mind-and brain-shaping environmental supports for cognition are collectively provided tools for thinking and are transmitted across generations.

The model of scaffolded mind presented above has strong affinities with material engagement theory. Epistemic artefacts, according to Sterelny, do not merely take on the cognitive function previously performed by internal, 'mental', mechanisms, but rather they transform human cognitive processes and afford new ways for interacting with the world, for making decision, for solving problems. However, it is important to briefly sketch out an important difference in the use of the term "epistemic" between these two theories. On the one hand, Malafouris conceives of the epistemic qualities of things as *emerging from* the interaction with them, or to put it better, from the act of interacting with them. Things impose their own physical properties on human agency without the need of representations or symbolic meanings, but only in virtue of their material presence. Thus, as seen above, beads, being permanently attached to the body, made possible the emergence of forms of self-knowledge and self-recognition that cannot result from the interaction with and production of stone tools. To make another example, the physical persistence of things, their decaying and ageing, makes possible to appreciate and to understand the qualities of time. On the other hand, drawing from niche construction theory, Sterelny emphasizes human attitude in transforming features of the environment for supporting intelligent actions. Artefacts, for Sterelny, are epistemic because they are intentionally built to facilitate, improve or transform access to relevant information. In this sense, whereas for Malafouris the epistemic qualities of artefacts cannot be determined a priori, but rather become manifest only in the interaction between human and objects, Sterelny understands epistemic qualities mostly in the context of problem-solving and decision-making activities. This implicit and explicit epistemic agency of cognitive artefacts can perhaps echo the two dimensions of agency involved in the interactions between humans and their material environment identified by Aydin et al. (2018). As they write, first, the relation between humans and things can be conceptualized as a dance in which human and material agency are interwoven and help to shape each other, but, second, there is also a form of human agency resulting from this interaction (Aydin et al. 2018, p. 14). The next chapter will thus be dedicated to enquire how this two-fold agency of epistemic artefact might inform my discussion on persuasive technologies and self-awareness.

Chapter 3

3.1 Persuasive ambient mirrors and self-awareness

Objects shape, change and scaffold human cognitive processes and abilities, they make possible the emergence of new forms of self-awareness and support action by transforming the informational character of the environment. This perspective offers a different view on the impact of persuasive technologies on human life and behaviour than those discussed in the first chapter. As we saw, the critical target of ecological approaches to human mind and selfhood are disembodied, dualist models of cognition, but it might be argued that the behaviourist design implicit in most of persuasive technologies represents the other side of the 'modernist' coin. Persuasive technologies are in fact designed upon an input-output model in which a certain stimulus is supposed to be followed by a desired action, or set of actions, which will lead to the target, desired behaviour. Microsuasive elements, in particular, decompose the persuasive process in smaller steps which guide and shape users' behaviors through the deployment of different persuasive strategies. Thus, in the case of persuasive ambient mirrors, looking at a distorted face or seeing the sea-level increasing aims at triggering specific psychological responses which will lead the users to modify their harmful behaviours. Legitimately, designers of persuasive technologies are not interested in the broader impact of their systems upon the cognitive life of their users, but, if MET and Niche Construction Theory are right, in the use of persuasive systems there is more at stake than their behavioural outcomes, or, from a moral view, than their interference with users' freedom(s). As Clowes (2018, p. 11) points out:

> «without considering the environment, artefacts and cultures of usage, the skilful practices involved in appropriating and repurposing technology, we risk seriously misunderstanding the capacity for the emergence of new cognitive faculties. Agency comes into focus here as an aspect of self-actualization and mind-design. Human beings use material culture to bring new aspects of themselves into being»

How, then, persuasive technologies transform self-awareness? Or, to paraphrase Malafouris, what is the causal efficacy of persuasive devices on the users' cognitive and self-system? To begin with, it is worth to remember that human self and mind exist as historically situated contingencies shaped by their social, physical and technical environments. As we have seen, according to Malafouris, human autonoetic awareness could only emerge, phylogenetically, through a process of material engagement with material culture, but, importantly, this autonoetic self needs to be instantiated anew, generation after generation, in space and cultural time (Malafouris 2008b, p. 1996). Persuasive technologies, and persuasive ambient mirrors, can then be treated as 'pieces' of material culture through which new experience of the self, new modalities of self-recognition can emerge. Of course, a significant qualitative and quantitative gap exists between the cases discussed by Malafouris and persuasive technologies. The latter depends, I suggest, on the proliferation of objects used and encountered by contemporary humans in comparison to early hominins or Mycenean warriors. Malafouris, for instance, underlines that Mycenean self-objectification occurred throughout a variety of material media, so that warriors' awareness emerged

through the gold signet ring, but also through his sword and possibly few other artefacts. It seems plausible to assume, however, that in the vast majority of societies humans are now surrounded by and interact with a considerably greater number of objects. The 'contemporary' self, in other words, seems to be extended and distributed along a richer variety of things, implying that the transformations brought about persuasive technologies co-exist, and perhaps overlap and contrast, with those caused by other elements of material culture. More importantly, however, there is a qualitative leap between the kinds of artefacts discussed by Malafouris and the features of current technological systems (Aydin et al., 2018, p. 11). First, whereas MET focuses above all on the tangible materiality of things, current technological devices and environments operate in hidden ways, requiring no attention and falling outside the domain of everyday phenomenal experience. The lifestyle tracking component of persuasive mirrors captures precisely this point, as it records and processes user's data without the involvement and attention of the user. Second, technological environments are increasingly able to perform as 'actants', actively interacting with users in modalities that cannot be found in low-tech objects. Captology is a good example of this trend for two reasons. On the one hand, recalling Fogg's functional triad, persuasive technologies replace and potentially enhance the function of social actors, giving advice and developing social bonds and expectations. On the other hand, in contrast to the 'static' nature of Malafouris' examples, the *interactive* dimensions of persuasive technologies enable them to emotionally or intellectually engage the users over time and in personalized ways.

Having done these considerations, how do the epistemic qualities of persuasive ambient mirrors shape and transform self-awareness? First, due to their physical presence and properties, persuasive mirrors offer the material support for the emergence of new modalities of self-recognition and self-objectification. However, unlike body-ornament or rings, persuasive ambient mirrors are not incorporated into the bodily schema, they do not touch the body and they are not corporeal extension of it. The process of enchainment between internal and external elements of memory, the transition from a first person to third person perspective, occurs then in a different way. Helpful, in this regards, is to reflect on the metaphor of the mirror encapsulated in these devices, since the very act of looking into a traditional mirror already represents a moment of self-knowledge and objectification. Indeed, when scrutinizing our face or body in a mirror, we are immediately detached from our first person perspective and experience and we are forced to *face* ourselves, to become the object of our attention. Moreover, this close relation between mirrors and self-awareness seems to be confirmed by Gallup's famous experiments, where animals' ability to correctly interpret mirrored information about themselves is associated to the presence of self-conception and to a sense of psychological continuity over time (Gallup 1982). But while traditional mirrors reflect back to us a real-time image of our bodily movements and appearance, persuasive ambient mirrors provide a diachronic representation of our actions and behaviours. In this sense, if it is true, as suggested by Malafouris that objects enable episodic memory and they strengthen the recollection of past events connected to their use, then the kind of "dynamic cognitive biography" resulting from persuasive mirrors is radically new. By tracking and

recording users' activities, persuasive mirrors are in fact part of that E-memory revolution that has dramatically redraw the limits of memory (Clowes, 2015). Like Malafouris, Clowes does not conceive of E-Memory systems as "impacting" on the virgin territory of our minds, but rather as «entering into, and restructuring the pre-existing, and always developing, bio-technological complex of our minds» (Clowes 2015, p. 262). Among the different aspects of E-memory cognitive technologies, the most psychologically novel is, according to Clowes, totality, as E-memory technologies can potentially record everyday activities on a scale and with a fidelity and completeness previously unknown. While focusing only on one specific portion of daily life, such as household CO2 emission, eating or walking habits, persuasive ambient mirrors promise to offer an accurate and complete overview of that target behaviour. As digital repository of our behaviours, persuasive mirrors can consolidate and digitalize moments of our past that we did not remember and, also, those that we do not want to remember.

But the agency of persuasive ambient mirrors is not limited to the silent processing and visualization of users' data, but rather, by deploying feedback and prompts, it aims at shaping and transforming users' agency in more direct and intrusive ways. Drawing from cognitive niche construction then, the use of a persuasive mirrors can be thought as an epistemic action aiming at supporting and enhancing intelligent behaviour by transforming the informational character of the environment. More precisely, persuasive ambient mirrors can be defined as epistemic technologies designed for improving decision making by transforming cognitive problems into perceptual ones through the use of visual feedback that either invite or inhibit specific behaviours. At first glance, then, persuasive ambient mirrors can be treated as cognitive artefacts through which users can policy their behaviours, through which they can have access to new information about themselves and receive guidance and suggestions, but keeping open the possibility to decide whether or not following the cues appearing on the screen. The multidimensional space delineated by Sterelny will allow me to delve into the specific type of functional relationships established between persuasive mirrors and their users.

To begin with, persuasive ambient mirrors show an interesting tension between individual and collective resources. On the one hand, while it is true that persuasive technologies are relatively young epistemic artefacts, and thus it is not yet appropriate to treat them as the result of an intergenerational process, they can be conceived as communal resources. The sensors for tracking users' actions, the algorithms used to process them, the feedback mechanisms employed and their representational nature apply equally to all their users and are designed to support behaviour in the same way. Persuasive ambient mirrors, however, are not individually produced but they are individually *used*. The visual feedback is in fact highly personalized and directed to a specific person, and thus it has informational value *only* for that person, since the feedback refers to the progress toward the target behaviour of the user whose actions are tracked. In this respect, the most impressive feature of persuasive ambient mirrors is their potentiality for individualisation and entrancement. Over time, persuasive mirrors become progressively more adapted to their users, by supplying them with accumulated feedback that display information about past actions and makes projection about the future. Finally, to what extent persuasive

mirrors are trusted and reliable resources? While it is true, as Sterelny argues, that externally stored information is different and less reliable than internally stored ones, persuasive ambient mirrors introduce a novel aspect. The data and feedback shown on the mirror give in fact access to qualitatively new information about users' actions than the information which can be retrieved from the brain. In fact, the ambient, life-tracking component of persuasive ambient mirrors enable to keep track and quantify eating or walking habits in ways that overcome the possibilities of the 'naked' brain. While this kind of information is not necessarily more accurate or reliable than the information internally stored, it provides users with new ways for perceiving and reflecting about themselves.

Putting all this together, persuasive ambient mirrors seem to provide a highly personalized and highly scaffolded form of autonoetic awareness, one in which multiple modalities of self-recognition are possible. Persuasive ambient mirrors strengthen and enhance the ability of self-recollection because they allow users to remember, with unprecedented degrees of accuracy, their past actions and behaviours. On the other hand, they offer new possibilities for self-projection, because, through their feedback, they constantly remind the users about the consequences of their behaviours and help them to reach their goals. Unlike the rings or beads discussed by Malafouris, persuasive mirrors anchor our consciousness and enable the passage to a third person perspective not only because they are invested with memories and events associated to them, but because they literally "contain" some portion of our life. Moreover, the feedback loop configured by persuasive mirrors in forms of visual images is used to trigger specific reactions in the users in order to make them align with the target behaviour. To make an example, when, in *Ecoisland*, users are confronted with rising sea-level, not only they become aware of their ecologically harmful behaviours, but they are also pushed to behave differently in the future. Similarly, in the case of *Wize Mirror*, the textual messages appearing on the screen do not only make users aware of their health condition but they also provide them with recommendations about how to conduct a healthier lifestyle. What results is a hybrid agency in which the feedback appearing on the mirrors generate in the users emotions or cues that incentive them to act toward a certain direction.

What is, however, the price to be paid for this highly personalized and scaffolded form of self-awareness? At least three different but interrelated objections can be raised against the picture I have delineated above. First, while there is no doubt that persuasive mirrors "modify the informational character of the environment", one might be sceptical about the authentic epistemic value of the feedback received. Sterelny's definition is in fact quite loose, as it seems to include in the notion of "epistemic" any information which help humans to improve their decision-making or to solve high-load cognitive problems. But, as we saw, persuasive ambient mirrors deploy, in order to steer users, emotionally engaging feedback which can be grounded upon operant conditioning or mutual surveillance. The risk, then, is that by operating mainly at a psychological level, persuasive mirrors, rather than allowing the emergence of new cognitive abilities and modes of self-reflection, might end up configuring a 'conditioned' and 'standardized' self whose actions and decisions follow predetermined patterns. With the expression 'conditioned self' I refer to the feedback loop involved in

the systems which administer rewards or punishments in order to shape the form and frequency of the desired behaviours. The expression 'standardized self' refers to the fact that, as commercial products, this feedback loop applies to all users, whose action are then shaped according to the same persuasive mechanisms and leading, ideally, to the same behavioural outcomes. In this respect, the artifactual mind thesis (Aydin 2015) allows to fully appreciate the potential consequences of persuasive technologies on human thinking. Aydin radicalises extended approaches to cognition, which, he argues still presuppose a 'movement' from inside to outside and thus perpetuate the distinction between an internal cognitive realm and the external domain of objects and artefacts. On the contrary, according to Aydin, «our thinking is not pre-given or naturally present in a presumed inside world but it unfolds itself by virtue of and through objects and artefacts» (Aydin 2015, p. 89). The properties of objects induce certain thoughts and frame the world in certain way, as shown by the terms used to describe cognition, such as 'processing' and 'computation', which, far from being neutral, are borrowed from specific technologies and technological context. In the case of persuasive ambient mirrors, the visual feedback appearing on the screen might then come to constitute not only a reason for action in certain ways, but, more deeply, become the very concepts through which users conceive of themselves.

A second objection points to the differences between highly scaffolded choices on the one hand and cases of more weakly constrained individual cogitation on the other hand (Clark 1997, p. 182). As Clark points out, advanced cognition depends on the ability to *dissipate* reasoning, so that by intelligently structuring the environment, it becomes possible to operate in it with less intelligence. But, Clark continues, human behaviour appears to be more predictable in highly scaffolded situations where the role played by external resources is significant and there is few room for personal beliefs and desires. So, while it is true that persuasive ambient mirrors could lead to richer and sophisticated form of self-awareness, it might also have an homogenization effect, levelling the psychological and cognitive differences among people. In this regards, and this is the third objection, ecological approaches to cognition seem to neglect and disregard consciousness and the phenomenology of thinking (Walsh, 2017). According to Walsh, ecological frameworks of cognition underplay the role of epistemic virtues such as understanding and self-reliance and, in doing this, they tend to obscure how relevant information become *phenomenally conscious*. In the case of Otto and Inga, what really matters is not the functional equivalence between Inga's brain and Otto's notebook, but the fact that Otto's access to the information is perceptual, whereas Inga retrieves information from her brain-based memory. This difference is particularly evident in the case of failure, when Otto will attribute the source of his mistake to the notebook, while Inga will be disposed to attribute errors to herself. This phenomenological difference can be framed, according to Walsh (2017, pp. 45-46), through the concepts of confidence and trust: Inga trusts her faculties, while Otto is confident in his artefact. Analogously, *understanding* is not the same of *knowledge*, as the former implies an explanatory ability that is not necessary in the latter. In concluding, Walsh invokes a phenomenologically informed approach to the design and implementation

of cognitive artefacts, one which pursues not only functional integration, but phenomenal integration as well (Walsh 2017, p. 48).

So far my discussion has focused on a narrow, specific category of persuasive technologies, but my conclusions can, to a certain extent, be generalized to other kind of self-imposed persuasive technologies. Particularly, what I think is a structural element of these kind of persuasive devices is the dialectic relation established between narrative self-awareness and minimal self-awareness. As we saw, Malafouris is interested in the transition from a minimal, pre-reflective experience of the self to the ability of consciously reflecting about oneself. However, one might wonder to what extent this transition is actually ever completed, both from a phylogenetic and ontogenetic point of view. There is, in fact, an interesting parallel between these two dimensions of awareness and the distinction, discussed in the first chapter, between System 1 and System 2 and between Berlin's ideal self and empirical self. Advocates of dual system theory, in fact, would reply that human cognitive processes are still characterized by a fast, automatic and unconscious system, an "old mind" which evolved early and share many features with animal cognition. The "new mind", they would continue, is able to solve new problems by reasoning about the future and by enabling consequential decision making. Similarly, Berlin describes a divided man in which the dominant self, «the self which calculates and aims at what will satisfy it in the long run», is contrasted with irrational impulses, uncontrolled desire, the pursuit of immediate pleasure. (Berlin 1969, p. 132). As I noticed before (p. 16), adopting a self-imposed persuasive systems operate precisely upon this relation: self-imposed persuasive systems in fact presuppose a narrative, ideal self, able to plan about the future and aware of behaviours that need to be changed, and a minimal, impulsive and automatic self unable to resist to the stimuli coming from the external environment. In turn, the persuasive mechanisms deployed by persuasive technologies attempt to reverse this process by triggering specific behaviours and patterns of actions in the users. The dominant, narrative self willingly undergo a persuasive process in which the minimal self will be shaped and scaffolded by objects which induce, with more or less intensity, certain actions. This is the function of Fogg's microsuasive elements: to choose the right time for inviting to action, or to display the correct emotional feedback, in order to engage the users and generating specific chains of intentions and actions, thereby transforming the feeling of agency at the core of minimal self-awareness.

How this dialectic process is articulated depends on the different design and persuasive strategies used and, I think, can be well shown by the two alternative branches of persuasive technologies discussed in Chapter 1, mindless computing and reflective computing. The triggers deployed in mindless computing, in fact, produce *immediate and automatic* responses which affect users' behaviour without being noticed or perceived by them. In this way, the minimal self is deprived of his sense of agency and it is left only with the feeling of self-ownership. The user, in other words, is aware that it is her body which is acting in a certain way, but she does not control the source of that action. To follow Walsh's observation, in this case the coupled system user-persuasive technology has a purely functional dimension, as the persuasive process occurs below the threshold of user's consciousness. On the other

hand, the idea of reflective computing seems to better resonate with Walsh's view. According to Munson, in fact, supplying users' with relevant information about themselves might suffice for reflecting on themselves and becoming more conscious of their behaviors, with no need of persuasive feedback mechanisms or goals set by the system. In the case of reflective computing, then, the effect of the persuasive system on the minimal self is weak, due to the lack of microsuasive elements, whereas the persuasive process is directed to strengthen the narrative self, to make it thicker, by giving users' the possibilities to learn about their behaviours but without the introduction of normative elements. This is not, however, a neutral process: the type of information available to the users, the way this information is shown, like graphs, numbers or text, inevitably influence how users' represent themselves. As Munson recognizes, any digital environment in which people make choices is inherently persuasive, as it favours certain behaviours over others. In this respect, the possibility, for users, to customize and modify the persuasive systems they are using might be a crucial factor in determining their impact on self-awareness. Clowes (2015), for instance, speaks of *epistemic possession* when the cognitive technology used by an agent is cognitively penetrable, policeable and revisable by the agent. In concrete terms, being in the condition to decide what kind of information is stored, how it is manipulated and made available, enable users to become epistemically responsible of the consequences of persuasive devices on their cognitive systems. Moreover, if users can personalise their devices according to their own preferences or needs, the risk of standardization discussed above might be substantially reduced.

3.2 Extending the ecological niche: persuasive technologies and adaptive lag

My discussion of persuasive technologies has so far focused on individual experience. However, it might be wise to consider the interaction between the user and the persuasive system not only in itself, but also in relation to their function in the ecological niches in which they are designed and used. As I briefly noted (p. 23), Malafouris (2008c) suggests that the Mycenean sword or rings were not mere tools, but rather they had a crucial function in the construction of new personal and cultural identities. Through the sword, the Mycenean warrior was able to identify himself with his social context and to make sense of the world in which he was living. The sword, Malafouris argues (2008c, p. 120) was the material actualization of a new Mycenean *ethos* with a focus on the warrior's body. Following this line of reasoning, we might ask, with the eyes of a historian from the future, how the self-awareness emerging from the use of persuasive technologies is related to the broader ecological context in which these technologies are developed and used. To be more precise, how do the highly-scaffolded form of self-awareness enabled by persuasive technologies fit within the life and daily experience of their users?

I suggest that the conceptual tools of niche construction and, particularly, the discussion about human adaptive lag are useful lenses through which framing this point. As we saw, human modification of the environment can lead to negative consequences in terms of selective pressure, like environmental pollution or the spread of new diseases, but further cultural niche construction activities are usually developed in order to restore an adaptive balance. I suggest that (some) persuasive technologies, like

those for health and wellness or environmental sustainability, can be seen as episodes of niche construction activities aiming at alleviating environmental, psychological or physiological problems caused by post-industrial and contemporary society. On the one hand, the rapid and disruptive changes caused by technologies have produced niches which deeply alter the ecological balance of the Earth, or the physiological balance of the human organism. On the other hand, one might continue, the short time frame of these changes have made obsolete previous forms of education and social learning, or have revealed the need for new ones, as the members of previous generations did not have to face similar challenges. Persuasive technologies might be interpreted as an attempt to fill this gap, as a means deployed by people for coping with the rapid cultural and technological transformations which occurred in human societies within the last couple of decades.

An example of this process can be the relation between persuasive technologies for health and wellness and the nutrition transition that occurred in modern societies in the last decades of 20th century. Significant shifts in diet and physical activity patterns have in fact interested Western society, and more recently developing countries, leading to a diet high in saturated fats, sugar, and refined foods but also to lifestyles characterized by lower levels of activity (Popkin & Gordon-Larsen, 2004). Lakdawalla and Philipson (2002) studied the connection between long-run growth of weight in post-industrial societies and technological change, arguing that the latter has simultaneously raised the cost of physical activity and lowered the cost of calories. Whereas in agricultural and industrial societies work is physically tiring and food is expensive, post-industrial societies have reversed this relation: household and market work are now mainly sedentary, while agricultural innovation reduced food price but increased access to calories. Similarly, Egger and Boyd (1997) call for "an ecological approach to the obesity pandemic", which is defined as "normal physiology within a pathological environment". According to them, obesity should not be understood anymore as a personal disorder requiring treatment and depending on metabolic defects or genetic mutations, but rather as a normal response to an abnormal ecological conditions. Downey (2016) explicitly draws from niche construction theory to discuss the phenotypic bias of urban niches, arguing that the daily habits made possible by urban niches, such as the access to high-energy food or reduced physical activity, have produced deep metabolic and physiological changes which have led to the spread of obesity. In discussing the urban foraging habits of Brazilian *meninos de rua* (street children), Downey underlines their lower propensity for obesity in compared to other, more privileged children, although it is not clear whether this is the result of higher activity levels or lower energy intake is not clear.

A similar logic can be applied to persuasive technologies designed to improve the management of energy consumption and to increase environmental awareness. Since industrial and post-industrial societies have increased the amount of CO2 emission to levels unsustainable for the planet, the need for more ecologically sustainable lifestyles has become more urgent. It is drawing from this evidence that some of the persuasive mirrors discussed above, such as EcoIsland or the car dashboard displaying leaves according to driving behaviours, have been developed. Particularly significant is the case of a

Digital Mirror for shaping smart citizenship (Tonelli & Converso, 2014) which tracks all the data about the behaviour of the house and then displays them on a timeline. Through the timeline, users can check and compare their energy consumption in order to increase awareness about the consequence of their action. In this way technology can guide the users to take informed decisions and to improve their skills in managing energy consumption. Significantly, the authors highlight an intergenerational perspective on the function of these persuasive systems. As they state, the goal of the systems is to bring back to contemporary context the awareness and know-how of our grandmothers, who knew very well how to operate their homes, but, on the other hand, the goal is also to involve the native digitals through the gamification of housing behaviours.

A final interesting and recent example regards what has been defined as *smartphone addiction*. In the last decade smartphones have become integral part of everyday life, transforming communication, access to knowledge, working habits and individual entertainment. Resting in pockets, bags or on the desk, smartphones are always accessible and within hand-reaching distance, leading to a reconceptualization of personal space (Hatuka & Toch, 2016). However, due to their capacity of providing quick access to rewards like social networking, communications, and news, smartphones have the potentiality to generate new habits of Internet use, such as increasing the frequency with which emails, social networks updates or news headlines are checked (Oulasvirta, Rattenbury, Ma, & Raita, 2012). These *checking habits*, or checking behaviours, consist of relatively short but repeated sessions which have the effect of increasing phone use and might eventually degenerate into addictive behaviours that interfere with everyday activities. In particular, checking habits threaten and hinder the achievement of a permanent and deep level of attention and concentration (state of flow), with negative effect both on work productivity and on personal life (Duke & Montag, 2017). To help smartphone users to focus on their activities and to manage more responsibly the time spent on smartphone and internet-related activities, a number of persuasive applications reached the market in the last couple of years. I already mentioned ForestApp, which deploy the visual metaphor of a growing tree in order to incentive users to stay focused. Apple has recently introduced a ScreenTime modality which tracks and reports the time spent by the users on each application and other similar products have been developed for laptop and tablets, such as Rescue Time or StayFocus. Clowes (2018, p. 15), for instance, described the latter as a potential agentive enhancement which «has taken on the immediate role of making sure I comply with my own previous policy. It has become part of my own extended mechanisms of self-control». In this case the relation with smartphones and internet is perhaps particularly redundant. One the one hand, computers and phones have become strong cues for behaviours, requiring attention via their visual or acoustic signals and affording fast access to a vast array of informational rewards. On the other, they become persuasive agents to which self-control is delegated in order to remain on task and to better manage the time spent with them.

What these examples show, I suggest, is the relation between the fast pace of technological change, with its possible negative externalities, and the activities of niche construction aiming at

mitigating the impact of such externalities. The point is not, as evolutionary psychologists would say, that human physiology and psychology are not adapted to the technological landscape of contemporary societies, but rather that these novel ecological niches need to be integrated with likewise new ways for living in them. Persuasive technologies might represent a means for dealing with these challenges, but it is important to remember that the construction of tools and epistemic artefacts represent only one dimension of niche construction activities. Sterelny, as we have seen, accentuates the role of social learning in ensuring the transmission of knowledge and expertise across generations, a process, he argues, that does not occur only through formal education, but also in more indirect ways, such as observational learning. In this respect, as I have anticipated, it might be the case that contemporary urban and post-industrial environments present problems relatively unknown to previous generations and thus require the development of new habits which, in turn, will provide new forms of social learning. It would be naïve, however, to conceive of educations and persuasive technologies as mutually exclusive. On the one hand, as Sterelny notes, epistemic objects can be spread and improved horizontally along different individual or groups. This is particularly true in the case of persuasive technologies running on laptops or smartphone, which, based on software, are relatively easy to reproduce and can become available to massive amount of people in a short time span. On the other hand, in discussing the difference between nudging and education, Gigerenzer (2015) points out that the latter represents a more enduring and sustainable solution, since it would allow people to *understand* the risks connected to specific behavioural patterns.

Conclusion

The goal of this work was to enquire the effects of persuasive technologies on human cognition and self-awareness. By stressing the evolutionary and developmental dimension of the human mind, I tried to show that self-awareness, far from being a static property, is constantly shaped and restructured by material culture and cognitive artefacts. It is in the interaction with objects, in the engagement with their materiality, that humans can detach themselves from the flow of subjective experience and can objectify themselves. Enchained and extended into the material environment, the human self becomes conscious of its temporality. But the environment is not fixed either. To improve their decision making, to orientate themselves into the world, humans engineer their environment by making tool for thinking that deeply transform their cognitive abilities. Persuasive technologies are an example of this constructivist attitude, as they support users to reach their goals through different persuasive strategies. I discussed the case of persuasive ambient mirrors which, by tracking users' activities, become a digital repository of the past, confronting users' with their past actions and, potentially, projecting their current behaviour into the future. In turn, the feedback loop of persuasive mirrors leads to a form of hybrid agency where users' actions are triggered by emotionally engaging images aiming at encouraging or discouraging patterns of behaviours. For this reason, I argued that persuasive ambient mirrors enable the emergence of a rich, highly personalised and highly scaffolded form of self-awareness, as they strengthen the possibilities of self-recognition and self-objectification and guide users towards their goals. On the other hand, by acting primarily on the psychological traits of the users, some of the persuasive strategies used seem to have a relatively low epistemic value and might tend to delineate a conditioned and standardized self where the sense of self-agency is progressively eroded. Finally, I tried to go beyond a purely individualistic approach to persuasive technologies by trying to understand what might be their function in the contemporary niches inhabited by human beings. I framed my discussion through the concept of adaptive lag: the technological development occurred in the last decades has produced some negative consequences on human fitness, such as obesity, smartphone addiction or increased levels of CO2 emission. Due to the short frame of these changes, previous forms of social learning have become outdated and new ones need to be developed in order to face emerging novel challenges. Persuasive technologies might be a (provisional) means for filling this gap by making users more aware of their choices and of the ecological, psychological and physiological implications of their habits and behaviours.

Having summarized the main points of this work, I wish to briefly linger on a couple of conceptual tensions and blind spots that need to be considered in order to evaluate the achievement of my discussion. A first tension that has characterized this work is between two different processes of human becoming, phylogenesis and ontogenesis. The use of theoretical frameworks rooted in archaeology and philosophy of biology has in fact particularly stressed the evolutionary dimension of human cognition, and one might wonder to what extent this perspective can be fruitfully applied to

contemporary phenomena. After all, other theories, less indebted with human evolution, such as the different waves of Extended Mind Theory or post-phenomenology, could provide the philosophical tools for discussing the impact of persuasive technologies on human life and cognition. However, the merit of the evolutionary approaches discussed here is, I believe, to show the deep connection between those two levels. Human phylogenesis has been characterized by the progressive reliance and incorporation of external resources which, in turn, come to constitute and determine the developmental possibilities available for individuals: by changing their material and social environment humans change themselves and their cognitive abilities, creating new ones but also losing previous ones. In this way, Material Engagement Theory and Niche Construction Theory allow to historicize cognition and thus to better appreciate, to make more visible, the impact of material culture.

A second tension which has remained in the background of my work is between persuasion and persuasive technologies on the one hand and social learning and education on the other hand. What is the difference between these two processes for shaping human behaviours and why do persuasive technologies seem at odds with more classical educational methods? In this respect, I have argued that the highly scaffolded awareness provided by persuasive ambient mirrors might produce a homogenization effect, configuring users who behave and react in uniform ways. However, one might wonder whether and to what extent this process is different from classical education. An answer to this question cannot be given here. As a provisional answer, however, I suggest that education, or social learning, can be conceptualised as information that is extracted from the external environment, people or texts, but it is then internalized, becoming more reliable and always accessible and, perhaps, more open to personal interpretation. Another question which the present project did not investigate concerns the learning potentialities enabled by persuasive technologies. In this regards, it might be interesting, for future research, to explore the relationship between gamification, persuasive technologies and education. Gamification, namely the use of game design elements like points, levels or badges in non-game contexts, is in fact often criticized for being a powerful rhetoric tools used by marketers and industries to commercialize their products by taking the least essential aspects of games and presents them as the most relevant ones (Bogost 2012). However, an alternative view might be to consider gamification as the deployment of an *alternative pedagogic system* developed in gaming environments and then applied to non-game contexts (Tulloch 2014). As a consequence, according to Tulloch, gamification should be treated as a form of training and learning built upon the techniques inherited from games rather than on traditional pedagogy.

Finally, I wish to briefly come back to the ethical debate surrounding persuasive technologies. I stated in the introduction that this work might contribute to the debate by providing a preliminary anthropological and philosophical analysis of the impact of persuasive technologies on human cognitive systems. How is this translated in moral terms? Are self-imposed persuasive technologies a form of moral enhancement or, on the contrary, they replace and inhibit moral faculties? According to the result of my work, the answer depends, I suggest, on what is the kind of self-awareness at stake in this question.

Certainly, by operating upon the sense of agency of the minimal self and pushing users towards the morally right choice, persuasive technologies seem to negatively impact on the ability of moral deliberation. The conditioned self I delineated above might lead to a moral agent who makes the right choices but with small awareness of the reasons behind them. On the other hand, however, by providing new means for self-recognition and self-objectification, self-imposed persuasive technologies might give the users a more comprehensive perspective of the consequences of their actions, giving access to information that would not be available otherwise. In this way, moral reasoning and moral responsibility might be significantly enhanced. Another question remained unanswered is to what extent my conclusions can be applied to persuasive technologies imposed by third parties. Crucial, in this regards, might be whether or not individuals know that they are being subjected to a process of persuasion.

To conclude, an important blind spot and limitation of this work is the lack of an empirical and sociological analysis concerning the ways users appropriate persuasive technologies. To what extent users comply to the feedback received? How do users conceive of themselves after prolonged interaction with a persuasive system? In addition to this, an in-depth sociological analysis of the origin and development of persuasive technologies might also validate or dismiss the hypothesis I proposed in the last section of this thesis.

Acknowledgments

This final project is not the result of a solitary effort, but it benefitted from the help and support of various people. I wish to thank my first supervisor, Prof. Aydin, who taught me the importance of *thinking* before writing and for his helpful and insightful feedback that helped me to give structure to this project. I am grateful to my second examiner, Dr. Michael Nagenborg, for his acute comments. His glosses to a previous version of this text were very precious. A special thanks to Margoth, for our discussions about Niche Construction and Material Engagement Theory and for her constant moral support. Also, I am very grateful to the PSTS staff, for the hospitable and highly stimulating intellectual environment in which I had the possibility to grow in these years. A special thought goes to Prof. Lantz Miller and Prof. Lissa Roberts, who helped me to sharpen my writing skills and critical thinking.
Finally, I wish to thank my parents for the constant, tireless support and encouragement I have received throughout these years of study.

References

Aydin, C., González Woge, M., & Verbeek, P.-P. (2018). Technological Environmentality: Conceptualizing Technology as a Mediating Milieu. *Philosophy & Technology*, 1–18. https://doi.org/10.1007/s13347-018-0309-3

Aydin, C. (2015). The artifactual mind: overcoming the 'Inside-Outside' dualism in the extended mind thesis and recognizing the technological dimension of cognition. *Phenomenology and the Cognitive Sciences*, 14(1), 73–94.

Bateson, G. (1972). *Steps to an ecology of mind*. New York: Balentine Books.

Berdichevsky, D., & Neuenschwander, E. (1999). *Toward an ethics of persuasive technology*. Communications of the ACM, 42(5), 51–58. https://doi.org/10.1145/301353.301410

Berlin, I. (1969) Two Concepts of Liberty. In *Four essays on liberty*. Oxford, Oxford University Press

Bertolotti, T., & Magnani, L. (2017). Theoretical considerations on cognitive niche construction. *Synthese*, *194*(12), 4757–4779. https://doi.org/10.1007/s11229-016-1165-2

Bogost, I. (2011). *Persuasive games: exploitationware*. Gamasutra. Retrieved from http://www.gamasutra.com/view/feature/134735/persuasive_games_exploitation ware

Brey, P. (2005). Freedom and privacy in ambient intelligence. *Ethics and Information Technology*, 7(3), 157–166. https://doi.org/10.1007/s10676-006-0005-3

Clark, A.: 1997, *Being There: Putting Brain, Body and World Together Again*, MIT Press, Cambridge, MA.

Clark, A. (2005a). Word, niche and super-niche: How language makes minds matter more. *Theoria-Revista De Teoria Historia Y Fundamentos De La Ciencia*, *20*(3), 255–268. https://doi.org/10.1387/theoria.561

Clark, A. (2006). Language , embodiment , and the cognitive niche, *10*(8), 6–10. https://doi.org/10.1016/j.tics.2006.06.012

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7–19.

Clowes, R. (2015). Thinking in the Cloud : The Cognitive Incorporation of Cloud-Based Technology, *Philosophy & Technology*, 28:261–296. https://doi.org/10.1007/s13347-014-0153-z

Clowes, R. W. (2018a). Immaterial engagement: human agency and the cognitive ecology of the internet. *Phenomenology and the Cognitive Sciences*, 1–21. https://doi.org/10.1007/s11097-018-9560-4

Clowes, R. W. (2018b). Screen reading and the creation of new cognitive ecologies. *AI and Society*, *0*(0), 1–16. https://doi.org/10.1007/s00146-017-0785-5

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). *From game design elements to gamefulness*. Proceedings of the 15th International Academic MindTrek Conference on Envisioning Future Media Environments - MindTrek '11, 9. https://doi.org/10.1145/2181037.2181040

Downey, G. (2016). Being Human in Cities: Phenotypic Bias from Urban Niche Construction. *Current Anthropology*, *57*(13), 52–64. https://doi.org/10.1086/685710

Duke, É., & Montag, C. (2017). Smartphone addiction, daily interruptions and self-reported productivity. *Addictive Behaviors Reports*, *6*(April), 90–95. https://doi.org/10.1016/j.abrep.2017.07.002

Egger, G., & Boyd, S. (1997). An “ecological” approach to the obesity pandemic. *SPE Advanced Technology Series*, *315*(7106), 477–490. https://doi.org/10.2118/21677-pa

Est, R. Van, Timmer, J., Kool, L., Nijsingh, N., & Rerimassie, V. (2016). *Rules for the digital human park Two paradigmatic cases of breeding and taming human beings : Human germline editing and persuasive technology*. Retrieved from https://www.rathenau.nl/sites/default/files/2018-09/Est et al. -

Evans, J. S. B. T. (2014). Two minds rationality. *Thinking and Reasoning*, 20(2), 129–146. https://doi.org/10.1080/13546783.2013.845605

Fogg. B.J. (2003). *Persuasive Technology Using Computers to Change WhatWe Think and Do*. Morgan Kaufmann Publisher

Fogg, B. J., & Iizawa, D. (2008). *Online persuasion in Facebook and Mixi*. Proceedings of the 3rd International Conference on Persuasive Technology, 35–46. https://doi.org/10.1007/978-3-540-68504-3-4

Fuentes, A. (2015). Integrative Anthropology and the Human Niche: Toward a Contemporary Approach to Human Evolution, *American Anthropologist*, 117(2), 302–315. https://doi.org/10.1111/aman.12248

Gallagher, S. (2010). “Ambiguity in the sense of agency,” in Decomposing the Will, eds J. Kiverstein and T. Vierkant (Oxford: Oxford University Press), 1–17.

Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. Trends Cogn. Sci. 4, 14–21. doi: 10.1016/S1364-6613(99)01417-5

Gallup, G. G. (1982). Self-awareness and the emergence of mind in primates. *American Journal of Primatology*, *2*(3), 237–248. https://doi.org/10.1002/ajp.1350020302

Gibson, J. J. (1986). *The ecological approach to visual perception*. Hillsdale, NJ: Lawrence Erlbaum Associates

Gigerenzer, G. (2015). On the Supposed Evidence for Libertarian Paternalism. *Review of Philosophy and Psychology*, *6*(3), 361–383. https://doi.org/10.1007/s13164-015-0248-1

Guthrie, C. F. (2013). Smart Technology and the Moral Life. *Ethics and Behavior*, 23(4), 324–337. https://doi.org/10.1080/10508422.2013.787359

Hatuka, T., & Toch, E. (2016). The emergence of portable private-personal territory: Smartphones, social conduct and public spaces. *Urban Studies*, *53*(10), 2192–2208. https://doi.org/10.1177/0042098014524608

Heersmink, R. (2015). Dimensions of integration in embedded and extended cognitive systems. *Phenomenology and the Cognitive Sciences*, 14(3), 577–598.

Hutchins, E. (2010). Cognitive Ecology. *Topics in Cognitive Science*, 2, 705–715

Jeffares, B. (2010). The co-evolution of tools and minds : cognition and material culture in the hominin lineage, *Phenomenology and Cognitive Science*, 9,503–520. https://doi.org/10.1007/s11097-010-9176-9

Kimura, H., & Tatsuo Nakajima. (2010). *EcoIsland: a persuasive application to motivate sustainable behavior in collectivist cultures*. NordiCHI ’10 Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries, 703–706. https://doi.org/10.1145/1868914.1869009

Knowles, B., Coulton, P., Lochrie, M., & Wohl, B. (2014). *“ Convince Us ”: An Argument for the Morality of Persuasion*. Workshop on Co-Creating & Identity-Making in CSCW. Retrieved from http://ethicsworkshopcscw2014.files.wordpress.com/2013/10/knowles-convince-us.pdf

Lakdawalla and Tomas J. Philipson, D. N. (2002). Technological Change and the Growth of Obesity. *NBER Working Paper*, *Cambridge*,(2), 253–257.

Laland, K. & Brown, G. (2006). Niche Construction, Human Behavior, and the Adaptive-Lag Hypothesis. *Evolutionary Anthropology* 15, 95–104

Laland, K., & Kendal, J. (2007). The niche construction perspective : Implications for evolution and human behaviour. J*ournal of Evolutionary Psychology*, 5, 51-66 https://doi.org/10.1556/JEP.2007.1003

Laland, K., Matthews, B., & Feldman, M. W. (2016). An introduction to niche construction theory. *Evolutionary Ecology*, *30*(2), 191–202. https://doi.org/10.1007/s10682-016-9821-z

Lee, U., Han, K., Cho, H., Chung, K. M., Hong, H., Lee, S. J., Carroll, J. M. (2019). *Intelligent positive computing with mobile, wearable, and IoT devices: Literature review and research directions*. Ad Hoc Networks, 83, 8–24. https://doi.org/10.1016/j.adhoc.2018.08.021

Magnani, L., & Bardone, E. (2008). Sharing Representations and Creating Chances through Cognitive Niche Construction . The Role of Affordances and Abduction, S*tudies in Computational Intelligence* (SCI) 123, 3–40, 3–40.

Malafouris, L. (2004). The Cognitive Basis of Material Engagement: Where Brain, Body and Culture Conflate. *Rethinking Materiality: The Engagement of Mind with the Material World*, 53–62.

Malafouris, L. (2008a). Beads for a Plastic Mind: the 'Blind Man's Stick' (BMS) Hypothesis and the Active Nature of Material Culture Lambros. *Cambridge Archaeological Journal*, *18*(3), 401–414. https://doi.org/10.1016/j.cub.2015.10.018

Malafouris, L. (2008b). Between brains, bodies and things: Tectonoetic awareness and the extended self. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1499), 1993–2002. https://doi.org/10.1098/rstb.2008.0014

Malafouris, L. (2008c). Is it 'me' or is it 'mine'? The Mycenaean sword as a bodypart. In J. Robb & D. Boric (eds) *Past Bodies*. Oxford: Oxbow Books, 115-23.

Malafouris, L. (2013). *How things shape the mind. A theory of Material Engagement*. Massachusetts Institute of Technology

Maturana, H., & Varela, F. (1987). *The tree of knowledge: Biological roots of human understanding*. Boston: New Science Library

Munson, S. (2012). *Persuasion in Personal Informatics*. Workshop on Personal Informatics, 1–4.

Nagenborg, M. (2014). Surveillance and persuasion. *Ethics and Information Technology*, 16(1), 43–49. https://doi.org/10.1007/s10676-014-9339-4

Nakajima, T., & Lehdonvirta, V. (2013). *Designing motivation using persuasive ambient mirrors. Personal and Ubiquitous Computing*, 17(1), 107–126. https://doi.org/10.1007/s00779-011-0469-y

Nakajima, T., Lehdonvirta, V., Tokunaga, E., & Kimura, H. (2008). *Reflecting human behavior to motivate desirable lifestyle*. Proceedings of the 7th ACM Conference on Designing Interactive Systems - DIS '08, 405–414. https://doi.org/10.1145/1394445.1394489

Oulasvirta, A., Rattenbury, T., Ma, L., & Raita, E. (2012). Habits make smartphone use more pervasive. *Personal and Ubiquitous Computing*, *16*(1), 105–114. https://doi.org/10.1007/s00779-011-0412-2

Pocheville, A. (2015). *The Ecological Niche : History*. In T. Heams et al. (eds.), Handbook of Evolutionary Thinking in the Sciences, https://doi.org/10.1007/978-94-017-9014-7

Popkin, B. M., & Gordon-Larsen, P. (2004). The nutrition transition: Worldwide obesity dynamics and their determinants. *International Journal of Obesity*, *28*, S2–S9. https://doi.org/10.1038/sj.ijo.0802804

Poulsgaard, K. S. (2017). Enactive individuation : technics , temporality and affect in digital design and fabrication.

Rughiniş, C., Rughiniş, R., & Matei, Ş. (2015). A touching app voice thinking about ethics of persuasive technology through an analysis of mobile smoking-cessation apps. *Ethics and Information Technology*, 17(4), 1–15. https://doi.org/10.1007/s10676-016-9385-1

Shiraishi, M., Washio, Y., Takayama, C., Lehdonvirta, V., Kimura, H., & Nakajima, T. (2009). *Using individual, social and economic persuasion techniques to reduce CO 2 emissions in a family setting*. Proceedings of the 4th International Conference on Persuasive Technology - Persuasive '09, 1. https://doi.org/10.1145/1541948.1541967

Sinha, C. (2015). Ontogenesis, semiosis and the epigenetic dynamics of biocultural niche construction. *Cognitive development*, 36, 202–209

Spahn, A. (2012). And Lead Us (Not) into Persuasion...? Persuasive Technology and the Ethics of Communication. *Science and Engineering Ethics*, 18(4), 633–650. https://doi.org/10.1007/s11948-011-9278-y

Sterelny, K. (2010). Minds: extended or scaffolded ?, *Phenomenology & Cognitive Science*, 9, 465–481. https://doi.org/10.1007/s11097-010-9174-y

Sterelny, K. (2004). Externalism, epistemic artefacts and the extended mind. In R. Schantz (Ed.), *The externalist challenge. New studies on cognition and intentionality* (pp. 239–254). Berlin: de Gruyter.

Sterelny, K. (2006). Cognitive load and human decision, or, three ways of rolling the rock up hill. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind: Volume 2: culture and cognition* (pp. 218–233). Oxford: Oxford University Press.

Sterelny, K. (2007). Social intelligence, human intelligence and niche construction. *Proceedings of the Royal Society London* (series B), 362(1480), 719–730.

Sterelny, K. (2012). *The Evolved Apprentice. How Evolution Made Humans Unique*, The MIT Press Cambridge, Massachusetts

Tonelli, C., & Converso, S. (2014). Digital mirror: A method to shape smart citizenship. *Energy and Buildings*, *83*, 173–180. https://doi.org/10.1016/j.enbuild.2014.04.032

Tulloch, R. (2014). Reconceptualising gamification: Play and pedagogy. Digital Culture and Education 6 (4), 317-333.

Varela, F., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. Cambridge, MA: MIT Press.

Verbeek, P.-P. (2006). Persuasive Technology and Moral Responsibility Toward an ethical framework for persuasive technologies. *Persuasive*, 1–15. https://doi.org/10.1111/j.1600-0668.2005.00303.x

Verbeek, P. P. (2009). Ambient intelligence and persuasive technology: The blurring boundaries between human and technology. *NanoEthics*, 3(3), 231–242. https://doi.org/10.1007/s11569-009-0077-8

Wallach, E. (2016). Niche construction theory as an explanatory framework for human phenomena. *Synthese*, *193*(8), 2595–2618. https://doi.org/10.1007/s11229-015-0868-0

Walsh, P. J. (2017). Cognitive extension , enhancement, and the phenomenology of thinking, *Phenomenology and the Cognitive Science*, 16, 33–51. https://doi.org/10.1007/s11097-016-9461-