Freedom and Persuasion in the Attention Economy

James Wilson Williams
Balliol College
University of Oxford

Thesis submitted in partial fulfillment of the requirements for the degree of
D.Phil. in Information, Communication, and the Social Sciences
at the Oxford Internet Institute at the University of Oxford

Trinity Term 2017

53,175 words

James Wilson Williams
Balliol College, Oxford
Thesis Title: 'Freedom and Persuasion in the Attention Economy'
Degree: D.Phil. in Information, Communication, and the Social Sciences
Submitted in Trinity Term 2017

Abstract

In order to do anything that matters, we must first be able to give attention to what matters. However, as Herbert Simon predicted in the 1970s, the unprecedented abundance of information produced by digital technologies has resulted in a similarly unprecedented scarcity of attention. As the newly scarce resource, our attention is today the primary object of competition among our digital technologies. In the so-called 'attention economy' that has emerged, design wins when it captures as much of our attention as possible—and in order to win, it must increasingly exploit our non-rational psychological biases. This attentional capture is carried out with a view to directing our thoughts and behaviors toward predefined goals—goals that may or may not align with our own. Persuasion, then, is the dominant business model of the digital attention economy, and this persuasion is increasingly powerful, prevalent, and centralized. Major moral and political questions therefore lurk here: questions of *attention* mangagement that to date have been largely passed over in favor of questions about *information* management, such as privacy or surveillance. As a result, a full ethical grappling with the digital attention economy is long overdue. But how should we begin to think about, let alone protect, human freedom in such an environment of industrialized persuasion? In this thesis I approach this topic via four main questions. First, what *counts* as a 'persuasive' technology? Second, what risks to freedom do persuasive technologies pose? Third, how does advertising ethics need to change in order to be useful in the digital attention economy? And finally, how does specifically *digital* advertising differ in its nature and ethics from advertising as historically understood? I broadly conclude that the digital attention economy has produced several direct as well as systemic effects that pose serious challenges for user self-determination. These challenges have received little serious ethical attention to date and therefore warrant a sustained project of critical analysis and intervention. I then close with a brief, high-level discussion of possible paths for reforming the digital attention economy moving forward.

Acknowledgements

Deepest thanks are owed to my wife Julianne and to my parents, Drs. Don and Lynne Williams, for many years of loving support, encouragement, and inspiration.

I am grateful to Professor Luciano Floridi, my supervisor, for his tutelage and wisdom over the course of my research. I am also indebted to Dr. Victoria Nash, my secondary supervisor, whose support and guidance has been pivotal. This thesis was also made stronger through discussion and debate with many other brilliant minds at the Oxford Internet Institute; I particularly wish to acknowledge the contributions of Joshua Melville, Professor William Dutton, Professor Ralph Schroeder, Professor Vili Lehdonvirta, Professor Eric Meyer, Dr. Heather Ford, the OII DPhil seminars, and the Digital Ethics Lab. Thanks are also due to the OII PhilTech seminars, Dr. Elizabeth Dubois, Bendert Zevenbergen, and Dr. Brent Mittelstadt. I am also appreciative of the teaching opportunities afforded me during my D.Phil. by Dr. Ian Brown, Dr. Joss Wright, Dr. Mariarosaria Taddeo, and the Oxford Computer Science department. *Docendo, discimus*.

This thesis has particularly benefited from my ongoing affiliation with the Uehiro Centre for Practical Ethics in the Oxford Philosophy department. I am grateful to Professor Julian Savulescu, director of that group, for his gracious support and and hospitality. For particularly meaningful discussions about my work I wish to acknowledge Professor Roger Crisp, Professor Janet Radcliffe-Richards, Dr. Regina Rini, Professor Jeff McMahan and the Oxford Moral Philosophy seminars, Dr. Anders Sandberg, Dr. Carissa Veliz, the late Professor Derek Parfit, the Ockham Society seminar, and the Applied Ethics Work-in-Progress seminar. I have also very much valued my interactions with the annual Oxford-Bucharest Applied Ethics Workshop, in particular my conversations with Dr. Constantin Vică and Dr. Emilian Mihailov.

I also thank several friends and mentors at Balliol College for their support: Achas Burin, Thomas Møller-Nielsen, Nicola Trott, the 2012 Chalet reading party, Professor Adam Swift, and the Balliol Interdisciplinary Institute. Additionally, for their gastronomic adaptability as well as their long-term reliability in the provisioning of egg wraps, I warmly thank our neighbors at Jimbob's Baguettes, who have, quite literally, fueled the majority of the thinking contained herein.

Valuable conversations with many others, both during and prior to my research, have elevated the quality of this thesis. I cannot list them all, but I particularly wish to acknowledge the input, feedback, and support of: Benson Dastrup; Tristan Harris, Joe Edelman, Max Stossel, and other members of the the Time Well Spent community; Dr. Vint Cerf; Thich Nhat Hanh; Wael Ghonim; Professor Tim Wu; Georgi Kantchev; Professor David Runciman; Ernesto Oyarbide; Viviane Eide; Dr. Andrei Duta; Phil Bolton; and my brothers, Thomas, John, and Mark Williams.

Finally, I wish to thank my friends and colleagues at Google, with whom I spent over ten years of my life, and who encouraged and accommodated me as I set out on this project of research. That koan-like maxim which still serves as Google's guiding light for design—'Focus on the user and all else will follow'—also motivates the spirit of this work. If my efforts here enhance the capacity for that focus, even just a little, then I will consider them to have been a success.

Contents

Introduction: Technologies of Attention **5**
Prologue: The Faulty GPS 5
I. The Age of Attention 6
II. Persuasion and Freedom 12
III. Persuasive Technology 19
IV. The Synthetic Challenge: Aims and Methods 24

Chapter 1: What Is a 'Persuasive' Technology? **33**
I. The Emergence of 'Persuasive' Technology 33
II. Current Definitional Challenges 36
III. Toward a More Intentional Definition 44
IV. A Search for 'Defensible Metaphors' 52
V. Conclusion 56

Chapter 2: Autonomy & Persuasive Technology: Reactions & Distractions **58**
I. Common Autonomy-Based Criticisms of Persuasive Technology 42
II. 'Reactions': Separating Real from Apparent Autonomy Criticisms 62
III. Attention as a Common Framework for Autonomy and Dignity 72
IV. Three Types of 'Distraction' 79
V. Conclusion 86

Chapter 3: Reclaiming Advertising Ethics **90**
I. The Virtual Field of Advertising Ethics 90
II. Marginal Ethics on Marginal Time 94
III. The Agent-Orientation of Advertising Ethics 98
IV. The Mercurial Definitions of 'Advertising' 101
V. Attentional Blindness: Advertising Ethics and the Simonian Inversion 106
VI. The Urgency of Clarifying Digital Advertising 108
VII. Conclusion 110

Chapter 4: The Moral Character of Digital Advertising **111**
I. Digital Advertising: A Tectonic Shift 111
II. Differences of *Boundedness* 114
III. Differences of *Intelligence* 119
IV. From 'Underwriting' to 'Overwriting' 129
V. Some Ethical Implications 134
VI. Digital Advertising and Self-determination: Challenges, Opportunities, Obligations 137
VII. Conclusion 147

Conclusion: Stand Out of Our Light **148**
I. Diogenes and Alexander 148
II. Asserting and Defending the Freedom of Attention 152
III. The Benefit of Competence 157
IV. The Brightest Heaven of Invention 159

Bibliography **161**

'The empires of the future are the empires of the mind.'
– Churchill

Introduction: Technologies of Attention

Prologue: The Faulty GPS

Imagine that you have just bought a new GPS device for your car. The first time you use it, it works as expected. On the second trip, however, it takes you to an address a few blocks away from where you had wanted to go. On the third trip, you are shocked when you find yourself *miles* away from your intended destination, which is now on the opposite side of town. Frustrated, you decide to just return home—but when you enter your address, the GPS gives you a route that would have you drive for *hours* and end up in a totally different city. Like any reasonable person, you would consider this GPS faulty and return it to the store—if not chuck it out the window of your car. Who would continue to put up with a GPS they *knew* would take them somewhere other than where they wanted to go? What reasons could anyone possibly have for continuing to tolerate such a thing?

No one would put up with this sort of distraction from a technology that directs him or her through *physical* space. Yet we do precisely this, on a daily basis, when it comes to the technologies that direct us through *informational* space. We have a curiously high tolerance for poor navigability when it comes to our informational GPSes—those technologies that direct our thought, our actions, and our lives.

This tolerance is all the more astonishing when we consider the increasing persuasiveness of the technologies in our informational environment. As we spend ever greater amounts of time with our information and communication technologies, they are being increasingly designed to exploit our psychological biases in order to move us toward goals that may or may not align with our own.

In the short term, this can distract us from doing the things we want to do. In the longer term, it can distract us from living the lives we want to live, or, even worse, undermine our capacities for reflection and self-regulation, making it harder, in the words of Harry Frankfurt (1988), to 'want what we want to want.' As a result, deep ethical implications lurk here for human freedom, wellbeing, and even the integrity of the self.

The motivating question of this thesis is how—and indeed whether—individuals can remain autonomous in a media environment that increasingly shapes their attention and action in ways that produce conformity with predefined goals. My goal here is not to answer this question, but rather to lay essential groundwork for it: to carry out a multileveled, interdisciplinary analysis of persuasive technology that brings new and ethically salient issues to light, and contributes to the development of conceptual tools that future research may draw on as it grapples with this question.

I. The Age of Attention

In every age, the dominant technologies inevitably become dominant metaphors for the human condition. Man is a book, a ship, a machine. In our time, the computer is ascendant. It is through these metaphors that the myths of an age, and the perceived impediments to flourishing that they exist to help us navigate, come into full view. As computers, the impediments with which we grapple are naturally informational in nature—so much so that we refer to our era as the 'Information Age.' Dazed by the sheer volume of information, paralyzed by its variety, and knocked off our feet by its velocity[1], we take as our solemn mission the task of taming this new digital wilderness, and of restoring to the human collective our rightful dominance over the informational world.

[1] cf. IBM (2015), "The Four V's of Big Data"

Yet perhaps our computer's-eye view has led us to grapple with many of the wrong impediments. Perhaps we have even misjudged their fundamental nature. In the 1970's, Herbert Simon pointed out that in an environment of information abundance, a figure/ground reversal takes place which renders *attention* the scarce resource (Simon 1971). We are currently living through the pendulum swing of that reversal. True, we have more information than we know what to do with—discussion of 'big data' seems to be everywhere—but all the world's research still uses only 1% of the *one quintillion* bytes of data available (Wall 2014). As Luciano Floridi (2012) has pointed out, the true epistemological challenge of 'big data' is not the bigness of the data but rather the 'small patterns' we bring to the table when we try to deal with it—in other words, the smallness of our attentional capabilities. Thus the proper response to information abundance cannot lie in valuing everything we measure—an impossible task—but rather in making sure we measure everything we value.

As the scarce resource in the Information Age, attention is now the object of global competition among our technologies. Since Simon, digital technologies have come to form the background of our lived experience in an astonishing way. Networked computing devices have become more portable, more ubiquitous, and more natural in the way they interact with us. Furthermore, 'traditional' media such as television and radio have largely been digitally retrofitted, rendering the networked digital environment a constant presence in human life. Within this environment, attention is now the primary object of value, and thus of competition, among our technologies. The so-called 'attention economy' that has resulted (Lanham 2006) now functions as the default business model of the Internet. In fact, some 'attention economists want to say that attention is changing the fundamental rules of operation for *all* markets by making attention the primary measure of exchange value' (Rogers 2014, emphasis original).

The attentional impediments to flourishing that the attention economy poses are fundamentally challenges of self-regulation. In a sense, the psychological and behavioral adjustments that are necessary to help us adequately respond to these impediments require us to invert our usual understanding of what we mean by 'technology': rather than a method of overcoming barriers in the world, 'technology' increasingly refers to the way we put barriers in place. 'Silence is now offered as a luxury good' (Crawford 2015). These challenges of self-regulation also require that we take a more environmental approach to technology: the felt externalities of human life and behavior that have resulted from the attention economy have prompted calls for a 'grey ecology' (Virilio 2009, Broadbent & Lobet-Maris 2014); the magazine *AdBusters* describes itself as a 'journal of the mental environment.'

An instructive comparison with this environment of informational abundance is that of *economic* abundance, where similar challenges of self-regulation occur. Like informational abundance, economic abundance (i.e. 'affluence') 'may be characterized as a flow of new and inexpensive rewards. If these rewards arrive faster than the disciplines of prudence can form, then self-control will *decline* with affluence: the affluent (with everyone else) will become less prudent' (Offer 2006, emphasis original). This scenario should be familiar enough to anyone who has read the all-too-familiar stories about lottery jackpot winners who, though initially elated, soon find they have squandered their wealth—and often regret even winning it in the first place. In a sense, the mistake these individuals make upon winning the jackpot is to construe the primary question before them— 'What should I do with all this money?'—as a question of *money* management rather than one of *attention* management or self-regulation. They are turned outward at precisely the moment when they ought to be turning inward. In the same way, the dominant question we seem to have been asking ourselves thus far in the Information Age—'What should we do with all this information?'—has been framed as one of *information* management rather than *attention*

management. And, as in the case of the lottery winners, the rewards of our information seem to be arriving 'faster than the disciplines of prudence can form.'

There are four interrelated trends that amplify these challenges of self-regulation, and bring urgency to a full analysis of their dynamics:

1. First is the collapse—or jettisoning—of previous cultural 'commitment devices' on which we have historically relied to support our pursuit of desired values. By enabling the rejection of the 'package deal' of religion on the basis of philosophical or cosmological disagreements, secularism made it easy to reject the habits, practices, and rituals that evolved over centuries to reliably guide—or shall we say *design*—our lives in the direction of particular values. 'The left's project of liberation,' writes Crawford (2005), 'led us to dismantle inherited cultural jigs that once imposed a certain coherence (for better and worse) on individual lives. This created a vacuum of cultural authority that has been filled, opportunistically, with attentional landscapes that get installed by whatever "choice architect" brings the most energy to the task—usually because it sees the profit potential.' (Crawford 2015, emphasis original). Peter Sloterdijk (2013) has called for a reclamation of this particular aspect of religion, which he calls 'anthropotechnics.' In his book *You Must Change Your Life*, he writes, 'It is time to reveal humans as the beings who result from repetition. Just as the nineteenth century stood cognitively under the sign of production and the twentieth under that of reflexivity, the future should present itself under the sign of the exercise' (Sloterdijk 2013). Yet in the absence of effective commitment devices we are left to our *own* devices, and given the inherent scarcity of attention the resulting cognitive overload renders effective pursuit of such 'exercise' extremely challenging, if not prohibitive.

2. The second trend is the increasing tendency to view human life primarily as a design or optimization challenge. While the felt need for this perspective may originate in the loss of stable commitment devices, the capability that makes it really convenient is the wide (we might even say indiscriminate) repurposing of the language and concepts of software engineering across all domains of life. In this view, 'reality is broken' (McGonigal 2011), but of course it can be fixed by 'smart' technologies or sufficiently 'gamified' systems (Lobo et al. 2009). At the macro level, when this perspective intersects with the massive digital infrastructure that has emerged for measuring human behaviors, attitudes, and contexts, it results in the practice of analytics, which in turn enables faster and tighter feedback loops in the design and optimization of technologies. (In fact, when using an online service or platform, there may be so many experiments taking place at a given time that a user literally may never use the same product twice.) Similarly, at the micro level, consider the 'quantified self' movement and the quest for 'happiness by design' (Dolan 2014). In all cases, the dominant value at which this perspective aims is that of *efficiency*—though this value is rarely explicit, and is even more rarely given justification.
3. Third, advances in psychology and behavioral economics have continued to reveal new insights about the importance of non-rational processes in human cognition. This knowledge has rapidly been applied in the design of many technologies that compete for user attention. It was in the early- and mid-twentieth century that psychologists such as Freud and Skinner laid the early groundwork for the study of unconscious thought. Later, the first shots in the so-called 'rationality wars' were arguably fired by Kahneman and Tversky (1973) when they revealed the ways that people's intuitive heuristics can override more rational rules of statistical prediction (Samuels et al. 2002). Subsequently, the picture of non-rational human thought was further filled in as researchers identified more heuristics and biases that can broadly be classified into three areas: 1) the 'automatic effect

of perception on action,' 2) 'automatic goal pursuit,' and 3) the 'continual automatic evaluation of one's experience.' Bargh and Chartrand (1999) proposed that such automatic, non-conscious processes in fact form the majority of our experience of everyday life, suggesting that humans operate against a backdrop of what they called the 'unbearable automaticity of being.' In recent years, a burgeoning industry of books and consultants has emerged to help companies better target and exploit the many psychological biases that have been discovered by this line of research (Fogg 2003, Eyal 2014, Parr 2015). In the area of public policy, too, the notion of 'nudge' (Thaler & Sunstein 2008) has been prominently discussed as a way to help individuals make decisions that better promote their wellbeing. As Rogers (2014) writes, 'the rise of behavioral economics closed the theoretical gap between the psychological crowd and the market crowd, creating a unified field theory explaining that the macro-behavior (Black-Scholes) and micro-behavior of individual actors (Nash, Simon) were simply two views of the same coextensive process.'

4. The fourth trend is the increasing rate of technological change itself. Historically, our information and communication technologies took years, if not generations, to be adopted, analyzed, and adapted to. Today, however, new technologies can arrive on the scene and rapidly scale to millions of users in the course of months or even days, and indeed are incentivized to do exactly this by the market forces of the attention economy, where products are provided for 'free' (but where, as it is often said, 'the user is the product'). The constant stream of new products this unleashes—along with the ongoing optimization of features within products already in existing use—can result in a situation in which users are in a constant state of learning and adaptation to new interaction dynamics, familiar enough with their technologies to operate them but never so fully in control that they can prevent the technologies from operating on *them* in unexpected or

> undesirable ways. In addition, as both the demand and the potential for more intuitive, natural user interfaces continues to grow, the role of automation is playing an increasingly important role. The mantra 'AI is the new UI' is informing much of the next-generation interface design currently underway (e.g. Siri, Alexa, Google Home), and the more that this vision of computing as intelligent, frictionless assistance becomes reality, the more the logic and values of the system dynamics will be pushed below the surface of awareness to the automation layer and rendered obscure to users, or to any others who might question their design.

We say that we live in the 'Information Age,' but perhaps a better name for it would be the 'Age of Attention.'[2] The product of all these convergent trends is an informational environment whose dominant character is that of *persuasion*.

II. Persuasion and Freedom

Of all the ways human beings try to influence each other, persuasion might be the most prevalent and consequential. A marriage proposal. A car dealer's sales pitch. The temptation of Christ. A political stump speech. This document. When we consider the stories of our lives, and the stories that give our lives meaning, we find that they often turn on pivot points of persuasion.

But what exactly *is* persuasion? Despite its continued importance throughout history, we still don't have a solid definition. Since ancient Greece, persuasion has been understood primarily in its linguistic form: *rhetorike techne,* or the art of the orator. Aristotle identified three pillars of

[2] This is admittedly an individualistic and psychologized framing of the shift I am describing; for a discussion of similar trends at the societal level, cf. Schroeder's (2013) notion of our era as 'an age of limits.' (The concept of the 'technology-culture spiral,' a process through which technology displaces established cultural commitment devices, is particularly helpful.)

rhetoric—ethos, pathos, and logos—which roughly correspond to our notions of authority, emotion, and reason. And Aesop's fable 'The North Wind and the Sun' illustrated the power of persuasion over coercion: in it, the sun and the wind compete to force a traveler to remove his cloak; while the wind blows coercive gales to no avail, the sun warms him with gradual beams of persuasion and ultimately wins the day. Into Medieval times, the centrality of persuasion continued via its position in education, alongside grammar and logic, as one-third of the classical trivium.

More recently, Powers (2007) surveyed the range of definitions in use to identify several broad characteristics of persuasion: it is usually intentional, it includes 'one or more senders, a message, and one or more receivers,' it involves behavior or attitude change, and it occurs via psychological processes. But perhaps more importantly, persuasion also requires that the receiver be free to 'accept or reject the message' as they wish (Powers 2007). On this criterion, we tend to define persuasion by what it *isn't*. It is not coercive, it is not manipulative; it is not deceptive, it is not a brainwash. Rather, it lies in the space that is left over after we exclude all the ways we might try to change someone's attitude or behavior in a way that limits their ability to respond in a manner of their own choosing (Fogg 2003, Powers 2007). In other words, an essential criterion for persuasion is *freedom*. Persuasion is a deliberate attempt to influence that you are free to handle or engage with as you wish.

I use the word 'freedom' here in a broad sense: it's the 'independence and authenticity of desires' as well as the 'ability to act' on those desires (Zalta et al. 2012). Bernard Williams (2001) thought of freedom as a 'ratio concept': the relationship 'between what people desire to do and what they are prevented by others from doing.' In some contexts, a distinction is drawn between the freedoms of action and desire, with the latter often referred to as 'autonomy.' But here I will refer to both

variants as 'freedom' and 'autonomy' interchangeably, because the scope of persuasion includes changes both to behaviors and desires.

Historically, we have conceived of freedom in rational terms because we have understood the human mind as primarily rational. Plato and Aristotle took reason to be the essence of humanness, and in the eighteenth century Immanuel Kant, drawing on the ideas of Rousseau, defined autonomy as rational self-rule that exists free from all other influences. (Those influences could be external, like other people, or internal, like your own emotions.) In a Kantian world, 'to be an individual is to respond to the circumstances and to act on the basis of reasons' (Savulescu 2007). But of course everyone is different, and we all navigate our lives through different circumstances that afford us different reasons—and so John Stuart Mill (1859) later extended Kant's conception of rational autonomy to account for this, proposing that the act of making one's own decision is valuable *for its own sake*. For Mill, the very ability to choose for oneself, for any reason, forms a large part of our concept of human well-being.

Currently, the most influential account of autonomy is the one developed by Frankfurt (1988) and Dworkin (1988). Often called the 'hierarchical' account due to its emphasis on the multi-leveled structure of desires and volitions, it has two necessary conditions for autonomy: competency and authenticity (Zalta et al. 2012). Criticisms of the hierarchical account have tended to target the 'authenticity' condition, in particular via the so-called 'problem of origins' which asks how we can have truly authentic desires if we are completely shaped by our world. But requiring that autonomy exhibit a complete independence from all influences is to move its definitional goalposts into an unrealistic zone—one which is perhaps amenable to hypothetical scenarios, but which has very little potential to inform practical decisions about ethics in the real world.

In other words, even though freedom is about the absence of constraints, it is not about the absence of *all* constraint. *Some* constraint must exist if freedom is to be a meaningful concept at all. As Frankfurt (1988) puts it: 'what has no boundaries has no shape.' This is not necessarily a bad thing, as limits on autonomy can promote justice, help prevent harm to others, and advance the public interest (Savulescu 2011). Sometimes constraints can even *enhance* our autonomy: both love and reason, for example, 'require a person *to submit* to something which is beyond his voluntary control and which may be indifferent to his desires' (emphasis Frankfurt's). To this list we could add language, culture, or even certain technologies. Yet this distinction between *negative* and *positive* freedom (i.e. freedom *from* vs. freedom *to*), to use Isaiah Berlin's (1969) well-known terms, is too rarely made in contexts of technology analysis, use, or design. In his book *Moralizing Technology*, Peter-Paul Verbeek (2011) suggests that we should recalibrate our commonly used notions of freedom amidst technology to mean not only an absence of technological constraints, but also 'a person's ability to relate to what determines and influences him or her.' It is a justifiable question whether 'relate to' is too vanilla of a verb to describe something so essential as the ideal relationship to technological influences, but regardless Verbeek's point resonates clear: it is not only freedom *from*, but also freedom *through*, with which we should be concerned.

Because we have historically defined freedom in rational terms, we have approached the ethics of persuasion in the context of constraints on rationality. Thus the types of influence for which we have clear-cut ethical boundaries—and thus often specific, separate names—all limit rational thinking in some regard:

- **Coercion**—The use of force, and thus the clearest sort of non-persuasion. For Bernard Williams (2001), coercion is where the development of any understanding of freedom must begin.

- **Manipulation**—Influence 'by controlling the content and supply of information' (Pennock 1972) in such a way that a person is not aware of the manipulation (Strauss 1991, Kane 1996, Pereboom 2001, Sripada 2011, Smids 2012).
- **Deception**—The use of 'false or incomplete information' (Smids 2012, Strauss 1991)
- **Brainwashing**—The use of physical methods to disorient or confuse people and then prime them for psychological conditioning which makes them compliant (Powers 2007)

Other categories exist, but these are some of the major types of constraint that are widely agreed to diminish autonomy.

Yet psychological research continues to reveal new insights about the role of the non-rational mind in the persuasive process. For example, Robert Cialdini (2006) identified six main principles of influence that increase the likelihood of persuasive success: reciprocity, commitment and consistency, social proof, authority, liking, and scarcity. Albert Mehrabian (1971) found, in a widely cited (and mis-cited) investigation, that when persuading someone about one's own feelings, only 7% of the resultant liking stems from the words themselves; tone of voice accounts for 38% and body language accounts for 55%. Neuroscience has recently offered interesting insights in this area as well: Falk et al. (2010) identified certain 'neural correlates of persuasion' and showed that persuasion was not simply a subjective experience, but in fact could be linked to 'a distinct set of neural regions typically invoked by mentalizing tasks.' Falk et al. later illustrated (2010) that sometimes persuasion is not even subjectively experienced at all, and that neurological methods for predicting behavior change can produce even a more accurate prediction of persuasive success than the self-reported predictions of the person being persuaded.

While we currently have no grand unified theory of how persuasion works (Fogg 2003, Cameron 2009), Cameron surveyed fifteen choice-persuasive theories, models, and frameworks to define four broad categories:

- **Message effects models** – Models that focus on the role of particular variables in a persuasive message (e.g. the Yale Model of Persuasion (Hovland 1953), the Protection Motivation Theory (Rogers 1975), and Extended Parallel Process Model (Witte 1992))
- **Attitude behavior models** – Models that attempt to predict behavior from attitudes (e.g. the Theory of Reasoned Action/Theory of Planned Behavior (Sheppard et al. 1988, Ajzen 1991) and the Triandis Model of Interpersonal Behavior (Triandis 1977))
- **Cognitive processing models** – Models that emphasize the cognitive processes by which attitudes change (e.g. Dual-process models such as the Elaboration Likelihood Model (Petty & Cacioppo 1986), the Heuristic Systematic Model (Chaiken 1980), and Social Judgment Theory (Sherif et al. 1965))
- **Other models** – Various other models e.g. Inoculation Theory (McGuire 1964), which emphasizes resistance to persuasion, and other 'functional approaches,' which emphasize reflection on the underlying causes for existing attitudes and beliefs

Within individual disciplines, other more specific models of persuasion also exist. In technology studies, for example, frameworks such as the Technology Acceptance Model (Davis 1985) and the Unified Theory of Use and Acceptance of Technology (Venkatesh et al. 2003) support field-specific applications of persuasion. And in marketing and advertising, of course, conceptions of the persuasive process abound.

Some of these models do a better job than others at integrating our knowledge about the pervasive role of non-rational mechanisms in human cognition. In particular, the 'cognitive processing' models—especially the 'dual-process' ones—may ultimately prove most accurate and broadly

applicable. The term 'dual-process' there refers to the two cognitive routes by which persuasion may take place: the central (direct) and peripheral (indirect) routes. The two dual-process theories most widely quoted in the research literature are the Elaboration Likelihood Model (Petty & Cacioppo 1986) and the Heuristic Systematic Model (Chaiken 1980), but because of their close similarity the Elaboration Likelihood Model is regarded as the standard. (The Heuristic Systematic Model differs mainly by emphasizing the role of heuristics in decision-making.) One way to understand the two pathways of persuasion in dual-process models is by comparison to the two 'systems' of thinking that Daniel Kahneman articulated and popularized in his recent book *Thinking, Fast and Slow* (2011). Here, Kahneman's 'System 1' (intuitive thinking), which is fast, emotion-based, and more likely to employ heuristic conventions, corresponds to the 'peripheral' route of persuasion. His 'System 2,' on the other hand, is rooted in slow and methodical reasoning and corresponds to the 'central' route of persuasion.

In light of the importance of non-rational mechanisms in persuasion, how can we advance our ethical language and understanding to encompass them as well? Stanley Benn (1967) points out that 'liberalism has never taken much notice of how men come to want what they do want,' and that the 'traditional target for liberal critics . . . has been censorship, the monopolistic control of the supply of ideas, not the techniques used to persuade people to adopt some ideas rather than others.' For Powers (2007), such an ethical evaluation might involve clarifying the gray area she calls 'destructive persuasion,' which could include notions such as the engineering or manufacturing of consent (Lerbinger 1972, Herman and Chomsky 1988), propaganda (Bernays 1928, Ellul 1965) or subliminal persuasion (Strahan et al. 2002, Cooper and Cooper 2006, Legal et al. 2011). But there has been too little attention paid to these ethical questions that exist at these definitional edges of persuasion, and especially in a way that could have broad applicability across domains. If non-rational mechanisms really are responsible for the amount of persuasive effect that

the research suggests, then we should aim to know more, not less, about which types are ethical and which are not.

III. Persuasive Technology

In parallel with our advances in understanding the human mind, technologies have increasingly been designed explicitly to effect persuasive ends. If we consider 'technology' in the broadest sense of the word—the human shaping of the world in a particular way and for a particular end—it is clear that this is at core nothing new. Humans have long designed physical environments, for example, that explicitly aim to change people's behavior or attitudes: the placement of escalators in shopping malls, the music in grocery stores, or the layouts of cities (Goss 1993). In a certain sense, all design could be thought of as persuasive (Redstrom 2006). But recent technologies—particularly those arising from the large-scale interaction of digital computing systems on the Internet—are exhibiting new types of persuasive potential without precedent. And they are being developed and deployed across diverse domains such as health, education, and commerce—a trend that is only increasing as technology advances (de Ruyter and Pelgrim 2007, Lobo et al. 2009, Kosta 2010, Nakajima et al. 2011).

On the Internet in particular, the convergence of myriad technologies and data sets has greatly increased—and continues to increase—the potential for ever-greater persuasive power. Perhaps the most sophisticated and pervasive examples occur in commercial applications, though these technologies are by no means unique to that domain. In the global online advertising industry, a highly advanced infrastructure of behavioral measurement, predictive analytics, and adaptive delivery of persuasive cues has evolved. Online advertising may be the most targeted system of large-scale information delivery in the world today: some of the most sophisticated algorithms

known are managing the way in which little boxes of light drop in front of our eyes and beg us to buy soap, or ties, or plane tickets. This is the primary way in which we are now monetizing most of the information in the world.

Because online advertising has been leading the way in the evolution of persuasion on the Internet, it has been the context in which society has advanced the discussion of key ethical questions in areas related to:

- **Privacy**— Cranor & McDonald (2010) found in a survey of American Internet users that 69% believe privacy in the context of online advertising to be a right. Brown and Muchira (2004) also found that privacy concerns have a 'significant inverse relationship with online purchase behavior.'
- **Transparency**— The Internet Advertising Bureau has defined 'transparency' as the act of making legal papers available for download on websites, while the European Union recently enacted a mandate that site owners must inform users about the presence of 'cookies.' (Information Commissioner's Office 2012)
- **Trust**—Implications exist here not only for trust in the source of a persuader, but also for the information being monetized by way of that persuasion (Choi and Rifon 2002, Yang and Oliver 2005).
- **Attention and cognition**— Lincoln Dahlberg has called online ads' distracting power the 'corporate colonization of online attention' (2005). Significant implications exist here for definitions of media literacies, public policy decisions, and reducing harm for those without 'ad immunity' (e.g. children, new adopters in emerging markets) (Grant 2005, Calvert 2008).
- **Behavioral profiling**—Joseph Turow (2012), for example, outlined potential negative social effects of highly tailored information delivery based on behavioral profiling.

As we embed the Internet ever more deeply into our lives, more and more of the human experience will take place in environments of high persuasive prevalence and power—commercial or otherwise. Our experience of the Internet is rapidly becoming more mobile, natural (e.g. augmented reality and intuitive user interfaces), ambient, and connected—and the commingling of these trends creates a situation where persuasion can become more informed, more social, and ultimately more effective. Thus the evolution of persuasion in digital environments poses new and urgent ethical questions that require a holistic, cross-domain perspective to sufficiently explore.

Previous work in the moral implications of technology provides a helpful background for asking ethical questions about technological persuasion. Langdon Winner (1980) asked whether artifacts can be imbued with politics, giving the example of a bridge in New York State that some felt had been designed deliberately low so as to prevent inner-city school buses from passing underneath. Bruno Latour (1992) suggested that objects constitute the 'missing masses of morality' in a 'network of agents,' and that as mediators, rather than intermediaries, they actually function as part of people. The 'device paradigm' of Borgmann (1987) similarly held that things could be endowed with morality, and Floridi and Sanders (2004) outlined a conception of 'artificial moral agency' rooted in the interactivity, autonomy, and adaptability of a technology. More recently, Peter-Paul Verbeek (2011) has focused on the role of intentionality in our endowing technologies with morality, proposing that intentions are in fact 'distributed among human and nonhuman entities,' and that technologies' intentions are 'to be found in their directing role in the actions and experiences of human beings.' Verbeek's focus on intentionality for an understanding of the morality of technology is particularly useful for exploring the question of persuasion, because, as we recall, intentionality is a key criteria in its definition (Powers 2007).

Under the paradigm of technology as a *tool*, questions of design ethics, while still not simple, were much more straightforward. In that context, the reigning design philosophy has been 'user-centered design' (UCD), a perspective that emerged from the industrial engineering subfields of human factors and ergonomics. UCD primarily focuses on supporting the wants or needs of a user, giving particular attention to his or her primary tasks/goals, context of use, and other human factors that could impact effective use of the tool (Garrett 2010). Some approaches in particular that have been suggested to mitigate ethical risks in technology design are:

- The 'inscription method' – Offered by Jaap Jelsma (2002, 2006), this approach brings Latour's notion of an artifact's 'script' into the design process and suggests an eight-step method for the review and redesign of existing technologies.
- Value-sensitive design (VSD)—A widely-cited method proposed by Batya Friedman (1996, 2006), VSD proposes a 'tripartite methodology' consisting of conceptual, empirical, and technical techniques for eliciting the values at play in a product's design.
- Participatory design—Proposed by Floyd et al. in 1989, participatory design aims to maximize user involvement in the design process to democratize the product's creation. However, there appear to still be few applications in the context of persuasive technologies (Davis 2009).
- 'Disclosive ethics' (Brey 2000, Introna 2005)—Offers an approach for surfacing the intentional nature of information technologies to users.
- 'Moral imagination' (Verbeek 2011)—In which designers try to imagine all the possible effects and implications of their technologies at the time of design.

However, as B.J. Fogg (2003) points out, technology now functions not only as a tool or a medium, but also as a 'social actor'—a trio of uses that he has dubbed the 'functional triad.' On top of this, the very nature of persuasion as 'an attempt to change attitudes or behaviors' brings it into potential conflict with the aims of user-centered design. The user's goals are not necessarily the goals of the persuasive system. How can this conflict be reconciled?

Over the past decade an interdisciplinary field called Persuasive Technology has emerged to research the use of digital technology to intentionally change people's attitudes or behaviors (Fogg 2002, Fogg 2003, Oinas-Kukkonen 2010). The potential benefits of such a field are substantial: rather than examining persuasive technologies in the context of domain-specific applications, PT can systematically advance understanding of the underlying mechanisms used by these systems (Fogg 2003, Oinas-Kukkonen and Harjumaa 2008). This would enable the better sharing of knowledge of persuasive design across disciplines. In addition, PT can advance societal discussion and understanding of persuasive systems, a need that will only grow as they continue to increase in their persuasive capabilities. As a result, the cross-domain nature of PT makes it very well suited for the study of the ethics of persuasion.

Yet, curiously, ethics remains a relatively understudied aspect of the PT literature (Oinas-Kukkonen 2010). Berdichevsky and Neuenschwander (1999) initially provided a set of eight ethical principles for PT design using a rule-based utilitarian framework that culminated in the 'Golden Rule of Persuasion,' which suggests that 'the creators of a persuasive technology should never seek to persuade a person or persons of something they themselves would not consent to be persuaded to do.' Later, Spahn (2011) theorized about the application of discourse ethics to persuasive technologies but proposed few practical mechanisms for the design process. Finally, Kaptein and Eckles (2010) have utilized traditional philosophical methods such as 'reflective equilibrium' (which I will return to later on in the document) for judging the ethics of 'persuasion profiling' in particular.

But none of these attempts to clarify the ethics of persuasive technology have taken on the fundamental challenge of clarifying the nature of freedom amidst persuasion—especially in

contexts where the technology is more ambient, adaptive, ubiquitous, and socially aware. When does technological persuasion spill over into coercion, manipulation, deceit—or something else that may limit autonomy in just as objectionable a manner? According to Oinas-Kukkonen (2010) this research need is particularly acute when it comes to understanding the role of elements such as user consent, one's awareness of the persuasive process, the role of tasks or goals in the design of persuasion, and the importance of factors such as culture or gender in designing persuasion for user autonomy.

Such ethical questions about the gray areas of persuasion are by no means unique to the field of Persuasive Technology—nor, for that matter, to our time (Bobonich 1991). But they are urgent and practical for us in ways they could never have been for Aristotle or Cicero: the role and trajectory of current technological evolution represents the real-life implications of a question that could previously only have been hypothetical, one that has been posed by writers in areas as diverse as economics (van Tuinen 2011), law (Strauss 1991), health care (Barilan and Weintraub 2001), and more. Namely: where should the boundaries of persuasion lie? This question is immediately pressing in our time due to our increased understanding of the biases and pitfalls of the mind, as well as of the mechanisms that can exploit them for persuasive ends. In his book *Code* (2006), Lawrence Lessig wrote that cyberspace itself 'creates a new threat to liberty, not new in the sense that no theorist had conceived of it before, but new in the sense of newly urgent.' I believe that this is even truer about technologically augmented persuasion.

IV. The Synthetic Challenge: Aims and Methods

Thus the motivating question of this thesis is how—and indeed whether—individuals can remain autonomous in a persuasive media environment that increasingly shapes their attention and action

in ways that produce conformity with predefined goals. To address these questions holistically and in parallel with one another requires a truly interdisciplinary project spanning philosophy, ethics, psychology, sociology, politics, economics, design, and, to a lesser extent, other fields or subfields. Such a perspective does not reflect an insufficiently narrow scope, but rather an attack angle that is *problem*-oriented, as opposed to *field*-oriented. The problem to which this thesis responds is a very practical—and urgent—one, one that carries immediate implications for the most personal of all human considerations: namely, how we think and how we act. At such an apex of importance, many domains of human inquiry will inevitably commingle and find fruitful conversation. As Crawford (2015) writes:

> 'Clearly, no single discipline or body of thought is adequate to parse the crisis of attention that characterizes our cultural moment. There is a rich literature on attention in cognitive psychology, extending from William James's work of a century ago to the latest findings in childhood development. There are scattered treatments in moral philosophy, and these are indispensable. The fact has not been widely noticed, but attention is the organizing concern of the tradition of thought called phenomenology, and this tradition offers a bridge between the mutually uncomprehending fields of cognitive psychology and moral philosophy. What is required, then, is a highly synthetic effort—we can call it philosophical anthropology.'

When such a project of 'philosophical anthropology' is specifically aimed at clarifying the role of the technological environment in our current attentional crisis, it takes on a character akin to that of 'disclosive ethics,' as advanced by Brey (2010). Disclosive ethics 'makes transparent moral features of practices and technologies that would otherwise remain hidden, thus making them available for ethical analysis and moral decision-making.' According to Brey, disclosive ethics is distinct from 'mainstream computer ethics' in that the former focuses on 'morally opaque' aspects

of a technology (i.e. potentially morally salient aspects that are not immediately obvious), whereas the latter deals with 'morally transparent' (i.e. more readily identifiable) aspects.

Taking an approach that is broadly similar to that of disclosive ethics seems appropriate for the analysis of persuasive technologies for four reasons. First, we could say that persuasive technologies have more than one sort of 'opaqueness' that we want to make transparent: not only the behind-the-scenes mechanisms of the technologies, but also the nonrational psychological biases they exploit, which operate below the surface of the user's conscious awareness. Second, disclosive ethics seems particularly well-positioned to enable ethical analysis at the level of the overall ecosystem, which consists of 'different parties responsible for the design, adoption, use, and regulation,' as opposed to solely the level of individual technologies and their designers. Third, the aim here is not to arrive at anything remotely like a 'final' answer to these questions of freedom and persuasion in the attention economy—that would be a project for a lifetime, not for a thesis—but rather to lay the necessary groundwork for these questions to be productively advanced moving forward. Fourth, the multi-disciplinary character of disclosive ethics makes it a good fit for the nature of the challenge at hand.

Yet all disciplines are not created equal, and thus some will occupy privileged positions in this thesis. To determine this prioritization, I will take as my guide the framework of 'transdisciplinarity' provided by Max-Neef (2005), the hierarchically structured levels of which are four-fold: at the lowest level, we find disciplines that answer the question of 'what *exists*' (e.g. physics, chemistry); above those are the disciplines that tell us 'what we *can* do' (e.g. engineering, commerce); at the next level are the disciplines that tell us what we *want* to do (e.g. design, law); finally, at the top, we have those domains of inquiry that answer the question of 'what we *must* do' (e.g. philosophy, ethics). Max-Neef's framework allows for clear delineations to be made between descriptive,

technical, and normative considerations—but in a way that enables, rather than hinders, productive conversation between them. In this view, then, any truly holistic analysis is ultimately philosophical or ethical in its aims, regardless of whether those aims are explicit or latent, in a similar way to how attention to perceptual information occurs in a fundamentally goal-oriented manner. In a sense, Max-Neef's transdisciplinary framework could be seen as an executive-control function for human inquiry, a vision of 'top-down' attentional allocation at the societal or global level. Where science and technology overestimate their capacities and try to answer 'what *we must* do,' as Neil Postman described in *Technopoly*, this picture becomes inverted.

I will, however, make one enhancement to Max-Neef's model. I reserve for *language* a place outside the hierarchy altogether. Other disciplines (or subdisciplines) may also deserve a similar position of meta-ness (for example, metaethics, or the philosophy of disciplinarity itself). But language is still more *meta* than that, and, as I will show, is *meta* in a way that carries enormous implications for my topic of analysis in this thesis. True, linguistic *considerations* may exist within the levels of the transdisciplinary framework—e.g. the rhetorical techniques of a technology company's marketing efforts, or the emergence of new words for new experiences that let users of technology articulate grievances that otherwise would have remained mere vague intuitions—but language *itself* is the scaffolding of it all, the primary medium, the formal cause of all human inquiry. Thus, throughout this thesis I will continually return to the role of language, and in particular of metaphor[3], in shaping not only the answers to my questions but even the contours of the questions themselves. Taking as my mantra Wittgenstein's (1998) claim that '*die Grenzen meiner Sprache bedeuten die Grenzen meiner Welt*,' I will continually push on the ceiling of language in hopes of finding hidden doors.

[3] As Lakoff & Johnson (1980) point out in *Metaphors We Live By*, their seminal book on the topic, 'Metaphor is pervasive in everyday life, not just in language but in thought and action. Our ordinary conceptual system, in terms of which we both think and act, is fundamentally metaphorical in nature.'

However, while I will give particular attention to language in this thesis, it will not be my primary object of analysis. My primary object of analysis will be the negotiation of control over attentional processes between people and their informational environments. At times this may require focusing on a particular *type* of technology, or even a particular *instance* of a technology (much as we might say that 'Save Our Parks!' and 'Save Port Meadow!' are subsets of the more general exhortation to 'Save Our Environment!'). However, I do not see a need here for any elaborate definitional or taxonomic effort to specify what counts as a 'technology,' or how we ought to carve them up conceptually at the outset: such distinctions are ultimately incidental to the goal of understanding the informational environment as a whole. (Where such distinctions do seem warranted in the context of specific situations or technologies, I will make them there.)

The thesis is structured as follows:

Introduction: Technologies of Attention

Chapter 1: What Is a 'Persuasive' Technology?

Chapter 2: Autonomy & Persuasive Technology: Reactions & Distractions

Chapter 3: Reclaiming Advertising Ethics

Chapter 4: The Moral Character of Digital Advertising

Conclusion: Stand Out of Our Light

Bibliography

This introduction provides conceptual background that sets the stage for the core chapters of the thesis, providing an in-depth picture of the current state of research on topics that frequently appear such as autonomy, attention, persuasion, design ethics, user-centered design, and more.

In Chapter 1, 'What is a A "Persuasive' Technology?", I address the question of what should count as a persuasive technology, and carry out the conceptual and linguistic groundwork necessary at the outset of this thesis. I begin by discussing the lack of clarity that currently exists about what it means to say that a technology is 'persuasive.' I then propose a conceptual framework for persuasive technology that aims to contribute to the current understanding in three ways. First, I propose a way to consider more adequately the question of intention when determining a technology's 'persuasiveness.' Second, I disentangle normative conditions—in particular the requirement that persuasive technologies avoid coercion—from the descriptive conditions of their definition. Third, I suggest an approach to the language of persuasive technologies that will allow us to handle questions of intention, as well as normative considerations, in a more nuanced way.

In Chapter 2, 'Autonomy & Persuasive Technology: Reactions & Distractions,' I argue for two main theses. First, many criticisms of persuasive technology that appear to concern questions of autonomy in fact concern questions of dignity, and originate not in facts about manipulation of choice but rather in irrational and defensive psychological responses of *reactance*. Clearly distinguishing between the two carries large implications for ethical analysis of technology generally. Second, questions of both autonomy and dignity in the context of persuasive technology can be greatly clarified if viewed as questions of *attention*. I outline initial steps toward a framework for such a view, within which an expanded concept of *distraction* emerges as a useful construct for analyzing potentially problematic persuasive mechanisms. In order to typologize these persuasive mechanisms and establish a common language for describing them, I introduce a framework consisting of three main types of distraction, which I call Functional, Existential, and Epistemic Distractions.

In Chapter 3 I turn to the question of advertising. Although digital advertising is the dominant business model in the attention economy, to date it has received neither the amount of ethical analysis nor ethical guidance that it deserves. I suggest that this situation results, in large part, from the general failure of advertising ethics over the past century. Four dimensions of that failure are particularly relevant for engineering the future of digital advertising ethics: (1) because advertising ethics has received much less attention than it deserves, the field has failed to significantly guide the practice of advertising; (2) due to its framing as a subfield of business ethics, advertising ethics has suffered from an agent-centered ethical perspective; (3) work in advertising ethics has generally failed to account for the way in which information abundance produces a scarcity of attention; and (4) the definition of 'advertising' has been perennially vague. In response to the definitional problem, I develop what is, to my knowledge, the first user-centered definition of advertising: *a proactive appeal for a resource of value made in a way that overrides the dominant design goals for information delivery in that medium*. That is to say, when viewed as media dynamic, advertising functions as an *exception to the rule*. I close by discussing the urgency of advancing and accelerating ethical work on the unique challenges posed by specifically *digital* advertising.

In Chapter 4 I assess the nature and ethics of digital advertising and argue that it poses important challenges to human self-determination that have to date gone largely unacknowledged. I show how digital advertising radically departs from advertising as historically understood, particularly in its *boundedness* and its *intelligence,* such that it no longer functions as an *exception* to the rule; instead, it now *is* the rule. In traditional media, advertising was said to 'underwrite' the dominant design goals; in digital media, advertising now 'overwrites' them with its own. As the now-dominant design logic, it is thus arguable whether digital advertising ought to be described as a form of 'advertising' at all. This 'overwriting' tendency of digital advertising has produced several direct as well as systemic effects that pose serious challenges to user self-determination. However, these issues have

received little serious ethical attention to date and thus warrant a sustained project of critical analysis. I close with a discussion of the potential ethical obligations, as well as opportunities, that follow from this assessment of digital advertising— several of which sit in direct tension with other ethical interventions that have been proposed or implemented to date.

V. Conclusion

There is an opportunity right now to advance our ethical understanding about technological persuasion, particularly as it relates to the question of freedom. The field of Persuasive Technology is an ideal starting point from which to ask these questions because it is where the answers could be carried forward to have the greatest impact on our lives across a wide number of domains. These answers would, of course, carry large implications for any mechanisms we might develop to regulate or countervail the wrong sort of persuasion—mechanisms that could ultimately help us preserve, if not augment, our autonomy. In my conclusion, I will briefly consider a high-level view of the range of such regulatory interventions in the attention economy, and will ask how they might be usefully characterized and categorized. However, it is not my aim here to identify or elaborate these interventions comprehensively or in detail; to do this well would, of course, demand a thesis of its own. In the same way, it is beyond the scope of this, or any, doctoral thesis to produce a comprehensive consideration of the ethics of persuasion—or even just that of persuasive technology. My hope, however, is that this effort will serve as a firm step in the right direction.

We have understood for a long time the importance of persuasion in human life, and in recent years we have realized that technology has come to occupy a similarly central and prevalent role. So far, technology has been more like the North Wind in Aesop's story—brute, direct, and

coercive—but it is in the process of becoming subtler, more ambient, and more persuasive, like the Sun. As we look to the future, it is possible that persuasion and technology could be the two forces in which human life will become most irreversibly embedded. It is the goal of this thesis to explore how these two aspects of life should ideally work together in a way that preserves, if not enhances, human autonomy.

Chapter 1:

What Is a 'Persuasive' Technology?

While it has become popular to describe some technologies as being 'persuasive,' we lack clarity as to what this actually means. In this chapter I propose a conceptual framework for persuasive technology that aims to contribute to the current understanding in three ways. First, I propose a way to consider more adequately the question of intention when determining a technology's 'persuasiveness.' Second, I disentangle normative conditions—in particular the requirement that persuasive technologies avoid coercion—from the descriptive conditions of their definition. Third, I suggest an approach to the language of persuasive technologies that will allow us to handle questions of intention, as well as normative considerations, in a more nuanced way.

I. The Emergence of 'Persuasive' Technology

If John Stuart Mill was right when he wrote that 'all that makes existence valuable to anyone…depends on the enforcement of restraints upon the actions of other people,' then much of what makes existence valuable in our time depends upon the enforcement of *technological* restraints. In the digital age, much ethical analysis has grappled with technological constraints that relate to the management of information, such as user privacy or data protection. However, less attention has been given to constraints on attention itself, such as the way in which technologies exploit their users' non-rational psychological biases. The 'traditional target for liberal critics,' wrote Stanley Benn in 1967, 'has been censorship, the monopolistic control of the supply of ideas, not the techniques used to persuade people to adopt some ideas rather than others.' Two years later, Herbert Simon observed that information abundance in fact renders attention—the starting point

for human thought and action—the scarce resource. Despite the rise of the so-called 'attention economy' (Lanham 2006) and its role as the default business model of the Internet, we have yet to give attention its full due in our ethical questioning of the digital world.

Designing attentional constraints in ways that produce particular attitudinal or behavioral outcomes is nothing new. In the physical sphere, one need only look out the window to find speed bumps, street signs, stop lights, and so on. Entire environments are often constructed toward persuasive ends: consider the strategic placement of escalators in department stores to maximize exposure to products, the careful selection of ambient music in grocery stores to encourage consumption, or the layouts of cities to achieve particular efficiencies of traffic flow (Goss 1993). In the age of media and technology studies, persuasive systems have been addressed in the context of notions such as propaganda (Bernays 1928, Ellul 1965) the engineering or manufacturing of consent (Lerbinger 1972, Herman and Chomsky 1988), and more recently, the notions of 'gamification' and 'nudges' (Thaler and Sunstein 2008). And of course, any such persuasive endeavor traces its roots back to the tradition of classical rhetoric.

While *all* design could be seen as ultimately 'persuasive' in some sense (Redström 2006), recent technologies—particularly those arising from the large-scale interaction of digital computing systems on the Internet—are exhibiting new types of unprecedented persuasive potential. Perhaps the most sophisticated and pervasive examples occur in commercial applications, where a highly advanced infrastructure of behavioral measurement, predictive analytics, and adaptive delivery of persuasive cues has evolved; online advertising may be the most targeted system of large-scale information delivery in the world today.

Over the past decade an interdisciplinary field called Persuasive Technology (PT) has emerged to focus on the analysis and design of these technologies (Fogg 2003, Oinas-Kukkonen 2010). The potential benefits of such a field are substantial. For one, rather than examining persuasive technologies in the context of domain-specific applications, PT can systematically advance understanding of the underlying technological and psychological mechanisms used by these systems (Fogg 2003, Oinas-Kukkonen and Harjumaa 2008). This can enable the better sharing of the knowledge of persuasive design across disciplines. In addition, PT can advance societal understanding of, and debate about, persuasive systems—a need that will only grow as they continue to increase in their persuasive capabilities. Finally, such a cross-domain perspective also makes PT particularly well suited for the study of ethical questions, in particular those related to the freedom and autonomy of users.

At the same time, there is a lingering question of scope of in the field of PT. What ultimately *counts* as a persuasive technology? (Here I will use the phrase 'persuasive technology' to refer both to the field and its objects of analysis.) One cause of this definitional murkiness is confusion about the precise conditions for 'persuasiveness' in commonly used definitions of the term. On a deeper level, however, this confusion reflects a lack of common purpose among those who have defined and deployed the term; insufficient attention has been paid to the question of what we want the phrase 'persuasive technology' to do for us. Finally, these two challenges have been exacerbated by the extremely fragmented language that already exists for describing persuasion in general.

Regarding the conditions a technology must meet in order to be 'persuasive,' the general consensus seems to be that there are three: (1) there must be an intention on the part of its designers to (2) change, reinforce, or otherwise shape human behaviors or attitudes, or both, in (3) a manner that is not coercive.

However, these three conditions pose a number of problems. While they do carry an intuitive appeal, and even mirror the current definitional state of the concept of persuasion in general (Powers 2007), as currently given they serve as an unstable footing from which to classify, analyze, or even design technologies. Two of the three conditions have nothing to do with the technologies themselves; they deal instead with the human concerns surrounding their development. Yet even if we begin with a broader, more relational conception of technology as a socio-technical system, it becomes clear by looking at each of these conditions in greater depth that they will still require more clarification and revision to be of use.

II. Current Definitional Challenges

Intention

The first condition for PTs—the presence of a particular intention on the part of the designers—presents two main problems. First of all, *all* design can be seen as intentional—and even persuasive—to some degree. This has been widely acknowledged in the field of PT, but no satisfying distinctions about what *sort* of intentions should count have yet emerged. For example, Lockton et al. (2008) proposed situating PT within a larger field termed 'Design with Intent,' although such a characterization seems redundant because, at least in the realm of digital technologies, it is hard to imagine design *without* some sort of intention for its use. Furthermore, casting a definitional net so widely that *any* technology could conceivably count would seem to produce a distinction of marginal utility indeed.

In the PT literature, the boundary between PTs and non-PTs seems to exist in the same general zone as the distinction between 'design goals' and 'technological affordances.' For example, a shelf

in a refrigerator may *afford* a user the ability to place food on it, but that affordance is neutral as to whether the food is whole milk, Gouda, Hoegaarden, or whatever else. Conversely, the little round egg-holder in the refrigerator's door may impose a more restrictive set of constraints that makes it infeasible, and therefore less likely, to be used to store something that is not an egg. Therefore, using this same distinction (and assuming, for the sake of argument, that the designers intended these parts of the refrigerator to be used in the ways mentioned here), then *ceteris paribus* the egg-holder would be considered a more 'persuasive' technology than the shelf—if indeed the latter would be considered persuasive at all.

The problem with defining a persuasive technology in this way is that it actually has nothing to do with persuasion at all, neither in intent nor in outcome. In reality, this approach makes its central distinction on the basis of the *specificity* of design goals and their related technological constraints. Here, the more specific goal is taken to be more 'persuasive.' However, affordances can also be design goals, and the broader design goal of giving the user a multi-purpose shelf is no less persuasive simply because it is less specific about the particular conditions of what should be stored on it. In fact, a *less* specific design goal may in some cases be *more* persuasive than a more specific one because there is a greater percentage of all possible outcomes that could satisfy its criteria for success. There is no inherent correlation that I can see between goal specificity and degree of persuasive power.

The second problem with the intention condition for PTs is the arbitrary limitation of scope to the intentions of *designers*, which omits many important intentions that inform a technology's design. Most technologies, whether persuasive or not, emerge as a result of supervenient levels of organization—teams, companies, societies, economies, etc.—and inherit intentions that have trickled down from those higher levels. These broader organizational intentions may even remain

invisible to the designers themselves—or, if visible, they may be rejected or opposed by the designers but still be reflected in the design. In his foundational textbook on the topic, Fogg (2003) writes that PT should be focused on '*endogenous* intent, that is the persuasive intent that is designed into a computing product,' as opposed to '*exogenous* intent,' i.e. the persuasive intent a technology acquires when it is 'adopted for a persuasive goal the designers hadn't planned.' However, even if we agree to limit the scope of PT to endogenous intent, it is not clear why we should limit 'endogenous' intent to that of designers. To expand the notion of endogenous intent to include the full set of intentions that are reflected by a technology's design, we could simply reinterpret 'designer' as meaning any person or group whose intentions trickle down to, and are reflected in, the ultimate design of the technology. Or we could take the actual, hands-on designers to be representatives of the organizations in which they operate, and assume that their intentions and the intentions of their systems are one.

In any event, to interpret intention solely as a property of designers is to overlook a great many of the intentions that actually inform a product's design. In fact, for technologies that inhabit the vast and interrelated persuasive ecosystem we call the 'attention economy,' it may even omit the *majority* of intentions that shape their designs. It is curious that so many of the particular cases addressed in PT have arisen in academic research contexts, as though the aim were to place a particular technology in a test tube and study it in isolation from the broader systems with which it interacts in the real world. By focusing on these simpler cases, PT has largely avoided addressing many of the technologies that actually do persuade us on a day-to-day basis, especially ones with intentions that are ultimately commercial in nature. In addition, technologies with such commercial aims are also more likely to originate in a context of capabilities-based design, as opposed to the more needs-based design approaches that the PT literature often assumes informs a particular technology. For example, a commercial PT may at a given moment be optimizing its persuasive

mechanisms toward increasing the number of users and the frequency of their usage in order to pursue a higher-level organizational goal, such as securing more venture capital or being acquired by a larger company.

Behavioral or Attitudinal Change

The second condition for a technology to be persuasive—that the design intention be to shape (e.g. change or reinforce) human behaviors or attitudes—is largely problematic for its framing: it euphemizes the fact that designing a persuasive technology amounts to designing the user.

What should primarily give us pause here is the glossing over of crucial distinctions between the shaping of thought and action. The words 'behaviors' and 'attitudes' are often uttered in the same breath in definitions of persuasive technology, as though they were simply the two broad categories of possible outcome at which persuasive design could aim. In reality, the distinction between changing thought and behavior is much more profound. In fact, it represents the major blind spot in current models of PT and reveals the extent to which the metaphor of 'persuasion' has obscured sufficient consideration of user goals.

To illustrate the importance of this distinction, imagine that we are sitting on a beach and I decide I want to persuade you that the sky is green. Regardless of how I go about this task (e.g. by persuading you that your senses are unreliable, or that 'green' actually means 'blue' and that you have just been using the word incorrectly your whole life), my goal is clear. I want you to believe that the sky is green. But what is *your* goal? You may have peripheral goals for consenting to the persuasive attempt (e.g. you think I am setting up a joke for which you want to hear the punch line, or you want to spend time with me and grow our relationship regardless of how odd our conversations become). But with respect to your direct intentions for the persuasive exchange

itself, it would not be immediately clear what your goals are. In fact, provided that you take my persuasion as a serious attempt to change your view, your goals are only relevant insofar as they provide a playing field on which persuasion can operate. When persuading *that*, it is not to your goals, but instead to a general standard of truth, that I would appeal.

However, if I persuade you *to* take an action, your goals come to the fore and occupy a position of extreme relevance. Imagine that, after tiring of trying to persuade you that the sky is green, I decide I want to persuade you to jump in the ocean instead. Unlike the previous example, here I would not be appealing to the general playing field of your respect for truth: rather, I would be demonstrating how jumping in the ocean is salient to some other goal you currently hold. Regardless of what I take as my strategy—the central, rational route (e.g. 'you said you wanted to jump in the ocean five times this week, and you've only done it four') or the indirect, peripheral route (e.g. an appeal to scarcity: 'we'll only be on vacation one more day')—my attempt fundamentally appeals to your specific, existing goals in order to persuade you to create another goal.

So the difference between persuading *that* and persuading *to* is that the latter creates new goals in the persuadee as a direct result of the persuasion, whereas the former does not. In other words, persuasion to act is also persuasion to intend.

In light of this, the typical way of talking about behavioral 'change' in the field of PT, as primarily a difference in behavior between two points in time, seems insufficient to capture what really matters to either the persuader or the persuadee: namely, the salience of behavior change to some existing goal. In the end, the difference that really matters is the distance between possible futures—i.e. between desired and actual futures—and this is true for both persuader and persuadee. This view

of behavior change reframes persuasion *to* as an interaction of goal complexes—or, more properly, as a negotiation of hierarchical goal networks (Bagozzi 2003). It also makes the goals of users a much more prominent consideration than has been the case in research about PTs.

In technology design in general, the absence of any usable methods for understanding the full set of users' goals has resulted in insufficient attention on opportunity costs that emerge as a result of design. Ultimately, even User-Centered Design is only equipped to focus on a *slice* of the user, not the *whole* user: little guidance exists about how to consider user goals that do not relate to the immediate goals of a product. This is even more the case when it comes to persuasive technologies, where design goals have even greater prominence. Such 'behavioral externalities,' so to speak, can exist even when a PT moves a user toward a goal he or she wanted to move toward, as there still could have been a more valuable behavior the user could have carried out. (To the extent that a PT competes for a user's attention against internal stimuli, such as memories or goal representations, it may also produce 'internal' externalities as well (Kuhl and Chun 2014).) PTs often consider what a technology is pushing you toward, but not what it is pushing you away from (i.e. a user's unrealized goals). To a large degree, this is true of most technologies we use: they are typically bounded in a particular task domain and thus unable to account for the full set of our goals, which on the time scale of life essentially equates to our aspirations, our fulfillment, our wellbeing.

This distinction between behaviors and attitudes—between persuading *that* and persuading *to*—reflects a deep conceptual asymmetry in current models of PT. If truth is to persuading *that* as goals are to persuading *to*, then this asymmetry could carry a moral implication as well. When a persuader knowingly moves someone away from truth, we call it deception. When a persuader knowingly moves someone away from their goals, we call it distraction. But we do not place

deception and distraction on the same moral or ethical footing. Yet perhaps we should, if goals are the truth of action. There is a large opportunity to explore this territory further, which I plan to do in future work.

Noncoercion

The third condition, that PTs be noncoercive, poses three challenges. First, it defines persuasion negatively, not by what it *is*, but by what it is *not*. Namely, an attempt that does not limit a person from freely engaging with it in a manner of their own choosing (Fogg 2003, Powers 2007).

Second, it is not at all clear what counts as 'coercion,' especially when influence occurs via an indirect psychological route – i.e. intuitive, heuristic, automatic pathways (Kahneman & Tversky 1973, Chaiken 1980, Kahneman 2011). Bargh and Chartrand (1999) proposed that such automatic, non-conscious processes in fact form the majority of our experience of everyday life, suggesting that humans operate against a backdrop of what they called the 'unbearable automaticity of being.' With the emergence of so-called 'dual-process' models of persuasion, such as the Elaboration Likelihood Model (Petty & Cacioppo 1986) and the Heuristic Systematic Model (Chaiken 1980), it has become clear just how significant of a role the 'peripheral,' or non-rational, routes of persuasion play in the ultimate outcomes.

This rational vs. non-rational distinction is especially important insofar as it relates to cases where a technology might coerce a user toward certain lower-level, immediate goals in order to move them toward their higher-level, longer-term goals. For example, consider self-control technologies that prevent a user from accessing particular websites during particular times of the day. These technologies, where the persuasive mechanisms function as an extension of the user's own self-coercion, would seem to me very appropriate to include under the rubric of 'persuasive

technology'—not only because they ultimately promote the user's higher-order aims (which perhaps require a further exploration of how the distinction between freedom and autonomy should play a role here), but also because they in fact liberate the user from a deeper coercion—the coercion of habit, defined here as the default state that a PT would be trying to change his or her behavior or attitudes away from. So even if we *could* clearly define coercion, it is not clear that we would actually want to exclude all of it.

Again, this lack of a clear definition for coercion is not a challenge that is unique to PTs—it is a broader question—but it does make it difficult to answer the basic question of what counts as a PT.

The third problem with the noncoercion criterion for PTs is that I suspect it may really be functioning as more of a prescriptive than a descriptive condition. One major value of a field like PT is that it supports not only the analysis but also the design of technologies. However, my sense is that the emphasis on noncoercion here relates to the latter, representing an admirable exhortation to creators of PTs to ensure that their resultant designs remain ethical, yet nonetheless functioning as a case of normativity masquerading as description. In other words, there is a difference between saying 'persuasive technologies *must* be noncoercive' and 'persuasive technologies *are* noncoercive,' and perhaps the latter is often said when the former is meant. As a result, this confusion between prescriptive and descriptive modes may be omitting from consideration technologies that may be coercive in nature—and therefore more deserving of close analysis and criticism.

It is worth asking, however, whether this mixing of normative and descriptive considerations may actually be occurring by design. That is to say, might the noncoercion condition be intended to

influence not only the *construction* of these technologies, but also the *perception* of them as well? If we recall the linguistic fragmentation I described earlier—i.e. the myriad terms that exist for describing technologies that shape people's thought or action—we notice that they all essentialize around some specific aspect of the technology. (For example, the terms 'smart' technology and 'gamification' essentialize around automation and reward mechanisms, respectively.) Around what does 'persuasive technology' essentialize? It would seem that, like 'nudge,' the term 'persuasive technology' is designed to essentialize around *its own harmlessness*. Invoking a metaphor that already has its own highly nuanced language for describing freedom and control allows one not only to speak more clearly about those issues, but also to more easily deflect attention away from them via euphemism and meiosis.

In this way, the terms 'persuasive technology' and 'nudge' seem to euphemize the risks, if not the full effects, of the mechanisms to which they refer. While I suppose there is a pleasant irony in the fact that the word 'nudge' nudges, the pleasure is purely academic, and evaporates under the light of a clear description of what is going on: the design of human lives. This is not to say anything about the ethical appropriateness of 'nudges'; in many cases nudges are no doubt very appropriate, or even morally required. It is just to say that this sort of language pushes into the shadows of attention exactly the ethical considerations that, especially at this early stage of our understanding and exploitation of behavioral science, we ought to be calling forth into the bright light. Yet I find it extremely challenging to disabuse myself of the suspicion that the language here—the decision architecture of 'decision architecture,' if you will—has been designed to have precisely this effect.

III. Toward a More Intentional Definition

Given the benefits and challenges described above, we must ask whether 'persuasive technology,' as phrase and concept, still gives us a workable ground from which to grapple with the technologies that actually persuade—and often distract—us on a day-to-day basis. If so, what adjustments do we need to make in order to shore up its limitations? Or if not, what better options are out there?

I believe that the notion of a 'persuasive technology' is a useful construct, but there are four changes we must make to stabilize it conceptually:

1. We must develop a way of talking about the different types of intention that inform a technology's effects, and be clear about which ones are central to the definition of PTs.
2. We must develop an approach to intention that lets us deal with questions of goal specificity.
3. We must throw out the noncoercion condition.
4. We must develop more nuanced ways of describing PTs across the spectrum of normative considerations (in particular goal alignment and user freedom).

1. Categories of Intention

As described above, two problems with the way intention currently informs definitions of persuasive technology are that 1) all design could be seen as intentional, and even persuasive, to some degree, and 2) by limiting the scope of intention to that of designers, we omit a great deal of the intentions that actually inform a product's design. These two problems require us to expand our concept of intention beyond a narrow conception of designer intent—while also ensuring that we do not dilute its meaning beyond the point of usefulness in the process.

We must root any refinement of the concept of intention not only in an understanding of its essential characteristics, but also in a clear idea of what we want the concept to do for us. We are concerned with intention here in part because it is a condition for persuasion in a general sense, and because applying the metaphor of persuasion to technology requires that we also translate intention into this domain. But there is another, more practical reason we care about intention. All the influences that direct our lives, and that make it easier or harder to live the lives we want to live, can be classified into two groups: those where asking the question 'why' about their effects can produce an answer, and those where it cannot. We care about this distinction in part because asking 'why' reveals clusters of action in the world that we may manipulate in the service of our own goals: asking 'why' of the world can inform a 'how' for our own lives. The various types of answers we receive in response to the question 'why' could be broadly construed as different types of intentions.

One particular type of response to the question 'why' brings us to a conception of intention that aligns with Anscombe's. In fact, the question is so central to her analysis that at one point in *Intention* she goes so far as to write, 'the concept of voluntary or intentional action would not exist, if the question "Why?", with answers that give reasons for acting, did not.' Her view of intention, which we might roughly characterize as a description of an action that is taken as an instrumental step toward some further goal, broadly aligns with the standard understanding of intention in technology design. However, it still limits us to the realm of human agents—in order to have a world-to-word fit, you must first have words—and it is a long way indeed from the other varieties of intentionality with which we are concerned.

To develop a more comprehensive approach to intention, I suggest that we return to an earlier philosophical analysis for which 'Why?' was also the germinal question: Aristotle's analysis of the Four Causes. This model may serve as a useful tool for broadening the sense in which we ask 'why' of different aspects of the user-technology interaction, and for classifying the answers that are returned. For each cause—Material, Efficient, Final, and Formal—simply replacing the word 'cause' with 'intention' yields interesting results. In fact, Aristotle's Efficient and Final Causes arguably fit the description of 'intentions' better than 'causes' in the first place. The categories of Material and Formal Causes require a bit more tweaking, but are still useful for our purposes.

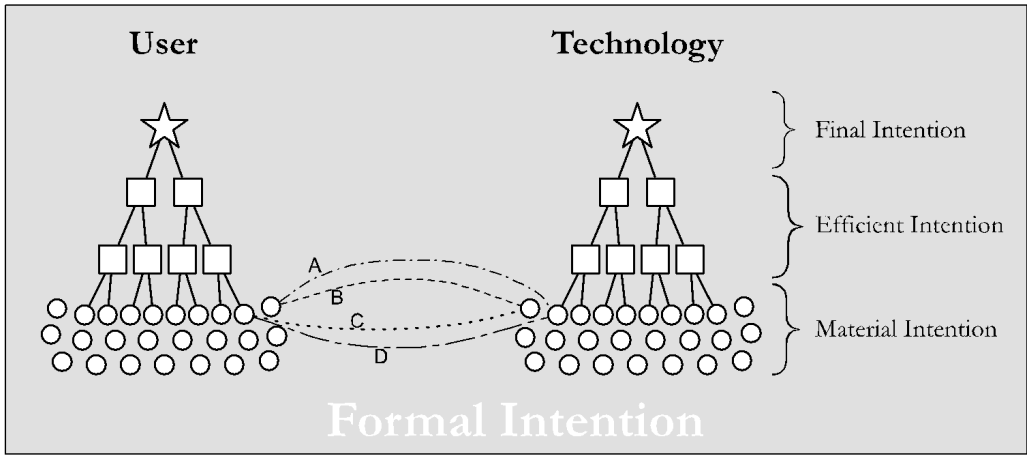
Translating the concept of intention more fully into the language of the Four Causes enables us to produce the following categories of intention:

> **Material Intention** – For a technology, Material Intention represents the elements of its composition or arrangement that have the potential to constrain, direct, or invite user action. It is akin to the broader view of technologies' 'intentionality' that, according to Peter-Paul Verbeek (2011), can 'be found in their directing role in the actions and experiences of human beings.' Similarly, the Material Intention of the *user* is the set of capabilities and biases that he or she brings to an interaction with a technology. These can be either physical or psychological (e.g. willpower, attention, habits) in nature.
>
> **Efficient Intention** – Efficient Intention is a representation of a possible outcome that may be served by, and/or inform the shape of, a technology's Material Intention. Efficient Intention is what is normally meant when discussing 'designer intent,' though it need not emerge directly from a human — it can exist wholly within a technological system. For

users, Efficient Intention is the hierarchical network of goals and tasks that serve higher-level ends.

Final Intention – Final Intention is like Efficient Intention but non-instrumental. It consists of goals that are pursued for their own sake. In the realm of technology design, Final Intention may consist of high-level values, organizational goals, or economic incentives. In the user realm, it may consist of a vision of the ideal self, 'being' goals, or *eudaimonia*.

Formal Intention – Formal Intention is intention in the background. It is the context of other Material, Efficient, and Final Intentions that gives rise to the possibility of the Material, Efficient, and Final intentions of both technologies and users. For a digital technology, this could include the intentions of the operating system, the programming language, HTML, TCP-IP, legal or economic systems, cybernetics, electronics, and so on. For a user, Formal Intention could include cultural norms and expectations, the media environment as a whole, education and knowledge, and even language itself.



One major benefit of this Four Causes approach to intention lies in the way it enables us to explore means-ends linkages between the different types of intentions. In the diagram above, Efficient Intentions (which may have any number of hierarchical levels, though I have only shown two for the sake of simplicity) are pursued as means to Final Intentions, represented here by stars. These top two sections of the diagram essentially correspond to goal networks as they are standardly described in the psychology literature (Bagozzi et al. 2003, Locke & Latham 2002, Austin & Vancouver 1996). By adding Material Intentions below them, however, we can link particular design features (in the case of technology) or capabilities/biases (in the case of users) to the broader goal network.

For example, in analyzing whether a particular technological influence should be considered a 'distraction' for a user, we could distinguish four different types of interaction based on the relationships between Material Intention and Efficient Intention on either side. In the diagram above, these interactions are represented by the four curved lines labeled A-D:

A - Intentional Distraction

B - Unintentional Distraction

C - Unintentional Support

D - Intentional Support

Not illustrated here are cases where the technology may influence elements of the user's formal intention (e.g. the cognitive capacities that undergird goal pursuit), either by 'brute-forcing' them over the course of repeated interactions or by adapting its persuasive mechanisms in response to behavioral feedback. (An example of this might be a game that varies the rewards it gives to the user based on the likelihood that the user is about to stop playing, thus increasing the addictive nature of the game and depleting the user's store of willpower.)

This approach to intention seems to meet many of the needs previously identified. It allows us to account for the immediate intentions of the designers (Efficient Intention), the affordances and 'directing role' of the technology itself (Material Intention), the broader organizational, economic, cultural, and other contextual intentions that inform the design and deployment (Formal Intention), and the end-directedness of the system as a whole (Final Intention). Furthermore, it allows us to bring user intentions to the fore and to move more easily between their analysis and the analysis of technological intentions.

2. Specificity of Intention

The other problem with intention in PT is that the part of the distinctively 'persuasive' character actually seems to be coming from the *specificity*, rather than the mere *presence*, of design intent related to user behavior. This requires us to develop a way of understanding these different degrees of specificity and then using them to clarify the definition of a PT.

The categorization of the Four Intentions provided above takes some steps in this direction. A further step could be taken by developing a formal model of a persuasive technology as a dynamic system. I will not construct such a model here, but will simply suggest three important considerations, drawing on the dynamic systems work of Terrence Deacon (2012), that are important in accounting for goal specificity.[4] First, does the persuasive system contain a representation of the goal toward which it is striving? Second, is the relevant outcome of its action—the behavioral or attitudinal change on the part of the user—measured by the system? Third, is there a mechanism within the system that assesses the outcome measurement against the represented goal (what in cybernetics is known as a 'comparator')?

[4] Deacon's notion of a 'teleodynamic' system warrants further exploration as a conceptual framework for persuasive technologies in particular, as well as artificial agents generally. He describes as 'teleodynamic' any system that has 'a dynamical organization that exists because of the consequences of its continuance' and that also contains 'within itself a representation of its own dynamical final causal tendencies.'

If the answer to the previous three questions is 'yes,' then, *ceteris paribus*, whether or not a system is 'persuasive' would depend on whether the specificity of the measured outcome is equal to or greater than the specificity of the representation of the goal within the system. (In principle, this condition is applicable for any of the four types of intention described above. In practice, however, we are most likely to see it satisfied for Efficient and Final Intentions.) As an example, imagine that you have a mobile app whose goal is to persuade you to go running at least once *every single day*. However, the feedback it receives about your behavior only tells it whether or not you have run once *in a given week*. Using the criteria above, this app would not qualify as a 'persuasive' technology because it does not measure, and thus cannot compare against its goal, at the required level of specificity.

These clarifications of the question of intention require further elaboration, which I will undertake in future work. For now, however, they provide a way to usefully grapple with the problems of intention I have described here.

3. Throw Out the Noncoercion Condition

As mentioned above, there is very little clarity about where the line between persuasion and coercion currently exists, especially as it relates to non-rational or 'peripheral' persuasive pathways. In fact, the search for one distinct line may be hindering us—across a wide range of domains of life—from developing a language and conceptual framework for talking about the myriad shades of influence to which we are all subject daily.

However, even if the line between persuasion and coercion *were* clear, it is still not obvious that we would want to exclude coercive technologies from our consideration. The normative goal of

ensuring that a PT functions in a noncoercive manner—assuming that such a position is desirable or even coherent—does not require that we exclude coercion at the first point of categorization. Indeed, doing so may even make us *less* aware of potentially coercive mechanisms, and thus less likely to give them our attention—and this could ultimately produce a *more* coercive technological environment than we would otherwise have.

Note that I am not suggesting here that *designers* should ignore or minimize the question of coercion whatsoever. (In fact, below I offer suggestions as to how they can take it into account more fully.) I am instead proposing that when it comes to the *analysis* of persuasive technologies—when we ask *what counts* as a PT— there is no good reason to omit technologies that employ coercion (however clearly we might be able to define the term) from our consideration. We in fact hinder progress toward the design goal of minimizing coercion if we banish it from the umbrella of PT, because we lose opportunities for showing designers what *not* to do.

IV. A Search for Defensible Metaphors

The final step in refining our understanding of what counts as a persuasive technology is to push at the edges of the metaphor of persuasion to determine whether or not it is the most appropriate way to frame the types of technologies we are concerned with here. Gordon Pask once referred to cybernetics as the 'search for defensible metaphors' (Von Foerster 2003), and in a sense our task here is the same: we are ultimately interested in identifying a linguistic way of carving up these technologies that is, if not 'true,' at least reasonably defensible.

A fruitful first step may be to question why 'persuasion' as a metaphor for technologically induced behavior and attitude change has caught on as it has. As a framework for this questioning, the

'tetrad' approach given by McLuhan (1988) seems a good candidate mechanism. A quartet of questions developed to navigate through the effects of media, the tetrad is interesting because of its inclusion of both figure and ground. For a given object of analysis, the tetrad asks, (1) What does it enhance? (2) What does it 'obsolesce' (i.e. push into the background)? (3) What does it retrieve (i.e. what previous state of affairs does it recall)? (4) What does it reverse into (i.e. what does it become when pushed to an extreme)?

To take the first question, that of enhancement: as mentioned above, 'persuasion' foregrounds concerns related to intention and freedom. However, it also suggests the rational consideration of 'arguments' (*ethos* and *pathos* notwithstanding), emphasizes the role of individual persuaders, and brings focus to individual outcomes (e.g. by suggesting a duel between two opposite arguments).

The second question asks what the metaphor of persuasion obsolesces, or pushes into the background. For one, it obscures the full bidirectionality of interaction in a persuasive encounter; indeed, there always seem to be effects on the user that go unmeasured in PTs. This fact relates not only to the question of context-awareness or formal cause, as mentioned above, but also to broader ethical questions relating to the final causes of fulfillment and wellbeing. In addition, the fact that persuasion has historically been understood to take place via language pushes questions of force, touch, contact, etc. into the background. (Perhaps it is a foregrounding of this awareness that partially explains the recent popularity of the 'nudge' metaphor.) Also in the background with the 'persuasion' metaphor are considerations relating to the persuasiveness of the overall environment (as opposed to that of one individual actor), as well as a consideration of multiple persuasive outcomes, whether real or potential (i.e. PT ultimately doesn't just enter a 'debate' between *do* and *don't*—it coordinates the battle between all the possible things you might do or think).

The final two questions in the tetrad ask what the metaphor of persuasion 'retrieves' (i.e. what previous state of affairs it recalls) and what it 'reverses into' (i.e. what does it become when pushed to an extreme). As for the first question, what the metaphor of persuasion 'retrieves,' all I can conjecture is that in the context of our technological environment, perhaps it activates some latent mythic sensibility that inclines us to interpret environmental forces as generally intentional, or perhaps even benevolent and god-like—which, if true, may resonate with several things: the way we psychologically relate to brands as though they were people, our predisposition to categorize technologies with respect to brand perception rather than an abstract analysis of their underlying features, or e.g. the rise of descriptions of persuasive technologies as being 'smart' (Lobo et al. 2009). As for what persuasion 'reverses into,' when pushed to its limits it seems that it would consist of various methods of indirect coercion, so to speak: e.g. brainwashing, manipulation, or propaganda.

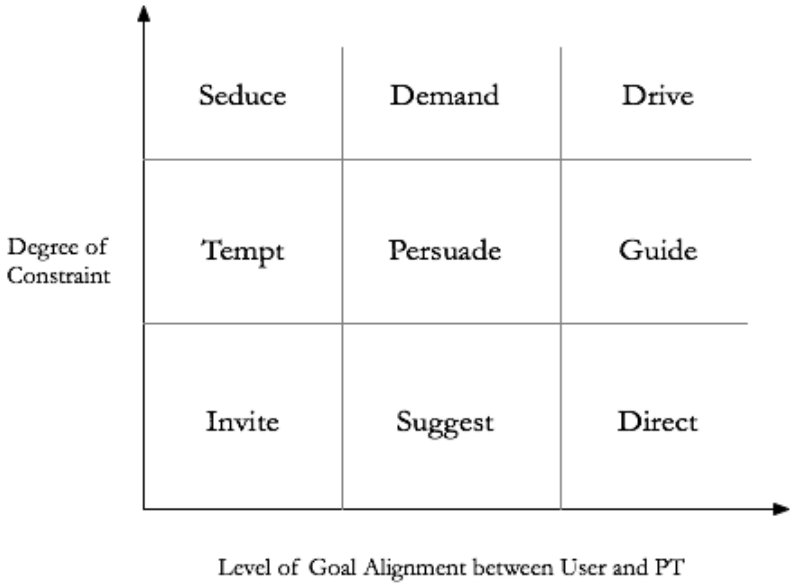
But the metaphorical options are literally endless, and at the end of the day no 'right' metaphor is waiting to be dug up. Instead, the question we must answer is which one seems true enough to the reality of the dynamics in play that it reveals the right things, and does not obscure important things.

What are the right things? This will be the topic of future work, but for now I want to suggest that two key normative concerns that have recurred throughout this analysis could serve as two starting points for mapping out a broader language of PTs.

> **Freedom** — To what degree does the PT constrain possible user actions?
>
> **Goals/intent** — To what degree do the goals of the system align with the goals of the user?

As mentioned above, the right metaphors need to be able to talk across the *spectrum* of both of these considerations.

If we take these two core considerations and plot them on axes, then each area of the resulting chart reveals an opportunity for different words or metaphors that could be used to describe a particular PT:



Using this linguistic framework, then, a 'Seductive Technology' would be one that has a low level of goal alignment with the user and places a high degree of constraint on them (e.g. an addictive game that a user wants to stop and regrets playing afterward). An 'Invitational Technology,' meanwhile, would have a lower degree of constraint. Similarly, we might say that a 'Guidance Technology' is one with a high degree of alignment with user goals and imposes a medium level of constraints (e.g. a GPS device).

These particular metaphors of course suggest various aspects of persuasion beyond the two axes listed here (e.g. whether the persuasion occurs via rational vs. nonrational pathways)—and more work is necessary to clarify how we would want to operationalize concepts like 'degree of constraint' and 'goal alignment'—but this framework serves as a useful starting point for taking the next steps toward that sort of linguistic mapping.

V. Conclusion

My goal here has been to clarify what counts as a 'persuasive technology,' primarily as a first step toward subsequent ethical analysis. The revisions provided above allow us to offer a new definition of a 'persuasive system':

Any system that sets and strives toward a goal that:

- consists of altering a user's relation to his or her goals
- shapes the system constraints toward that end
- matches the goal specificity with a specificity of the measurement of outcome

Most systems that satisfy these conditions will be socio-technical systems (e.g. systems where goals are represented in the human domain). However, it is possible for a wholly technological system to satisfy them on its own. Consider, for example, a software program that is able to define new sub-goals that inform its constraints: we could truly call such a system a 'persuasive technology,' because all conditions would be met by the technology itself. We could make a further distinction for systems that create new sub-goals in response to feedback about the outcome of a previous persuasive attempt; we could call such a system an 'adaptive persuasive technology.'

With this revised definition in hand, analysis of persuasive technologies can proceed on a much more stable foundation.

Chapter 2:

Autonomy and Persuasive Technology: Reactions and Distractions

Persuasive technologies are often criticized because they undermine users' autonomy, especially when they exploit irrational psychological biases. In this chapter I argue for two points. First, many objections that *appear* to be about autonomy are in fact primarily about *dignity,* and originate not in facts about manipulation of choice but rather in irrational and defensive psychological responses of *reactance* (more on this technical term presently). Clearly distinguishing between the two carries large implications for ethical analysis of technology generally. Second, questions of both autonomy and dignity in the context of persuasive technology can be greatly clarified if viewed as questions of *attention*. I outline a framework for such a view, within which an expanded concept of *distraction* emerges as a useful construct for analyzing potentially problematic persuasive mechanisms. In order to typologize these persuasive mechanisms and establish a common language for describing them, I introduce a framework consisting of three main types of distraction, which I call Functional, Existential, and Epistemic Distractions.

I. Common Autonomy-Based Criticisms of Persuasive Technology

Persuasive technologies are often criticized because they undermine users' autonomy, especially when they exploit irrational psychological biases (Smids 2012, Spahn 2011, Oinas-Kukkonen 2010, Verbeek 2009). Such biases include loss aversion (e.g. 'fear of missing out'), social comparison, the status quo bias, framing effects, and countless others (Kahneman 2012). A burgeoning industry of researchers and consultants exists to help designers create technologies that better target and exploit these psychological vulnerabilities (Fogg 2003, Eyal 2014, Parr 2015).

Consider the following three examples of persuasive technologies that operate by appealing to our irrational biases, along with brief descriptions of their effects. I will refer back to them periodically throughout this analysis.

1. Notifications — Notifications from systems such as email services, social networks, and mobile applications are widely encountered today. Each day, the Android mobile operating system alone sends over 11 billion notifications to its more than one billion users, each of whom check their phones an average of 150 times per day.[5] Often, as in Google's Gmail system, notifications are colored red and placed in the upper-right corner of the user's vision in order to better grab their attention and maximize the persuasive effect. This effect relies on the human reaction to the color red (Maier et al. 2009), as well as the cleaning/grooming instinct (Curtis 2007), which often makes it hard to resist clicking on the notifications. Users often find notifications distracting, and when a person is in a focus state and is interrupted, it takes on average twenty-three minutes for them to regain their focus. Furthermore, exposure to repeated notifications creates mental habits that train users to interrupt themselves, even in the absence of technological influences (Mark et al. 2008). The implications of distraction for wellbeing are often overlooked due to the bite-size nature of its influence; however, as Crawford (2015) writes, 'Distractibility might be regarded as the mental equivalent of obesity.'
2. Intermittent variable rewards — Randomizing the reward schedule for a given action increases the number of times a person is likely to take that action (Ferster and Skinner 1957). In his book 'Hooked: How to Design Habit-Forming Products,' Eyal (2013) identifies variable rewards as a crucial step in the process of creating habits and 'hooking'

[5] At the time of writing, Android is the most common 'smartphone' operating system, with 77% of the global market share. Apple, Inc.'s iOS comes in second, at 20%. (Llamas et al. 2014)

users on a product. This is the underlying mechanism at work behind 'infinite' scrolling feeds of information, such as those employed by Twitter, Facebook, Pinterest, and countless other news, social, or entertainment websites. It is often referred to as the 'slot machine' effect because it is the foundational psychological mechanism on which the machine gambling industry relies (and which generates for them over $1 billion in revenue every day in the United States alone) (Rivlin 2007). The addictive effects that these devices are designed to have can often be severely debilitating for users (Schüll 2014).

3. Facebook and 'emotional contagion' — The Facebook 'News Feed' is a filtered view of posts made by one's 'friends,' or reciprocal connections. In a widely discussed study, researchers at Facebook and Cornell (Kramer et al. 2014) carried out experiments using the Facebook News Feed to identify evidence of social contagion effects (i.e. transference of emotional valence). Over a one-week period, the experiment reduced the number of either positive or negative posts that a sample (n=689,003) of Facebook users saw in their news feed. They found that when users saw fewer negative posts, their own posts had a lower percentage of words that were negative. The same was true for positive posts and positive words. While the effect sizes were small (-0.19% for negative words and -0.07% for positive words), the results show a clear persuasive effect on the emotional content of users' posts. The study was widely publicized and some objected to it on the grounds that it was manipulative.

I base this analysis on these three examples in particular because they represent three of the most common concerns about persuasive technologies and autonomy — (1) distraction, (2) addiction, and (3) manipulation, respectively — that that could be seen as among the most invasive or aggressive in terms of their potential impact on autonomy due to their exploitation of non-rational psychological biases. These three concerns are by no means the only autonomy-related

considerations relevant to persuasive technologies; there are, for example, a wide range of mechanisms that may straightforwardly be described as 'coercion.' Smids (2012), for example, describes how the *persistence* of a car's incessant beeping can function as a type of coercion when it finally wears down its driver's capacity for self-control and forces him to fasten his seatbelt.[6] This example is particularly useful because it reminds us that, while more clearly coercive mechanisms are more likely to be apparent to the user *as* coercion when they occur, they may still influence via indirect means.

It is also important to clarify how I view the interrelation of these three types of influence. I consider them as distinct, but not mutually exclusive, types of persuasion. Each term emphasizes a different set of features associated with the persuasive exchange: distraction emphasizes the relation between the outcome and the user's goals; addiction emphasizes the habitual nature of usage in the context of the user's life as a whole; and manipulation, while admittedly a broader term with a wider range of potentially defensible definitions, emphasizes the selective presentation of information in a way that diminishes the user's freedom (Sripada 2011). In the specific persuasive technologies I have chosen to use as examples here, these three types of persuasion largely exist in independence from one another. However, in many cases they can, and will, coexist: an addictive technology can also be a distraction, distractions can also be addictive, and either (or both) of them can exist as an instance of manipulation. But any of the three can also occur in the absence of the others.

[6] In addition, 'autonomy can be impaired by persistent *desires* in cases which do not involve addictive substances' (Levy 2007, emphasis mine)

II. Reactions: Separating Real from Apparent Autonomy Criticisms

Some concerns about such persuasive technologies that *appear* to be about autonomy are, in reality, about *dignity*. It is, as I will show, extremely important to distinguish between the two. We are most likely to see these dignity-based objections made in the context of concerns about manipulation, though they may also exist in response to addiction, distraction, or other types of technological influence such as coercion or temptation.

'We Are All Lab Rats'

Consider the example of the Facebook 'News Feed' study mentioned in section #3 above, which generated an unexpected outcry about its methods. Some objections focused on ethical questions about informed consent, the standards for which are still open for conversation insofar as they pertain to large-scale research carried out by online service providers.[7] However, many of the objections were about the mere fact that Facebook had manipulated its users in the first place. Clay Johnson, founder of political marketing firm Blue State Digital, wrote (apparently unironically) that "the Facebook 'transmission of anger' experiment is terrifying." *The Atlantic* described the study as 'Facebook's Secret Mood Manipulation Experiment' (Meyer 2014). A member of the UK parliament 'called for a parliamentary investigation into how Facebook and other social networks manipulated emotional and psychological responses of users by editing information supplied to them' (Booth 2014). And privacy activist Lauren Weinstein took to the 'microblogging' platform Twitter to write, 'I wonder if Facebook KILLED anyone with their emotion manipulation stunt. At their scale and with depressed people out there, it's possible.' (qtd. in Goel 2014).

[7] One month after the study's publication, the Editor-in-Chief of *PNAS* published an 'Editorial Expression of Concern' in response to these objections (Verma 2014).

Manipulation is widely seen to pose threats to autonomy not only in the context of persuasive technologies but also in wider discussions about 'big data' (Schroeder & Cowls 2014), decision architecture (Sunstein 2015), and human interaction in general (Kane, 1996, Pereboom, 2001). While there is broad disagreement about what the criteria for manipulation are, the definition given by Pennock (1972) captures its most common elements: it is influencing a person 'by controlling the content and supply of information' in such a way that the person is not aware of the manipulation (Powers 2007, Strauss 1991, Kane 1996, Pereboom 2001, Smids 2012).[8]

However, what does seem curiously consistent is the way in which individual responses to manipulation — whether real or imagined — are described. These descriptions lead to the true roots of many of these objections, which lie in questions of *dignity*. If you read definitions or descriptions of manipulation, you will find the phrase 'treated like an X' recurring, where X is something other than 'adult human.' For Berlin (2002), manipulation is being treated 'as if I were a thing, or an animal…' For Wilkinson (2013), it is to be treated as 'puppets' or 'tools and fools.' In covering the brouhaha over the Facebook experiment, no less a restrained voice than the *New York Times* ran with the lede, 'To Facebook, we are all lab rats' (Goel 2014).

But the most common variant seems to be 'like a child.'[9] This metaphor is of course present in the term 'paternalism,' along with its pejorative variants (e.g. 'nanny state'). It is perhaps worth asking: what is particularly bad about being treated like a child? To have others giving attention to our needs, wants, and best interests; to have others willing to protect us from unwelcome influences; to give us love, care, and devotion; to even put our needs above their own — what qualms could one possibly have with such a thing? If you heard a friend complaining about a 'paternalistic'

[8] In an empirical investigation of people's intuitive judgments about manipulation, Sripada (2011) found not only 'corrupted information' but also 'deep self discordance' to be of particular importance.
[9] cf. Raz (1986), p. 422.

technology, and it was the first time you had ever heard the term, it would not be totally unreasonable to assume that their ultimate grievance was that the technology was only *partially* paternal — or perhaps that it was paternal in the wrong ways, or paternal like a bad parent is paternal.

The reality, I believe, is that the particular nature of the metaphor — whether child, puppet, lab rat, tool, or fool — is of little consequence. The important point is that it represents a way in which one does not wish to see oneself — i.e. as someone with diminished freedom — and is thus seen as *insulting*.

This creates a psychological response known as 'reactance.' Recall the famous scene from the film *Taxi Driver* when Robert DeNiro's character, Travis Bickle, says to an imagined interlocutor in the mirror, 'You talkin' to me?' This is an example of reactance. Reactance is a defensive response directed at an agent perceived to be depriving a person of their freedom. Substantial evidence from moral psychology illustrates a range of irrational biases that inform our judgments via System 1 processes, of which reactance is but one (Haidt 2003, Greene 2014), some of which influence our judgments about freedom and autonomy in particular (Phillips & Knobe 2009). This effect is not unique to manipulation; Cass Sunstein points out that even if people 'are constantly reminded that a due date is coming, they might feel as if they are being treated like children' (Sunstein 2015).

Why do the responses to concerns about manipulation that I have mentioned here seem to be cases of reactance? For one, the objections are often not based in the facts about possible versus actual choices made. They rarely take the format, 'P manipulated me into choosing X, but I would rather have chosen Y.' When they do, even if reactance occurs, there are of course serious

questions about autonomy.[10] But most of the time, they take the form of hypothetical or imagined situations: 'P would manipulate me into choosing X, but I won't be treated like a child!' Of course, it is certainly possible that there may be cases where manipulation has not yet occurred, but where there is good reason for thinking it *might* occur if it is not objected to beforehand. In those cases, reactance-based responses can serve a purpose and help mitigate future manipulation. But even in cases where manipulation is likely to occur, such a response is not necessarily justified. Returning to the example of the Facebook study, while the effects were statistically significant, they were nonetheless extraordinarily small. The interventions only decreased the frequency of negative words in users' posts by two out of every one thousand, and of positive words by half a word out of every one thousand (Kramer et al. 2014). Compared with other influences on the emotional valence of people's communications—many of which also occur by design—such an effect seems scarcely worth caring about. This is not to say that it is not important *at all*—only that there are other whales to fry before we start cooking up the minnows.

Another reason such responses seem to be rooted in reactance is that they are highly malleable. Schechter and Bravo-Lillo (2014) carried out a study to assess people's judgments about how acceptable they found several real-life experiments, including the Facebook experiment discussed above. However, only some of their respondents received questions about the Facebook study based on the way it was actually carried out, while other respondents received variants of the question with one element or another altered. The variations included replacing the name 'Facebook' with 'a social network' or 'Twitter,' adding text saying that Facebook had promised not to use the findings to inform the design of their advertising, or changing the described

[10] Even if the objections are based in fact, they may still be based on a partial view of the full set of influences actually affecting one's life. Attributable influences may not be biggest influences, nor the ones most deserving of our critical attention. (Consider the question of whether humans ought to intervene in nature to prevent animal suffering, which is where most of it happens.)

methodology from 'removing' posts on their News Feed to 'inserting' posts. Several interesting results emerged:

- Respondents who received the unaltered description of the Facebook experiment were nearly twice as likely to say it should not be allowed to proceed (46% vs. 24%) if they had already been aware of the experiment.
- Among respondents who were unaware of the Facebook experiment:
    - A full one-third *more* people (32%) said the experiment should *not* be allowed to proceed if Facebook were to 'promise not to use it for advertising'
    - Fewer people (19%) said the study should not be allowed to proceed if 'Facebook' was replaced with 'a social network' in their question
    - Only 17% said the study should should not be allowed to proceed if Facebook were to 'insert posts instead of hiding' them (and only 14% said 'no' if the posts were inserted rather than removed, *and* they were all positive)

These results show the significant effects of outside influences, the role of the Facebook brand name in their decision-making, and the way in which users largely perceive their information, rather than their attention, to be the thing managed (e.g. assuming one were to read a finite number of posts, whether a post is inserted or removed would make no difference). Perhaps the most interesting result here is that people were *less* likely to deem the study acceptable if Facebook promised not to use it to inform the design of their advertising products. One could view this as the result of priming effects reminding them (or informing them, if they were not already aware) that Facebook is an advertising company, which could put them on the defensive against possible persuasion or manipulation. (It could also indicate trust effects; it is unfortunate that the researchers did not include a variant of the 'no advertising' condition with Facebook's brand name removed.)

Furthermore, manipulation is actually much more likely to be happening when there is a perception that one's freedoms are being *increased* rather than *decreased*. For example, when asking someone to do something for you, if you add to your request the phrase, 'but you are free to accept or refuse,' then they are twice as likely to comply (Carpenter 2013). This effect is seen in e-mail interactions as well as in-person ones (Eyal 2014). In the realm of persuasive technologies, an example of this is the skippable 'TrueView' video advertisements on YouTube. Previously, users were required to watch a full advertisement in order to access monetized video they were trying to see, but when YouTube began allowing them to skip the ad after a few seconds, it resulted in 388% *higher* engagement, an increased emotional response within the first five seconds, increased brand favorability, and increased purchase intent (Ipsos/Innerscope 2011). These cases are particularly interesting—and scary—because it is the very act of affirming someone's autonomy that may serve to deprive them of it. Such misplaced trust is of even greater ethical consequence in an era when we have left behind many of our previous cultural commitment devices that have historically helped us protect our autonomy.

These counterintuitive dynamics have parallels in the psychology of attention as well. Kuhl and Chun (2014) write that 'perceptual learning may actually be *greater* when unattended information is weak (e.g. weakly coherent motion of visual stimuli) relative to when it is strong. Putatively, when unattended information is strong, attentional mechanisms detect and successfully suppress this distraction; but when it is weak, suppression is not elicited and learning occurs.' In other words, when information is perceived as distracting or annoying, it may in fact have *less* influence on a person's future attention or actions than when it pings their attention more subtly and unconsciously. The indirect route of persuasion is often the most effective.

As the setting of Travis Bickle's famous mirror monologue in *Taxi Driver* suggests, reactance relies on reflection. At its core, the effect comes not from recognizing oneself as having been manipulated, but in *imagining oneself as being manipulatable*. A person looks in the mirror of technology and sees a reflection of himself that he does not recognize, and this undermines the integrity of his own self-image. 'Nothing is at last sacred,' wrote Emerson, 'but the integrity of your own mind.' Being forced to take a third-party perspective of yourself that you do not recognize violates that integrity by creating a dissociation, or a cognitive dissonance, that is perhaps akin to the 'uncanny valley' effect in robotics (Mori 1970).

The mirror of technology gives us back both *dignified* and *undignified* reflections. The former align with our projects of self-authorship, and the latter do not. Dignified reflections show us the goals, preferences, and values that we identify with, whereas undignified reflections show us a distorted self, as though looking in a funhouse mirror. The difference is that when we are looking in a funhouse mirror, we know it. With technology, we often forget the mirror is there at all. (Perhaps an even deeper threat to dignity lies in the possibility of not even recognizing one's reflection *as a reflection*, which is of ever greater concern the more rapidly our habits of technology use evolve.[11])

The Monster and the Bank

Of course, it could be asked why is this is a problem worth mentioning at all. What does it matter if a concern about autonomy is ultimately about dignity, or if it has its roots in a response of reactance? The answer is not because autonomy is more important than dignity. (While this is something I happen to believe, it is unnecessary for my argument here.) Rather, there are four reasons why failing to distinguish between illusory and real autonomy concerns is important.

[11] cf. 'Narcissus as Narcosis,' in McLuhan (1964)

First is the fact that it is simply not a rational way to carry out ethical analysis. This is not to say that irrational biases are inherently problematic for ethical analysis—indeed, as in human cognition broadly speaking, they are unavoidable and can often be extremely helpful—but rather that there is not necessarily a correlation between the ethical considerations that are psychologically most rewarding and the ethical considerations that are most deserving of our attention.

Second, our attention is finite, and giving attention to one ethical concern means that we must withhold it from another. The opportunity cost of addressing issues of dignity *as issues of autonomy* is that we are less likely to give attention to *actual* issues of autonomy. When we add to this the knowledge that, according to Brehm (1966), 'the more important that free behavior is to the individual, the greater will be the magnitude of reactance,' it means that those most concerned about freedom and autonomy are most likely to be distracted by reactance.

Third, to the degree that the designs of persuasive technologies spark, nudge, or amplify impulsivity rather than an even-handed consideration of ethical concerns, they may actually contribute to the degree of reactance-based responses. This is particularly the case when we consider the rise of outrage cascades on Internet communications platforms. In such cases, objections to the system dynamics could be viewed as products of the system dynamics themselves. (I am reminded of McLuhan's observation that the only way to critique a medium from within that same medium is via parody.)

Finally, a vigilance in our delineation of such reactance-based objections—which operate via our need to *blame* someone—helps us remember that, in many situations (especially in the realm of technology design), *there is no one to blame*. At 'fault' are more often the emergent dynamics of complex multi-agent systems rather than the internal decision-making dynamics of a single

individual. John Steinbeck captured well the frustration we feel when our moral psychology collides with the hard truth of this reality in *The Grapes of Wrath*, when tenant farmers are evicted by representatives of the bank:

> "Sure, cried the tenant men, but it's our land...We were born on it, and we got killed on it, died on it. Even if it's no good, it's still ours...That's what makes ownership, not a paper with numbers on it."
>
> "We're sorry. It's not us. It's the monster. The bank isn't like a man."
>
> "Yes, but the bank is only made of men."
>
> "No, you're wrong there—quite wrong there. The bank is something else than men. It happens that every man in a bank hates what the bank does, and yet the bank does it. The bank is something more than men, I tell you. It's the monster. Men made it, but they can't control it."

The bank isn't like a man, nor is the technology company, nor is any other brand nor signifier that we might use to represent the boundary conditions of the technologies that shape our lives. *There is no one to blame.* Knowing this, however, presents us with two paths to choose from. Do we conjure up an image of a 'monster' at whom to direct our blame, and take a path which, while psychologically rewarding (and sometimes even necessary), is likely to distract from the goal of enacting real change in the real world? Or do we take the second path, and look head-on at the true nature of the system, messy and psychologically indigestible as it seems to be?

The first path would seem to lead us toward a kind of digital mythology, in which we engage in imagined relationships with personified dynamics of our informational environment, much as the

ancients did with their physical and emotional environments.[12] Yet if we take autonomy seriously, we cannot help but note that in Steinbeck's example it is not the displaced farmers, but rather the bankers, who invoke the idea and, we might say, the *brand* of the 'monster.' Similarly, in the realm of digital technology, it is less often users than companies who produce the representations that serve as the primary psychological and emotional points of connection. In fact, these brands and representations may be the elements of technology design over which users have the least amount of control of all. What this path would entail, then, is acquiescence to a mythology that, while psychologically satisfying, would be (and in many cases already is) even more engineered than the products they represent, or than the decisions that those products are designed to induce.

The second path would entail looking the 'monster' in the eye, and seeing it for the complex and multifaceted environment that it is. Such an approach would be akin to what Floridi (2013) has called 'infraethics,' i.e. attention to the infrastructural, 'first-order framework of implicit expectations, attitudes, and practices that *can* facilitate and promote morally good decisions and actions,' In a sense, the perspective of infraethics views society itself as a sort of persuasive technology, with a persuasive goal of maximizing moral actions.

Perhaps blazing a third way—a zigzag of trails between the two paths that acknowledges the infrastructural realities of the ethical situation, yet enables us to engage with it on human terms—is ultimately in order. If so, my intuition is that it would require us to take more seriously, and design more consciously, the underlying narrativity of our lives.

[12] Research on Facebook users' perceptions of the algorithmic nature of the News Feed has noted some respondents' tendency to view it as 'a river or a force of nature … rather than a road or some kind of infrastructure that can be altered by human intervention' (Rader and Gray 2015).

III. Attention as a Common Framework for Autonomy and Dignity

In order to assess the implications of persuasive technologies for issues that actually *do* concern autonomy, we must have some conception of what autonomy is. This presents a challenge, because many different versions of the concept exist. Some conceive of autonomy in terms of reasons (à la Kant), others in terms of an individual's motivational hierarchy (e.g. Frankfurt, Dworkin), and still others in terms of the concept of 'self-authorship' (Raz 1986).[13] Beyond these three groupings, still others exist that are not easily categorizable. One wonders whether, given the widespread disagreement about even the most fundamental aspects of autonomy, it actually functions less as an organizing principle than as a magnetic north for a cluster of loosely-related considerations.

One approach to routing around the conceptual traffic jam that is autonomy would be to identify one or more criteria on which most theories agree and then assess how particular cases of persuasive technology infringe upon those specific criteria. While this is an option, and is indeed common in applied ethics (Schaefer et al. 2013), it would by its very nature be a fragmented view. Such a fragmented approach would, I believe, fail to capture what seems intuitively most valuable about autonomy in the context of persuasive technologies. Furthermore, there is no assurance that the overlap points would ultimately represent the necessary, and not merely the sufficient, conditions of autonomy.

The other approach would be to simply select a widely used theory of autonomy that seems appropriate for my purposes and run with it. This is also an unsatisfying option, as it would only speak to the particular considerations of that one theory, and would thus be of limited wider value.

[13] As far as off-the-shelf conceptions of autonomy go, Raz's is the most enticing, as it incorporates many aspects of the others. (The notion of 'self-authorship' is also intuitive, although Raz seems to avoid the most interesting implications of this metaphor.)

Is there a meaningful third option here? Is there another way we can frame the motivating question in the particular context of persuasive technology such that issues commonly associated with autonomy—and potentially even dignity, as discussed above—are clarified, and problematic or tangential criteria withdraw into the background? With autonomy it seems as though the word often precedes the concepts. So perhaps we could begin by returning to where the word began. Since Kant, the term has accumulated such psychological force that we feel it must be *somehow* important—in the same way that we assume an old, distinguished-looking guest at a social event must be *someone* important—even if we do not exactly understand who they are.

The origins of the word 'autonomy' lie in the idea of 'governing oneself,' and the term was originally applied to political groups before it was applied to individuals. The core of the idea was that undue 'claims or demands' should not be made on a particular group, and then later, on a particular person (Darwall 2006). What sort of 'claims or demands' do persuasive technologies make on their users? Another way of asking this question is: what aspects of their users do persuasive technologies find most valuable? One clear answer, it seems to me, is the resource that most of the persuasive technologies we encounter on a day-to-day basis relentlessly optimize and compete for: our attention.

Attention

Herbert Simon's observation that information abundance results in attention scarcity came over forty years ago. Since that time, the advance of digital technologies has produced in an environment where we are now barraged with information from every angle. Human attention is now the resource that most of our digital technologies compete for by design, though many users still have not come to terms with this new paradigm, in which 'the user is the product.'

Attention is like autonomy in that no one seems quite sure what it actually is. Remarkably, this disagreement occurs not only where we might expect to find it—i.e. across the centuries, or across diverse domains of modern-day society—but even within the highly specialized psychological literature on the topic that has been published in the past decade.

In 1890, William James in *The Principles of Psychology* articulated something akin to a definition of attention. However, one wonders if he was perhaps a little too keenly aware of the gap between his goal and his attempt, because he felt the need to preface the definition with the sentence, 'Everyone knows what attention is.' He continues: 'It is the taking possession by the mind in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought...It implies withdrawal from some things in order to deal effectively with others, and is a condition which has a real opposite in the confused, dazed, scatterbrained state.' While several aspects of James's characterization of attention are relevant here and invite comment, I want to note two shifts in particular—shifts in his attention, we might say—that take place over the course of this definition. The first shift is seen in the way he pivots from thinking of attention as a general capacity at first to later treating it as the 'effective' application of that capacity.[14] But what is especially interesting here is how, in the particular cadence of James's definition, this pivot seems to occur in synchrony with his second shift: from defining attention by what it *is* to defining it by what it is *not*. Indeed, his verbs alone seem to suggest a stair-stepping retreat away from the head-on certainty with which he begins (i.e. 'Everyone knows…') and toward an attempt at some negative definition ('it is' becomes 'it implies,' which then becomes it 'has a real opposite'[15]).

[14] These two senses of attention mirror the two senses of 'distraction' discussed below: the broader sense, which refers to its mechanism, and the narrower sense, which refers to its outcome.
[15] Attention's 'opposite' is not distraction, however, as I will discuss below.

What becomes apparent when reading others' attempts at defining attention—or even just their attempts at securing some kind of conceptual foothold—is the way, like James, they seem to recognize the need to approach the topic from several different angles, and make several different attempts, just to begin to understand its dimensions, yet then are ultimately clearest when they define attention by what it is not.[16] In his book *The Attention Complex*, Rogers (2014) provides a wide-ranging survey of the history and evolution of the concept of attention that embodies this method of triangulation. As his analysis winds to a close, these different angles spiral inward with increasing speed such that, by the book's final page, in a breathless, nearly ecstatic prose, he variously describes attention as 'a transactional reality,' 'a set of theoretically polyvalent concepts that operate in the sphere of practical reason,' and 'a transduction zone between the self and external forms of power.' But, like James, he is ultimately clearest when he describes what attention is *not*: it is 'not a state or a system.' Similarly: 'There are no theories of attention, only practical applications of attention as a means of regulating individuals and helping the individuals regulate themselves.' One could replace 'attention' in the preceding sentence with 'autonomy' and it would not be wholly indefensible.

In the recent psychological literature, there is similarly little consensus about how attention ought to be defined. The epilogue to *The Oxford Handbook of Attention* observes that "getting from notions to definitions [of attention] has been problematic' … 'it must be admitted that the field of attention could do better in terms of providing explicit and consistent definitions of its topic . . . Across the literature, 'attention' can indicate prioritization, target selection, mental effort, a mental state, the availability of resources, executive control functions, awareness, or simply 'thinking"'(Nobre & Kastner 2014). Yet psychological definitions of attention *do* converge on two key points. The first is that attention is a way of describing information processing. The second is that attention

[16] In his study of Kant's conceptions of attention and distraction, Gasche (2008) concludes that 'becoming distracted is … the only way of paying attention to the phenomenon of attention.'

operates in a fundamentally goal-salient manner. 'If one were to distil a core definition of attention out of the contemporary literature, it would be something like: the prioritization of processing information that is relevant to current task goals.'

Three key distinctions about attention emerge from the psychology literature that are extremely useful for an ethical analysis of persuasive technology.

1. Top-down versus bottom-up attention — This distinction refers to the direction of information processing. 'Top-down' (System 2) attention involves the executive control function, which is rational and goal-oriented, whereas 'bottom-up' (System 1) attention involves impulsive and irrational processing. This is a particularly important distinction for many conceptions of autonomy because a person's top-down attention is generally seen as involving the processes that serve the considered goals of their planning self.
2. Perceptual versus reflective attention — This distinction refers to whether the information attended to is external (e.g. sensory stimuli) or internal (e.g. goals or memories). This is an important distinction because it emphasizes the informational nature of internal psychological constructs, reminding us that external information not only competes for our attention against other external information, but also against other internal information as well. (This will have significant implications for assessing the potential harms of distraction.)
3. Shifts in attention versus ongoing information processing — This is an important distinction because it will let us distinguish between *acute* and *chronic* cases of distraction, which may differ not only in their implications for a user but also in the types of interventions that may be necessary to mitigate their harmful effects.

The Centrality of Goals

As the criterion that attentional mechanisms use to prioritize information, goal-relevance is important for many existing conceptions of autonomy. Austin & Vancouver (1996) define a goal as 'the internal representation of a desired state.' A substantial literature on the psychology of human goals exists, and the topic is an area of active research (Locke & Latham 2002). Of immediate importance, however, is research on the psychology of goals that has shown that they exist in a hierarchical network structure.[17] In such a structure, immediate goals (i.e. tasks) inhabit the lowest level of the hierarchy and the goals that they serve inhabit higher levels. (The methodology that Bagozzi et al. (2003) used to identify this structure during their work with the Italian military—a repeated series of 'why' questions in response to existing information about their goals—will later prove instructive for the implications of this analysis for design processes.) What emerges from this picture is a construct of the 'goal' that can have wide applicability as a general-purpose unit of analysis of the human will. Bagozzi et al. even recognize this potential when they write:

> "The highest levels of our representation of superordinate motives correspond to Carver and Scheier's system concepts (i.e. the ideal self). The ideal self might be comprised of a small number of such higher order motives as fulfillment, economic independence, and being a person who does the right thing" (Bagozzi et al. 2003)

One is reminded here of Kierkegaard's phrase 'purity of heart is to will one thing.' Perhaps 'one' is overdoing it, but it is on the right track.

Using 'goals' in this broader sense has a significant implication for our definition of 'attention.' Taking into account *all* types of goals, from immediate tasks to higher-level goals, to life-level aspirations and even values, requires us to also expand our notion of 'attention' to cover both short

[17] Technically, the psychological structure of goals is more accurately described as a *heterarchical* network, since one sub-goal can serve multiple higher-level goals (cf. Bagozzi et al. 2003). However, because the distinction carries few, if any, implications for this analysis, I will use the more familiar term 'hierarchical.'

and long timeframes. We do not normally use the term 'attention' when speaking about the success or failure of our information processing with respect to our longer-term goals; we tend to use words like 'dedication' or 'commitment' instead. However, there is no reason why we should not think of attention as being operative on these longer time scales. We have already said that attention is a *process*—this just allows it to be a longer one.

In fact, I believe the concept of a 'goal' is elastic enough to have even wider applicability. I suggest that it could be extended to include, on one hand, actions to which our irrational, impulsive, System 1 selves are attracted (e.g. the 'goals' of your biological self)[18], as well as, at the other extreme, our highest-level metacognitive goals (i.e. goals about goals), which correspond to Frankfurt's 'second-order' desires. Henceforth, whenever I use the term 'goal,' unless specified otherwise it will be as a general-purpose umbrella term that encompasses these, as well as the above, descriptions.

In sum, what I am proposing here is that attention may serve as a useful ethical framework through which to analyze practical considerations about persuasive technologies that are normally explored in terms of autonomy and dignity. I suggest that such a framework would take as its scaffolding the psychology of human goals, as well as the three distinctions I have outlined here. In the same way that the construct of a 'goal' allows us to speak with one language across the entire hierarchy—integrating tasks, goals, values, and reflective capacities into one view—it can help us speak across different ethical considerations as well. In fact, this approach may also help unite considerations across ethical theories broadly speaking, because whether the aim is to maximize an action's alignment with rules (deontology), with outcomes (consequentialism), or with character

[18] cf. the notion of 'the proto self' advanced by Damasio (1999) and Churchland (2003), which refers to 'very basic bodily information, from various sources' that is 'integrated in the brain stem' (qtd. in Levy 2007), as well as the notion of 'picoeconomics' as advanced by Ainslie (1992), on which view the self may be viewed as the locus of competition among a variety of 'subpersonal interests'

(virtue ethics), in all cases that aim takes the form of some *goal*. And whenever an influence moves us away from a goal, we call that influence a *distraction*.

IV. Three Types of 'Distraction'

There is a broad sense in which the word 'distraction' can carry either positive or negative valence, or even none at all. In this expansive meaning, the term refers to the direction of one's attention away from *something*, where that something could be *anything*. In this usage, the term hews closely to its etymological root: *distrahere*, simply 'to pull apart' or 'to separate.' We routinely encounter the positive usage in colloquial language. Consider, for example, the phrase 'a welcome distraction.' As in: 'Josh's invitation to play ping-pong was a welcome distraction from answering my emails.' Here, if my authentic goal is actually to play ping-pong rather than answer my emails, then Josh's 'distraction' is a directing of my attention that prompts me to act in alignment with that goal and is thus a positive distraction.[19] This positive sense of 'distraction' can extend into the moral domain as well: e.g., 'Josh distracted the masked man from murdering his neighbor.' But if we were to change both of these examples so that the distractions were negative, the word 'distraction' would be no less appropriate. So in this general sense, the term 'distraction' derives its meaning from the nature of the mechanism rather than from the nature of the outcome.

I am not concerned here with that broad sense of 'distraction.' Rather, I am interested in a narrower sense—one that *does* take into account the outcomes of distraction, especially those that carry implications for a person's ability to make their life go well. In this sense, distraction is not just about the direction of attention, but also whether the object of attention matters to us. In

[19] Levy (2007) points out how, in the famous 'marshmallow test' carried out by Michel (1981) to study self-control in children, 'the ability to delay gratification [depended] crucially on self-*distraction*' (emphasis original).

other words, I am interested in the direction of one's attention away from information that is salient to one's authentic goals. (In this sense of the word, for example, you would never say, 'The fireman distracted me from my workout by telling me the building was on fire.')

Three Types of Distraction

I believe that we can expand the term 'distraction' to define a framework that is useful for analyzing and understanding concerns about persuasive technologies that address not only issues commonly associated with autonomy, but also issues of dignity, such as I described above. This is necessary because considerations of autonomy and dignity overlap, and so a conceptual pivoting may illuminate the relationship between the components of each. As in the previous section with 'attention,' this requires us to expand the term 'distraction' beyond its common use, which primarily refers to the redirection of one's moment-to-moment awareness.

I define three distinct types of distraction, which I will call Functional, Existential, and Epistemic distraction. Each type of distraction directs us from a different category of action—from doing, being, and knowing, respectively. Functional Distraction pertains to instrumental goals, Existential Distraction pertains to terminal (i.e. non-instrumental) goals that are pursued for their own intrinsic value, and Epistemic Distraction pertains to goals about goals (i.e. the maintenance and development of the reflective and deliberative capacities that enable us to set goals in the first place).

Although these categories of distraction are distinct, they are not mutually exclusive. In the same way that distraction, addiction, and manipulation may co-occur as described above, these three types of distraction may result in any combination from a particular persuasive technology.

Below I discuss each type of distraction in detail.

Functional Distraction

Functional Distraction is the direction of a person's attention away from information or actions relevant to the pursuit of their immediate tasks or goals. It is what is commonly meant by the word 'distraction' in day-to-day use. Functional distraction usually directs a person away from lower-level goals that are instrumental in nature and undertaken in support of some other higher-level goal. For example: 'I was going to turn on the kettle so I could make some tea, but Candy Crush reminded me that I haven't played in a few days.' In short, Functional Distractions make it harder to do what you want to do.[20]

Interruptions are the paradigmatic example of functional distractions. In the context of persuasive technologies, they often take the form of notifications, as described above. They could also take the form of person-to-person communications, such as receiving an instant message from a friend when you are trying to write an essay. (Whether this would count as a 'persuasive' technology would be a matter of definition; regardless, maximizing the amount of messages sent is certainly a conceivable design goal for such software.)

The effects of Functional Distraction are not limited to the opportunity costs of not pursuing one's desired goal during the moment of distraction itself. A functional distraction may make it harder for them to return to the previous object of their attention due to the effect of the 'inhibition of return' (Posner et al. 1985). In addition, as noted above, repeated functional distractions may create new habits that train us to interrupt ourselves (Mark et al. 2008), which may carry us into the realm of Epistemic Distraction, as discussed below.

[20] In defining the conceptual (and to some degree linguistic) features of these three types of distraction, I have drawn inspiration from the hierarchical conception of autonomy as advanced by Frankfurt (1988).

Existential Distraction

Existential Distraction is the direction of a person's attention away from information or actions that promote the pursuit of their ultimate values or 'being goals.' It results in a state of 'deep-self discordance' between a person's actions and identity (Cova et al. 2012), and ultimately makes it harder for them to 'be who they want to be.'

There are three ways Existential Distraction operates. The first is by directing the person toward a lower-level goal that has no 'line of sight' to any ultimate values (i.e. intrinsically valuable goals) that they want to maximize — or, vice versa, by directing them *away* from a goal that *does* have such a 'line of sight.' An example of this would be supporting a person's pursuit of 'avoidance goals,' i.e. goals that do not have a link to any ultimate values that reflect the person's desired identity. Avoidance goals have been associated with lower wellbeing (Elliot et al. 1997, Ryan & Deci 2001). For example, a persuasive technology may only know what a user's intentions are within a particular task domain, such as the context of Facebook, and may be ignorant about the way in which that technology fits into their life in a broader sense. As a result, if a user has an avoidance goal to spend more time using Facebook instead of doing homework, then when the technology deploys a persuasive mechanism such as intermittent variable rewards in the News Feed to maximize that behavior, it is not just a functional distraction, but an existential one.

The second way Existential Distraction operates is by *removing* the 'line of sight' between a goal or action and the ultimate goal or value it serves. This has the effect of making lower-level goals *seem* like ultimate, intrinsically valuable ones. Consider the idealistic college graduate who goes off to work in industry and begins to treat making money as something valuable unto itself. Or the computer salesman who persuades you to double the default amount of RAM on your new laptop,

even though none of your programs can actually make any use of the additional memory. When this type of Existential Distraction occurs, the effect is that of *pettiness*, i.e. where a goal or action is taken to be more important than it actually is. This dynamic often results from a change in one's environment, when the goal-striving processes of the old environment are still operative in the new one, before goal-setting processes have had a chance to catch up. (An extreme case of this is our biological goal of storing and retaining as much energy as possible, which in the environment of an African savannah led to survival; in an environment of Netflix and La-Z-Boys, it can lead to sickness and death.) When this type of Existential Distraction is pushed to extreme levels such that it affects one's reasoning capacities themselves, it spills over into Epistemic Distraction.

The third way Existential Distraction operates is by causing the person to *believe* that their immediate goal or action does not have a 'line of sight' with their ultimate 'being goal' (regardless of whether or not this is actually the case). Even if the person *is* acting in accordance with their values, Existential Distraction could make them believe that they are *not*, and this could cause them to judge that they are not 'being who they want to be.' (Whether or not they *would* actually be the person they wanted to be, even if they were not aware of it, is a question that I will sidestep for the purposes of this analysis.)

An example of this third type of Existential Distraction might be the case of the Facebook emotional contagion experiment that I discussed above. If a person were to interpret Facebook's alteration of their News Feed as unacceptable manipulation, and object to the image—the 'undignified reflection' — of themselves as someone who is not fully in control of their decisions about what they write in their own posts, then they would see their use of Facebook as incompatible with, and unsupportive of, the ultimate 'being goal' they have for themselves. The sense of a precipitous sliding backward from that ultimate goal would, as discussed above, have the

effect of undermining that person's sense of self-integrity, and would thus reduce their sense of dignity.

Epistemic Distraction

Epistemic Distraction is the diminishment of underlying capacities that enable a person to define or pursue their goals. This includes reflective capacities such as memory, prediction, reasoning, and goal-setting, as well as perceptual capacities such as the ability to identify goal-salient information in one's environment. In the context of persuasive technologies, Epistemic Distraction often occurs via the creation of new rules and associations in the executive control function, which can make it harder to 'integrate associations across many different experiences to detect common structures across them.' These commonalities 'form abstractions, general principles, concepts, and symbolisms that are the medium of the sophisticated, 'big-picture' thought needed for truly long-term goals' (Miller & Buschman 2014). In the absence of this capacity to effectively plan one's own projects and goals, the automatic, bottom-up processes of System 1 take over. Thus, at its extreme, Epistemic Distraction produces what Frankfurt (1988) refers to as 'wantonness,'[21] because it removes reflected-upon, intentional reasons for action, leaving only impulsive reasons in its wake.

I call this type of distraction 'epistemic' for two reasons. First, it distracts from knowledge of the world (both outer and inner) that is necessary for someone to be able to function as a purposeful, competent agent. Second, it constitutes what Fricker (2007) calls an 'epistemic injustice,' in that it harms a person in their ability to be a 'knower' (in this case, a knower of both the world and of oneself).

[21] Or in Raz's phrasing, being 'one who drifts through life unawares' (Raz 1986)

While my goal here is not to provide a comprehensive list of all capacities that Epistemic Distraction inhibits, the following may be considered representative examples:

- **Knowledge** — Reactance, as described in the section above, may cause imagined reflective information about a situation (e.g. 'Facebook's trying to manipulate my mood!') to crowd out perceptual information (e.g. the actual details of experimental results or effect sizes), which may inhibit knowledge about the world. It could also consist of inhibiting a user's understanding of the true nature of the economic tradeoff they are making by using a 'free' product that monetizes their attention (i.e. the 'user is the product' paradigm).
- **Reasoning** — A manipulative app or website may introduce information in such a way that it corrupts one's ability to reason effectively.
- **Intelligence** — A Hewlett-Packard study found that the 'IQ scores of knowledge workers distracted by e-mail and phone calls fell from their normal level by an average of 10 points—twice the decline recorded for those smoking marijuana' (Hemp 2009).
- **Expression/Language** — e.g. A person may feel manipulated by headlines that are highly optimized to result in their clicking on them, but not have the language (e.g. 'clickbait') to describe it.[22]
- **Reflection** — Notifications or addictive mobile apps may fill up moments in the day that a person might otherwise have used to reflect on their goals and priorities.
- **Willpower** — An addictive app that employs intermittent variable rewards could create ongoing dopaminergic effects that deplete a person's willpower and result in akratic behavior.

[22] A person's linguistic poverty may have other second-order effects on other capacities, e.g. their self-efficacy: 'Our annoyance dissipates into vague impotence because we have no public language in which to articulate it, and we search instead for a diagnosis of our*selves*: Why am I so angry? It may be time to adjust the meds.' (Crawford 2015, emphasis original)

- **Memory** — When an app notification or instant message from another person interrupts your focus or 'flow' (Csikszentmihalyi 2008), it may introduce information that crowds out other task-relevant information your working memory.
- **Physiology/Stress** — A phenomenon known as 'email apnea' has been shown to occur when a person opens their email and sees many unread messages, and a 'fight-or-flight' stress response is activated which causes the person to stop breathing (Stone 2008).

Like Existential Distraction, Epistemic Distraction also has an impact on both autonomy and dignity. It violates the integrity of the self by undermining the necessary preconditions for it to exist and to thrive, thus pulling the carpet out from under one's feet, so to speak. In fact, if we take capacities like reflection, memory, intelligence, etc. to be the hallmark of humanness, there is a very real sense in which one could say that Epistemic Distraction *literally* dehumanizes.[23]

V. Conclusion

I have argued here that: (1) in the context of persuasive technologies, many autonomy-based objections are in fact rooted in questions of dignity, especially when manipulation is the concern. Vigilance about the way our moral psychology frames the ethical questions we ask of persuasive technologies is extremely important. (2) Questions of both autonomy and dignity in the context of persuasive technologies can be usefully viewed through a common framework rooted in the psychology of human attention. Within such a framework, concepts such as 'attention' and 'distraction' can take on expanded roles, the latter of which I have typologized in three categories of potentially problematic persuasive mechanisms.

[23] As Levy (2007) writes, 'Bypassing our rational capacities in order to change our minds might carry a cost that is potentially far greater than merely passing up the opportunity to gain self-knowledge: it risks the very existence of our *self*, as it has traditionally been understood' (emphasis original).

Gordon Pask referred to cybernetics as 'the art and science of manipulating defensible metaphors' (Von Foerster 2003). As our information technologies increasingly draw on knowledge from psychology and behavioral economics to exploit users' cognitive vulnerabilities, they become more 'persuasive' in nature and play an ever greater role in our decision making. At present, the questions of what end goals these systems should strive toward, and who should define them, are extremely urgent. Because the end goals of these systems already significantly influence our lives, these questions warrant a sustained project of societal discussion, reflection, and creativity.

Which metaphors are most defensible? At the moment, the default perspective seems to be an economic one, within which some high-level, composite metric of 'value' or 'utility' becomes the object of optimization. But this is neither the only, nor the best, option. Alternate framings may include 'happiness' (whether of the narrowly psychological or the broadly eudaimonic sort), 'effective time use' or 'time well spent' (a repurposing of the business concept of productivity), 'fulfillment' (a goal- or preference-centric metaphor), and others.

Here I have sought to advance 'attention' as a possible alternative. We say that we live in the 'Information Age,' but we could just as justifiably call it the 'Age of Attention,' since this is now the scarce resource for which markets and technologies so vigorously compete. Yet in design ethics, this reality has not yet been fully taken into account. We have, as Aldous Huxley remarked of the defenders of freedom in his time, 'failed to take into account man's almost infinite appetite for distractions.' We have little understanding about how to think ethically about, measure, or design technologies to respect, users' attention.

What shall be maximized? This is ultimately the question. There may not be one answer, but where persuasive technologies exist we will need to give them one. My intuition is that we will need to go beyond 'attention' to find a really satisfactory answer. Instead, what is needed is a way of representing meaning in technological systems that mirrors the way we already represent it to ourselves. Such a representation would have close links to the way we construct our identities. It would also need to have wide validity across space, time, and cultures. And it would need to allow us to grapple successfully with the ethical considerations we consider most salient in the context of persuasive technologies (and decision architecture generally). The only thing I can think of that would satisfy these requirements is *story*. Is a good life a good story?[24] If so, will persuasive technologies be collaborative co-authors? It is my hope that this investigation has at least made that prospect a little bit more likely.

[24]'What I think is that a good life is one hero journey after another.' (Campbell 2004)

Chapter 3:

Reclaiming Advertising Ethics

Although digital advertising is the dominant business model in the attention economy, to date it has received neither the amount nor depth of ethical analysis it deserves. In this chapter I suggest that the absence of ethical attention to digital advertising has resulted, in large part, from the general failure of advertising ethics over the past century. Four dimensions of this failure are particularly relevant for engineering the future of digital advertising ethics: (1) because advertising ethics has received much less attention than it deserves, the field has failed to significantly guide the practice of advertising; (2) due to its framing as a subfield of business ethics, advertising ethics has been ill-served by an agent-centered ethical perspective; (3) work in advertising ethics has generally failed to account for the way in which information abundance produces a scarcity of attention; and (4) the definition of 'advertising' has been perennially vague. In response to the definitional problem, I develop what is, to my knowledge, the first user-centered definition of advertising: *a proactive appeal for a resource of value made in a way that overrides the dominant design goals for information delivery in that medium*. That is to say, when viewed as media dynamic, advertising functions as an *exception to the rule*. I close by discussing the urgency of advancing and accelerating ethical work on the unique challenges posed by specifically *digital* advertising.

I. The Virtual Field of Advertising Ethics

On Oct. 26, 1994, if you had fired up your 28.8k modem, double-clicked the icon for the newly released Netscape Navigator web browser, and accessed the website of *Wired Magazine*, you would

have noticed a curious new rectangle at the top of the page. In it, tie-dye text against a black background would have asked you, 'Have you ever clicked your mouse right HERE? *You will.*'

Whether intended as prediction or command, this message—the first banner ad on the web—was more correct than its creators could have imagined. Online advertising (which I will refer to here as 'digital' advertising, for reasons I will discuss presently) soon took off—and now, in 2017, digital advertising is by far the dominant business model for monetizing information on the internet. Its economic footprint is immense: digital ad spend is projected to pass $223 billion and continue to grow at double-digit rates until at least 2020 (eMarketer 2017). In the US alone, online advertising revenues in 2016 amounted to $27.5 billion, a 21.8% increase over 2015 (PricewaterhouseCoopers 2017). In fact, Nielsen projects that 2017 will be the first year in which global online advertising expenditures surpass those of offline advertising (Nielsen 2017). Across all media, it is estimated that the average adult is exposed to anywhere between two hundred fifty and three thousand advertisements per day (Abu-Saud 2013). As users' time spent with digital media continues to rise, an increasing percentage of these ad experiences will take place via digital media: according to eMarketer, between 2010 and 2014 for US adults this time spent rose from three hours eleven minutes to five hours and forty-six minutes—an eighty-one percent increase, to over one-third of waking life (Fisher 2014). For children, this time spent is even greater: in the US, the average child over eight years of age spends more than seven hours each day looking at screens—and over eighty-seven percent of popular children's websites contain advertising (Rideout et al., 2010, Cai & Zhao 2010).

Digital advertising has been, and continues to be, the driving force in the creation and evolution of many digital technologies. Today, many of the largest digital technology companies are primarily advertising companies (e.g. Google, Facebook, and Twitter), although the nature of their business

model is not well understood by their users. Furthermore, digital advertising is also arguably the largest and most sophisticated project for shaping human behaviors and attitudes in human history. Due to its dominance, many of the world's top software engineers, designers, and statisticians now spend their days figuring out how to direct people's thinking and behavior toward pre-defined outcomes that may or may not align with the goals they have for themselves. As Jeff Hammerbacher, Facebook's first research scientist, remarked: 'The best minds of my generation are thinking about how to make people click ads...and it sucks' (qtd. in Einstein 2016). Today, the greatest focus of these minds is on improving the performance of mobile device ads—a space in which just two companies, Facebook and Google, garner fifty-seven percent of total ad spend (Johnson 2017).

Given the scale, ubiquity, and importance of digital advertising—not to mention its rapidly increasing persuasive power—one would expect that, by now, a vigorous and sophisticated project of ethical guidance would be underway to steer this large-scale shaping of human thought and behavior in the direction of human wellbeing. While digital advertising has received some attention from advertisers and ethicists, very little work has resulted in changes to the way digital advertising is understood or carried out. (Ha 2008, Drumwright & Murphy 2013). Some initiatives emerging from the digital advertising industry have employed ethically-toned language in the context of identifying needed changes to advertising practices: efforts to identify 'responsible' advertising methods or promote 'acceptable ads,' for example (Acceptable Ads Committee 2017). However, these efforts have largely appeared as defensive measures aimed at preserving or at most slightly tweaking the status quo, and not seriously assessing the deeper ethical issues that digital advertising presents.

This lack of traction in the ethical steering of digital advertising is, in part, a continuation of failures in the field of advertising ethics broadly speaking. Over the past century, the advertising ethics literature has failed to clarify or influence the practice of advertising in any meaningful way. Why has this been so? For one, an extreme (though unwarranted) disinterest characterizes ethicists' views of advertising, as well as advertisers' views of ethics. Additionally, advertising ethics has been a victim of bad timing: because the modern advertising industry developed much of its sophistication in the mid-twentieth century, when foundational research in psychology—and particularly on indirect, non-rational psychological mechanisms of persuasion—was only emerging, ethicists lacked the conceptual and linguistic toolkits necessary for understanding, analyzing, and responding to these new persuasive methods. In any event, whatever the causes for advertising ethics' failure to launch, the lack of rigorous work on the topic, in combination with the new ethical challenges posed by digital advertising, make advancing the field now doubly urgent.

What does advancing the ethics of digital advertising require of us? First, it requires that we revisit the foundations of advertising ethics—its assumptions and definitions—in order to ensure a stable starting point. Doing so reveals four key ways in which advertising ethics has thus far failed—pitfalls that are essential to sidestep if we want to advance *digital* advertising ethics:

1. The way in which advertising ethics has received a marginal and tokenistic sort of attention, sharply limiting its influence
2. The industry-oriented—and, more generally, agent-oriented—approach of most advertising ethics research
3. The perennial vagueness of advertising's definition
4. The persistent failure to account for the way in which information abundance has inverted the relationship between information and attention.

If we are to avoid merely repeating these mistakes in a digital context, we must: (1) pivot advertising ethics from an agent- to a patient-centered perspective, (2) clarify and update the often vague definition of 'advertising,' and (3) fully account for (as well as slightly revise) Herbert Simon's observation that information abundance renders attention the scarce resource. After retrofitting these foundations, I will then build upon them in the following chapter to show how specifically *digital* advertising differs from advertising as it has been historically understood. In particular, I will identify a series of ethically salient differences that fall into two broad categories—differences of *boundedness* and differences of *intelligence*—which pose a package of serious and growing challenges for user self-determination that have so far gone under-addressed.

II. Marginal Ethics on Marginal Time

It is striking, especially when compared with other domains of applied ethics, how little attention advertising ethics has received in light of advertising's ubiquity and centrality in human life. A metaphor from the advertising world is perhaps apt for describing this inattention. In advertising parlance, the phrase 'remnant inventory' refers to a publisher's unpurchased ad placements, i.e. the ad slots of *de minimis* value left over after advertisers have bought all the slots they wanted to buy. In order to fill 'remnant' inventory, publishers sell it at extremely low prices and/or in bulk. One way of viewing the field of advertising ethics is as the 'remnant inventory' in the intellectual worlds of advertisers and ethicists alike.

This general disinterest in advertising ethics is doubly surprising in light of the verve that characterized voices critical of the emerging persuasion industry in the early-to-mid twentieth century. Notably, several of the most prominent critical voices were veterans of the advertising

industry itself. In 1928, brand-advertising luminary Theodore MacManus published an article in *The Atlantic Monthly* titled 'The Nadir of Nothingness' that explained his change of heart about the practice of advertising: it had, he felt, 'mistaken the surface silliness for the sane solid substance of an averagely decent human nature' (McManus 1928). A few years later, in 1934, James Rorty, who had previously worked for the McCann and BBDO advertising agencies (Newman 2002), penned a missive titled *Our Master's Voice: Advertising*[25] in which he similarly expressed the sense that some fundamental human interest was in the process of being violated:

> [Advertising] is never silent, it drowns out all other voices, and it suffers no rebuke, for is it not the voice of America? ... It has taught us how to live, what to be afraid of, how to be beautiful, how to be loved, how to be envied, how to be successful. ... Is it any wonder that the American population tends increasingly to speak, think, feel in terms of this jabberwocky? That the stimuli of art, science, religion are progressively expelled to the periphery of American life to become marginal values, cultivated by marginal people on marginal time? (Rorty 1934)

The prose of these early advertising critics has a certain tone, well embodied by this passage from Rorty, that is impossible for our twenty-first century ears to ignore. It is a sort of *parrhesia*, expressive of disbelief and even offense at the perceived aesthetic and moral violations of advertising, further tinged by a plaintive style of interrogation familiar from other Depression-era writers (James Agee in particular comes to mind). Yet from our historical vantage point there also seems to be an implicit optimism present here as well, i.e., in the mere fact that serious criticism is being leveled at advertising's existential foundations *at all*. Indeed, reading Rorty today requires a conscious effort to avoid projecting our own rear-view cynicism onto his apparently sincere prose.

[25] Rorty's title refers to 'His Master's Voice,' the famous painting of a terrier listening to his dead master's voice being replayed on a wind-up gramophone. The phrase later became the name of a British record company; today, both the phrase and image persist in the name and logo of the entertainment retail company HMV.

While perhaps less poetic, later critics of advertising were able to more cleanly circumscribe the boundaries of their criticism. One domain in which neater distinctions emerged was the logistics of advertising: as the industry matured, it advanced in its language and processes. Another domain that soon afforded more precise language was that of psychology. Consider Vance Packard, for instance, whose critique of advertising, *The Hidden Persuaders* (1954), had the benefit of drawing on two decades of advances in psychology research after Rorty:

> The most serious offense many of the depth manipulators commit, it seems to me, is that they try to invade the privacy of our minds. It is this right to privacy in our minds—privacy to be either rational or irrational—that I believe we must strive to protect.

Packard and Rorty are frequently cited in the same neighborhood in discussions of early advertising criticism. In fact, the frequency with which they are jointly invoked in contemporary advertising ethics research invites curiosity. Often, it seems as though they are invoked not so much for the content of their criticisms, nor for their antecedence, but for their tone: as though to suggest that, if someone were to express today the same degree of unironic concern about the foundational aims of the advertising enterprise as they did, and to do so with as much conviction, it would be too embarrassing, quaint, and optimistic to take seriously. Perhaps Rorty and Packard are also favored for their perceived hyperbolizing, which makes their criticism easier to dismiss. Indeed, as Wu (2016) points out, the ad industry at the time responded to these (and other similar) attacks as communist in nature—as attacks on capitalism itself. Finally, it seems to me that anchoring discussions about advertising's fundamental ethical acceptability in the distant past may have a rhetorical value for those who seek to preserve the status quo; i.e., it may serve to imply that any ethical questions about advertising's fundamental acceptability have long been settled.

Yet the absence of a rigorous, sustained project of advertising ethics has left many questions unresolved, if not unasked. Indeed, as Drumwright & Murphy (2009) write, 'Despite attention to issues of advertising ethics through the decades, it would be a mistake to assume that advertising ethics has received coverage commensurate with its importance. While advertising ethics has been recognized for some time as a mainstream topic (Hyman et al. 1994), research is thin and inconclusive in many important areas.' This dearth of attention to advertising ethics extends beyond research and into pedagogy as well, for example in university coursework and textbooks on advertising (Drumwright & Murphy 2009).

One reason for this inattention is undoubtedly the perceived interestingness of the topic. To be sure, there are few subjects initially as boring to the ear as 'advertising ethics.' I say this not to malign the field, but rather to suggest boredom as a serious hypothesis for why so little serious work has to date been done on such a serious topic. In fact, advertising ethics could almost be said to be a 'virtual' field of inquiry, in that much of it gives the impression that it is written for an audience that does not exist. Is it the case, perhaps, that the 'advertising' part has kept the ethicists away, and the 'ethics' part has kept the advertisers away? It appears that at least the latter is true. Reflecting on their survey of ad industry professionals, Hyman et al. (1994) indicated that 'lack of practitioner interest' was the largest impediment to advancing research in advertising ethics. It is highly plausible that this disinterest flows in the other direction as well. For a philosopher to spend scarce attention on the goals and concerns of what is often perceived as a narrow, practical, and *commercial* domain—to walk in a back alley of applied ethics as opposed to the grand gardens of, say, metaethics—would seem a petty and dispiriting prospect. It could only lead, one might imagine, into poorly lit cul-de-sacs of legalism and wrist-slapping.

Where philosophers *have* given attention to advertising, it has often been done against a background feeling of futility—a sense that while we might imagine new ways for advertising to be, any major change is unlikely. Likewise, advertisers have incentive to grapple with ethics as minimally as possible, as an insurance policy rather than a guiding light. In practical contexts, the interest is to be *seen* to be 'ticking the box' and to be 'doing ethics.' As a result, ethical analysis from an industry perspective has tended to function as defense against, or minimization of, ethical issues. The end result of these challenges on both sides has been a minimal, sporadic literature that has wandered without any real vigor or direction—and that is when the right conversations are even taking place. As Drumwright & Murphy (2009) write, 'Disagreement is not the problem; avoidance of the topic and/or failure to engage in a collaborative dialogue is.'

III. The Agent-Orientation of Advertising Ethics

A second reason for the minimal impact of advertising ethics to date relates to the agent-centered way in which it has almost universally been framed. By 'agent' here I mean, following Floridi and Sanders (2001), 'a system, situated within and a part of an environment, which initiates a transformation, produces an effect or exerts power on it over time.' Correspondingly, 'patient' is defined as 'a system that is (at least initially) acted on or responds to' an agent.

In advertising ethics to date, the agents have largely been the industry practitioners, in particular those closest to the logistics of ad campaign implementation (i.a., advertisers within companies or media buyers at advertising agencies). As a result of this agent- and industry-centered view of advertising, its ethics have generally been viewed as a subfield of business ethics. For example, in Dow (2013) we read that 'The ethics of advertising is … concerned with what considerations advertisers should take proper account of in the course of their work.' One unfortunate effect of

this view is that advertising ethics has been frequently conflated with law (Drumwright 1993). Within the advertising industry, a widespread view exists of ethics as a 'brake pedal,' a constraint on action to be avoided, rather than as an accelerator, or perhaps a steering wheel, that helps inform action. At Google's European headquarters in Ireland, for example, the wall of one floor contains a large screen-printed phrase that reads, 'If you are not breaking any laws, then go ahead and do it!'

To be sure, if we understand 'advertising ethics' in this agent- and industry-focused sense—as, e.g., 'the development of frameworks and guidelines that help advertisers promote their products to consumers in a more ethical way'—then the paths forward seem narrow, parochial, and uninteresting indeed. However, if we understand it instead as 'an effort of critical analysis, imagination, and co-creation aimed at steering the largest project of attitudinal and behavioral persuasion the world has ever seen—and the primary incentive structure of our first truly global communications medium, which determines the information most people see every day—toward human flourishing,' then it takes on a quite different—and breathtaking—shape.

In order to avoid the narrow parochialism that has hindered the field of advertising ethics to date, as well as take a more holistic view of digital advertising's nature and effects, it is necessary to pivot from an agent- to a patient-centered perspective. In a patient-centered view, 'a process or an action may be right or wrong irrespective of its consequences, motives, universality, or virtuous nature, but because it affects positively or negatively its patient and the infosphere' (Floridi 1999). In a patient-centered ethics there may be no one to *blame*, but there may still be something that can be *done*. A patient-centered approach is particularly appropriate as a foundation for specifically *digital* advertising ethics, for two reasons. For one, digital advertising's *direct* effects on users increasingly occur as the results of interactions across multiple touchpoints, and are often orchestrated by

multiple agents (both human and artificial), such that attributing and distributing accountability for their effects is difficult, if not impossible. The second reason is because digital advertising's *systemic* effects on users—for instance, via influencing the incentive structures that drive design changes in the underlying structure of the media themselves—often emerge as epiphenomena, and have no definable agent to whom we can attribute causality or accountability.

In design terms, we might say that this shift from an agent- to a patient-centered approach is akin to applying the principles and assumptions of User-Centered Design (UCD) to ethical questions. UCD emerged from mid-century human factors and ergonomics research and has been a dominant design paradigm in the digital technology industry where it has, in a sense, afforded the development of new 'patient-centered' approaches to technology design in a variety of contexts. However, it is noteworthy that despite UCD's wide adoption by digital technology companies, it has been minimally applied to the design of advertising experiences (that is, it has rarely been applied to advertising *as* UCD; in the advertising context, UCD techniques and principles have more often been deployed in the service of advertisers' and platforms' persuasive design goals, rather than the user's personal goals). A *de facto* mantra of UCD can be found in the first item of Google's list of core company principles, titled 'Ten Things We Know to Be True,' which reads: 'Focus on the user and all else will follow.' When it comes to the design of users' advertising experiences, this maxim still largely functions as an aspiration or motivation rather than a statement of established process. Regardless, at least it motivates in the right direction. Yet ideally, this mantra ought to inform and motivate not only the design and evaluation of products, but also the ethical perspectives that guide their design.

Shifting to a patient-centered ethical perspective has many important implications, but a particularly significant one is the way it requires a shift in the definition of advertising. This shift

allows (if not requires) us to reconceptualize advertising as a media dynamic (and, I would argue, as the most consequential media dynamic in our informational environment).

IV. The Mercurial Definitions of 'Advertising'

Agent-centered approaches to advertising ethics—which typically take the actions of advertising industry actors as their starting point—depend on, and also reinforce, agent-centered definitions of 'advertising' as a whole. Pivoting to a patient-centered ethics thus requires a corresponding pivot to a patient-centered definition of advertising as well. Such a definitional pivot is important not only for clarifying the conceptual foundation from which the present analysis of digital advertising can proceed, but also for identifying a baseline set of criteria for advertising in general against which the new features and affordances of *digital* advertising can be compared.

Curiously, even from an industry-oriented perspective the term 'advertising' has never had a straightforward definition. There are two reasons for this. One is the multi-purpose nature of the term 'advertising.' The word can be used to describe a particular *message* (e.g. the famous 'I'd Like to Buy the World a Coke' television ad), a particular *type* of message in a given medium (e.g. television commercials generally), part or all of the advertising *industry* itself (including e.g. ad agencies, media buyers, and advertising platforms), and the *processes* of communication and persuasion that the advertising industry deploys. The other reason is the question of the definition's *goal*. Why do we want to clarify our definition of advertising at all—what do we want it to *do* for us? Most definitions of advertising have arisen in contexts where the goal, whether implicit or explicit, has been to enhance advertising's effectiveness in some way: e.g. in advertising industry associations, business school textbooks, or research undertaken by advertising practitioners. As for those on the receiving end of advertising's effects, i.e. consumers themselves,

there has historically been very little need (or, perhaps, little *perceived* need) for a sharply defined concept of advertising.

In one of the few papers that directly grapples with the definition of advertising (which is done, as one might expect, from an industry perspective), Richards & Curran (2002) write that 'a survey of recent advertising and marketing textbooks makes it obvious there is no widely adopted definition at this time.' The authors produce, via a Delphi study of advertising industry professionals, a consensus set of elements that they propose should be included in definitions of advertising. Those elements are: '(1) it is 'paid,' (2) it is 'nonpersonal,' (3) it comes from an 'identified sponsor,' (4) it is communicated via 'mass media,' and (5) it aims to 'persuade or influence.' Of course, by limiting their interview participants to professionals in the advertising industry the authors ensure from the outset that their results will reflect a practitioner-oriented, rather than user-oriented, perspective. That said, their five criteria do seem *internally* consistent with other definitions rooted in an industry perspective, which include the research literature (e.g. Rodgers & Thorson 2012, Turow 2012) as well as industry associations and reference points (e.g. *The Common Language Marketing Dictionary* 2016). Thus, even though the advertising industry may not be able to *precisely* define the nature of its craft, it seems they generally know advertising when they see it.

Yet one definitional criterion common across these industry voices is flatly wrong, even when viewed solely from an industry perspective: the criterion that an advertisement must be 'paid.' To illustrate why this requirement is unnecessary, imagine that a seller of advertising space (e.g. a newspaper or search engine) began giving 10% of their ad inventory away for free (which advertising platforms sometimes do, for various reasons). Would that free 10% of served ads cease to be 'advertising'? Clearly not. While the existence of a payment is essential for commerce, and most advertisers operate in a commercial context, payment is not a necessary component of the

mechanisms of advertising. Rather, it is a by-product of the need to override the dominant design goals of a medium that guide its normal mechanisms of information delivery. For example, consider word-of-mouth advertising on social media platforms (e.g. Facebook or YouTube), which typically involves a company giving a product for free to a user who is identified as an 'influencer.' When another user sees the influencer with that product, and perhaps later buys it as a result, that user has still seen an 'advertisement' even though they are unaware of it, because the normal mechanism of information delivery (i.e. the products the influencer would have otherwise used) is being overridden. This is similar to product placement in movies (e.g. being impressioned by the particular Sony phone you notice James Bond using to get himself out of a jam). Furthermore, there is the broader point that *all* communication is 'paid' in *some* way. For instance, consider the practice of search engine optimization (SEO), in which the owner of a website pays specialists to optimize it (e.g. by minifying its JavaScript or adding 'meta' tags to its HTML headers) in order to increase the site's 'organic' (i.e. non-paid) traffic. SEO is a 'paid' effort of communication that would arguably meet all five criteria for advertising given by Richards & Curran (2002) above.

Apart from recasting the 'paid' criterion of advertising (which I will define explicitly below), there are two attributes of advertising sitting in the background of the industry-centered definitions given above that a patient-centered perspective requires us to bring into the foreground. The first is a distinction widely made between two general types of advertising—a distinction that reflects the nature of the advertiser goals it serves. These two types of advertising often go by different names: 'branding' vs. 'direct-response,' 'persuasive' vs. 'informational,' 'demand creation' vs. 'demand generation,' etc. At root, this distinction reflects the fact that advertising is not only concerned with shaping *behavior*, but also with shaping *attitudes* (Crisp 1987). Landa (2005) describes the first ad campaign to use this modern approach to branding:

> Uneeda biscuit, a packaged brand-name cracker made by the National Biscuit Company (now Nabisco), hired the advertising agency N. W. Ayer & Son to create an integrated brand campaign for their product. The agency suggested the brand name, the character (a little boy in a raincoat to suggest air-tight freshness and crispness), and the slogan 'Lest you forget, we say it yet, Uneeda biscuit.' This historic campaign, launched in 1899, was the first multimillion-dollar ad campaign; it would change everyone's perception of the critical role of branding and advertising. (Landa 2005)

At a high level, these two types of advertising are understood to work together as part of the overall customer 'funnel' or 'journey'—the steps on the 'path to purchase'—yet their mechanisms, methods, and effects are quite different from one another. (As I will discuss later, this difference has been widened even further by the unique affordances of digital advertising.) 'Brand' advertising, which at its most basic level increases cognitive recall of a particular brand and associates it with positive affect, engages our non-rational minds by way of the 'availability bias.' (Kahneman & Tversky 1973)

The second factor implicit in these industry-centered conceptions of advertising that should be brought to the foreground is advertising's *proactivity*. Imagine that you are in the Sunday Market and you see a man sitting quietly in his booth behind the table of handmade porcelain bowls he is selling. Would we say the man is 'advertising'? Typically we would not, because *you* would have to approach *him*. However, if the man were to yell at you, or walk over to you and grab your hand to make you feel the bowl's texture, we probably *would* say he is advertising. Now contrast that case with a situation in which you are reading the classifieds section of your local newspaper, and you see a supermarket's offer for 50% off the price of avocados. Is *that* advertising? In some sense, maybe—yet it still seems more akin to the old man at his booth in the market: *you* had to go to *it*, and you *knew* that sort of message would be there. So it seems intuitive to say that a very strong tendency of advertising (though probably not an airtight rule) is that it *advances* toward *us*, rather

than we toward it. Framing this distinction in erotetic terms may be productive: in such a framing, we might say that advertising does not function as a *response* to a question, but rather *poses* a question to us, in the form of 'why not…?'

Thus I propose a patient-centered definition of advertising as: *a proactive appeal for a resource of value made in a way that overrides the dominant design goals for information delivery in that medium.*

- **'Proactive'**—I use this term in the organizational psychology sense, which implies behavior that is 'anticipatory, change-oriented, and self-initiated' (Öncel 2014).
- **'Appeal'**—The connotations (and etymology) of this term include that of *addressing*, *calling upon*, and *asking for*.
- **'Resource of value'**—I use the term 'resource' loosely here. On a broad view, it could refer to the user's purchase of a product (money), brand favorability (e.g. regard or affect), or some other action the advertiser wants the viewer to take (effort, time, etc.). At the very minimum, it includes the user's time and attention that the advertiser wants them to give to the advertisement.
- **'Dominant design goals for information delivery'**—The medium's overarching design *goals* (or design *reasons*) that determine what information is surfaced to the user, and how. It could be objected that some media have *no* overarching design goals, or perhaps that they have *multiple* competing design goals. For example, newspapers contain both 'news' and 'opinion' articles, and they (usually) draw a clear boundary between the two. (At *The New York Times*, for instance, news and opinion writers even sit on different floors; writers on one side are physically prevented from accessing the office space of the other.) Would opinion articles, in this view, not count as an 'exception to the rule' similar to how I am describing advertising here? Perhaps they would—and perhaps some of them, if they

satisfied my other criteria here, would even qualify as 'advertising.' (In fact, full-page advertisements are frequently purchased in the *NYT* for just this purpose, to serve as *de facto* opinion pieces.) It is certainly true that a medium may contain subsections having separate, even competing, design goals. However, this does not mean that the medium as a whole has *no* overarching design logic. Every medium does, even if its designers are not conscious of it.

- **'Medium'**—Since the emergence of software it has become difficult to speak of distinct 'media' with any sort of clarity. It is most possible to do so where affordances are relatively stable over time; this typically means that hardware is where it is easiest to talk in terms of 'media,' then at the operating system level, and then decreasingly so as the system is further virtualized. In the world of software, and especially digital services, the *perceived* boundaries of a medium are more often semiotic and psychological in nature (e.g. defined by brand identifiers) as opposed to being grounded in the dynamics of the technological infrastructure.

To summarize, this patient-centered view of advertising reframes it as a media dynamic that functions as an *exception to the rule* for information delivery in a given medium. (e.g., newspaper ads vs. articles, billboards vs. street signs, or TV commercials vs. programs)

IV. Attentional Blindness: Advertising Ethics and the Simonian Inversion

The fourth important blind spot in advertising ethics is one it shares with most ethical analysis—and indeed most analysis generally—of information technologies. It is a failure to account for the inversion in the relationship between information and attention in the era of

information abundance. In 1971, Herbert Simon was arguably the first to draw widespread attention to this dynamic:

> ...in an information-rich world, the wealth of information means a dearth of something else: a scarcity of whatever it is that information consumes. What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it. (Simon 1971)

Across countless domains of life and inquiry, we are only really beginning to understand the implications of this reversal, which I will refer to here as the 'Simonian Inversion.' Advertisers, their business being the competition for human attention, are naturally far ahead of everyone else in understanding and adapting to the Simonian Inversion. However, given advertisers' significant lead over the rest of us in this regard, it is fascinating that advertising *ethics* has not likewise benefited from this awareness. In fact, when it comes to specifically *digital* advertising vis-a-vis the state of its ethics, the gap in accounting for the Simonian Inversion is even more pronounced. In work on digital advertising ethics, focus has predominantly gone to questions about the management of information or data, rather than the management of attention or behavior: the *content* of advertisements have taken precedence over the *dynamics* of its interactions. These informational questions have included things such as: the ethics of advertising particular types of products (e.g. cigarettes); the ethics of targeting advertisements at particular audiences, such as children (e.g. Nairn & Dew 2007, Calvert 2008); or the ethics of tracking and storing users' data (e.g. information about their web browsing behaviors collected via tracking 'cookies'). In Hyman et al. (1994), deception was the topic that their study's respondents thought was most important to the study of advertising ethics. (Nothing about manipulation, persuasive power, or freedom/autonomy even appeared in their list of topics.)

While important, such informational ethical questions only reflect a portion of the spectrum of relevant ethical questions. Yet they comprise the vast majority of work done in the field to date. As a result, digital advertising ethics has merely addressed the figure, and not the ground, of its dynamics. It has tended to address questions of *direct* causation (e.g. the way advertisements themselves shape people's behavior) to the exclusion of questions of *systemic* causation (e.g. the way advertising creates incentives that shape the design of media or platforms as a whole)[26][27].

However, these attentional questions, and questions of systemic causation, raise extremely important ethical questions about autonomy, self-determination, and respect, among other things. To be sure, these questions are not new (consider e.g. the mid-century paranoia about 'subliminal' advertising). However, given the current scale, dominance, and persuasive power of digital advertising, they are newly urgent—and essential to ask.

VI. The Urgency of Clarifying Digital Advertising

Why is it so urgent to reexamine digital advertising *now*? What is particularly new about the present situation? For one, advertising is a *pervasive* force in human life. This was already the case prior to the digital revolution, and it is even more so now. Second, digital advertising—and indeed digital technologies in general—are becoming ever more *persuasive* in their aims and effects. The testing and iteration of persuasive design patterns, along with more effective application of knowledge about non-rational human psychology, is enabling this persuasiveness to rapidly advance.

[26] The language of 'direct' versus 'systemic' causation is drawn from Lakoff (2015).
[27] As Luciano Floridi (2017) has pointed out, this also parallels the distinction between an 'ethics of things' and an 'ethics of relations.'

Advertising ethics has not received the attention it deserves given advertising's centrality and ubiquity in human life, nor has it significantly steered the overall direction the practice of advertising has taken. This has left us with a situation in which our most effective efforts of behavioral and attitudinal change now function as the primary and default business model for the first truly global communications network in history. This alone would be cause for urgently prioritizing work on the ethics of advertising. But there is much more. The advent of digital technology and networked software systems has enabled advertising to advance in certain respects, such that digital advertising departs in key ways from advertising as historically understood. In particular, with respect to its *boundedness* and its *intelligence* (as I will describe in the following chapter), digital advertising has fundamentally evolved beyond advertising as we have historically understood it. The effect of this is that, rather than being an *exception* to the rule of information delivery in a given medium, advertising now *is* the rule. Digital advertising is now best understood as the primary design logic of the informational environment. In the past, we would say that advertising was 'underwriting' the content that was already designed according to other goals. Now, however, advertising is 'overwriting' the design goals of our media, and putting its own goals in their place.

The extent to which this is the case is not widely understood, in large part because advertising is already something that operates behind the scenes. It is also because, as I have said above, while advertising has effectively marketed itself as an interesting and important force in society, it has also marketed itself as a force that needs little attention and is best left alone to function autonomously, much like the 'man behind the curtain' in *The Wizard of Oz*. But we cannot leave the curtain undrawn now that advertising also serves as the primary set of incentives behind the design of emerging technologies such as artificial intelligence and virtual reality. Furthermore, many of the enormous challenges facing humanity depend, crucially, on *changing minds* in order to be

surmounted. Thus it is time to take a new look at advertising—at what it has been, what it has become, and what we want it to be.

VII. Conclusion

I have argued here that the ethical analysis of digital advertising today begins at a disadvantage, in large part due to failures it has inherited from the field of advertising ethics broadly speaking. In particular, those failures include: chronic inattention to serious ethical problems; vagueness about what *counts* as advertising in the first place; rooting its perspectives in the processes and concerns of industry rather than the experiences and needs of users; and a failure to reevaluate the nature and purpose of advertising in the new environment of information abundance. To help overcome these challenges, and to clear the ground for further work in digital advertising ethics, I have offered what is, to my knowledge, the first user-centered definition of advertising: *a proactive appeal for a resource of value made in a way that overrides the dominant design goals for information delivery in that medium.* This definition will inform the analysis of digital advertising that I undertake in the next chapter.

Chapter 4:

The Moral Character of Digital Advertising

In this chapter I assess the nature and ethics of specifically *digital* advertising and argue that it poses important challenges to human self-determination that have to date gone largely unacknowledged. I show how digital advertising radically departs from advertising as historically understood, particularly in its *boundedness* and its *intelligence,* such that it no longer functions as an *exception* to the rule; instead, it now *is* the rule. In traditional media, advertising was said to 'underwrite' the dominant design goals; in digital media, advertising now 'overwrites' them with its own. As the now-dominant design logic of our information environment, it is thus arguable whether digital advertising ought to be described as a form of 'advertising' at all. This 'overwriting' tendency of digital advertising has produced several direct as well as systemic effects that pose serious challenges to user self-determination. However, these issues have received little serious ethical attention to date and therefore warrant a sustained project of critical analysis. I close with a discussion of the potential ethical obligations, as well as opportunities, that follow from this assessment of digital advertising—several of which sit in direct tension with other ethical interventions that have been proposed or implemented to date.

I. Digital Advertising: A Tectonic Shift

Perhaps the most striking thing about digital advertising is how profoundly—and how widely—it is misunderstood. Two types of misunderstanding occur most frequently (and often co-occur). The first type misunderstands the 'digital' part; the second misunderstands the 'advertising' part. In the first case, digital advertising is viewed as simply an extension of the existing practice of advertising

into a new medium called 'the internet' that sits on a conceptual shelf in one's mind next to the 'traditional' media such as television, radio, and newspapers. Though a newer arrival, the internet is viewed as being the same general *sort* of media as the others. This error is commonly embodied in the organizational structures of companies who cordon off their digital marketing and advertising efforts into a 'silo,' i.e. by creating a separate 'digital' team to deal with 'online' matters. The academic literature, too, has suffered from this 'online'/'offline' dichotomy: as Ha (2008) observes, 'the bulk of online advertising research views online advertising as an alternative to offline media advertising.' The error in this thinking lies, of course, in a failure to understand the extent and nature of digital technology; it is also a testament to the strength of habits.

In the second type of misunderstanding, the nature and historical import of digital media may be understood well enough, but 'advertising' is viewed as being broadly analogous to advertising as it has existed historically. The possibility that digital media may have altered the practice of advertising in fundamental ways is seldom considered; digital advertising is seen as different in degrees, but not in kind. For instance, Goldfarb (2014) views the main difference of digital advertising as being about 'a substantial reduction in the cost of targeting' (and on this basis he identifies the online advertising literature's primary foci as 'understanding advertising effectiveness, auctions, privacy, and antitrust'). The first scholarly article on online advertising was published by Berthon et al. (1996) in *Journal of Advertising Research*, and drew on 'the metaphor of an electronic trade show and a virtual flea market' to identify 'five advantages of online advertising: 1) Awareness efficiency, 2) locatability, 3) contact efficiency, 4) conversion efficiency, and 5) retention efficiency' (Ha 2008).

In reality, the mechanisms and effects of digital advertising represent a radical departure from previous forms of advertising. The most significant of these differences may be viewed as falling

into two high-level categories: differences of *boundedness* and differences of *intelligence*. Advertising's boundedness is dramatically decreased in digital contexts primarily due to its basis in software, which not only renders the design constraints on both the media and advertisements themselves more fluid but also enables the wholesale dissolution of boundaries that previously demarcated ads from non-ad content. Two important secondary forces that also reduce the boundedness of advertising in digital media are the fragmentation of content (e.g. the 'unbundling' of newspapers on the web so that articles, not editions, are the units of monetization) as well as the global scale of both audiences and advertisers.

When these new forces of decreased boundedness emerged, they created new possibilities for advancing advertising that were keenly understood by many in the advertising industry. In the mid-1990's, Edwin Artzt, the CEO of Procter & Gamble, gave a speech to advertising industry leaders in which he pleaded with them to advance the use of internet technologies for advertising. He said, 'We may not get another opportunity like this in our lifetime. ... Let's grab all this technology in our teeth once again and turn it into a bonanza for advertising.' Joseph Turow, whose book *The Daily You* (2012) provides a valuable narrative account of the emergence of digital advertising, writes, 'What worried P&G's chairman primarily was not that new technologies would encourage more targeted advertising. Rather, it was the 'chilling thought' that emerging technologies were giving people the opportunity to escape from advertising's grasp altogether.'

Soon, a sophisticated infrastructure of advertising technology (or 'ad tech' for short) emerged to help advertising advance in its *intelligence* to keep step with the new opportunities. This infrastructure included a new complex of measurement, analysis, and optimization capabilities referred to as 'analytics,' which crucially included the development of the behavioral tracking 'cookie' that ultimately enabled the rise of web analytics platforms (which allowed media and

advertising analytics to be housed in the same place, and to align in their methods of reporting, analysis, optimization, and experimentation). Two other important forces advancing advertising's intelligence were automation (which enabled so-called 'programmatic' advertising), and personalization (which enabled the tailoring of advertising messages at scale). All these forces of boundedness and intelligence have variously motivated, enabled, and amplified each other's operation.

Before I turn to examine these differences in detail, a point on terminology is necessary. The type of advertising I am discussing here has been known by various names; 'online,' 'digital,' and 'internet' advertising are the most common. Here I will use the term 'digital advertising.' I find the terms 'online' and 'internet' advertising problematic because the online/offline duality has already been obsolesced by the 'onlife' conception of the infosphere (Floridi 2014). Admittedly, 'digital' is not the ideal adjective either: the presence of digital computation *per se* is sufficient, but not necessary, to enable networked computing systems that permit the forms of advertising I am discussing here. In principle, any form of non-analog computation—whether digital, quantum, biological, or otherwise—would suffice. Yet to demand a more precise conceptual boundary here would be to over-engineer the term relative to my present purpose; for now, 'digital advertising' gets the job done.

II. Differences of *Boundedness*

Compared with previous forms of advertising, digital advertising is far less bounded in its design and operation. This unboundedness takes three main forms. First, having a basis in software has made its design constraints—and the constraints that would otherwise separate it from the design of the underlying media as a whole—far more fluid. Second, the fragmentation of content in digital

media has meant that advertising's unit of monetization is far smaller than it was previously. Third, the scale of competition among persuasive actors in the digital advertising environment is now global and far more vigorous.

a. Fluidity of Design Constraints

Before software and virtualization, experience design within a given medium was less flexible and more constrained by certain non-negotiable boundaries in the medium's underlying socio-technical infrastructure. These boundaries included the physical capabilities *across* the medium as a whole (e.g. limitations on sensory modalities or contexts of use) as well as constraints on the degrees of design freedom *within* advertising elements (e.g. strict limitations on available or acceptable ad formats). This meant that, even if the effectiveness of an advertising effort *could* be measured, there was a built-in limit as to how much subsequent optimization efforts could influence the underlying design of the medium itself. These built-in boundaries thus kept pre-digital advertisements as *exceptions* to the rule of existing design goals.

However, in a software-based environment, there are very few aspects of digital media that are ultimately non-negotiable in the face of demands from digital advertising. For one, this means that many different ad formats can be rapidly created & tested. Since *Wired*'s first banner ad in 1994, many new types of advertising have emerged on the web: search, contextual targeting, video ads, interactive ads, referral, social network ads, influencer marketing (e.g. giving products to influencers on YouTube), product placement, animated ads, homepage takeovers, interstitials, and many more.

In addition, this loosening of design constraints enabled a wide variety of new payment mechanisms to emerge. Initially, most digital ads were sold on a CPM (cost-per-thousand-views) basis. Soon, however, CPC (cost-per-click) advertising emerged and was particularly popularized by its use in Google search ads, which further incorporated an auction component and a behavioral signal of ad relevance (called 'Quality Score'). Today, improved measurement across the so-called 'purchase funnel' has enabled the development of payment methods such as CPA (cost-per-action), in which the advertiser pays based on the number of people who ultimately take the desired action after seeing their ad.

Finally, and perhaps most importantly, this loosening of boundaries on advertising in digital media has led to a blurring of the lines between advertising-oriented and non-advertising-oriented design. On a surface level, this manifests as new ad formats that make only token, trivial gestures toward identifying themselves as ads. At present, the most noteworthy of these ad formats is so-called 'native advertising.' A native advertising unit is an advertisement with a similar, perhaps even identical, look and feel to the usual content in that medium. For instance, it may be a news story, written by a professional journalist, on a topic and in a style similar to those often covered by the publication it appears in. The only difference may be the presence of a signal—usually small and peripheral in nature—identifying it as 'sponsored' or 'promoted' content. The guidance offered by the Internet Advertising Bureau (IAB) on native advertising disclosure (which is extremely general in nature) makes it clear that the IAB regards the emergence of native advertising as an important dismantling of a boundary: 'First and foremost, display advertising has been freed from the right rail and leaderboard to which it has long been confined and now has license to settle anywhere on the page. Moving forward, display advertising will not be forced back into solely those positions.' (IAB 2016)

On a deeper level, the loosening of boundaries on advertising is significantly influencing the direction of the underlying infrastructure of media itself. This is true of both back- and front-end design. On the back end, this can be seen in, for example, the way retailers' product 'feeds'—i.e. the structured data sets that reflect their product offerings and availability—are now being used as a source for automated advertising (e.g. in Google Product Ads, Hotel Ads, or Flight Search) and thus increasingly adopt feed specifications primarily on the basis of how likely they are to increase the success of those ads. On the front-end, this is seen in the emergence of the phenomenon known as 'clickbait,' i.e. articles with topics selected, and headlines engineered, specifically to maximize the number of pageviews they receive.

Looking ahead, as more and more aspects of the world become digital and networked (e.g. as ubiquitous computing and miniaturization give rise to the physical internet), the assumptions and capabilities present in the design of networked software environments are being extended to the design of physical environments as well. This digital translation has already happened with so-called 'traditional' media such as television and radio. The implications here are that a similar lack of firm design constraints, extended to these contexts, could enable digital advertising to commandeer the design goals of a wider range of media and have a similar influence across a greater portion of our lives.

b. Fragmentation of Content

Prior to the internet, content creation was largely tied to content distribution. This meant that (a) it was *bundles* of content (such as a newspaper or magazine), rather than atomized *units* of content (such as an individual article or video clip), that were being monetized, and (b) advertisers had to interact separately with many different platforms in order to buy and run their advertisements. On

the web, however, content creation became separated from content distribution. Now, individual pieces of content rather than bundles of content have become the units of monetization. For example, rather than receiving an entire daily issue of a physical newspaper, a user on the web may now only access a single article that a given newspaper produces.

Furthermore, users are likely to find these single units of monetized content via the new gatekeepers of content distribution, in particular via search engines and so-called 'social media' platforms. As a result, content producers must increasingly optimize for the sort of content these gatekeepers select for—which tends to be content that performs well against 'engagement' metrics such as the number of 'clicks' or 'shares' they receive. In this way, the fragmentation of content has served to pit longer and more nuanced material against lighter fare such as 'clickbait' that is designed to elicit automatic, impulsive responses from users.

Accompanying, and in large part produced by, this fragmentation of content has been the fragmentation of content publishers themselves. New so-called 'long-tail' publishers have emerged to provide niche content tailored for more specific audiences, something that in a pre-digital world would have prohibitively expensive to do.

c. Scale of Competition

Before digital media, advertising competition was typically local or national at most—but regardless, it wasn't (truly) global. There was also a relatively high bar to entry (e.g. it took a lot of money to be able to advertise; it couldn't be done piecemeal). Furthermore, the advertising industry was large but relatively stable in terms of its main companies and agencies, as well as its typical processes for creating and buying ads.

Now, in a digital context, the bar to being able to advertise is very low: for just a few dollars, one can run their own advertising campaign on a variety of platforms and in many formats. This lower bar to entry has led to increased competition: all sizes of companies can now compete—even the so-called 'long tail' of companies—and from anywhere in the world. Also, given the pace of innovation, the digital advertising industry is now highly volatile, with new technologies and companies emerging all the time. That said, a few particular companies known as 'stack players'—companies such as Facebook or Google, who aim to own the end-to-end pipeline from ad creation to delivery—have emerged as forces of centralization.

In a wider sense, the scale of information itself—and, in particular, the amount of *choices* available for users to consider—is also dramatically larger. Similarly, the number of users (i.e. the amount of attention available for advertisers to monetize) is enormous and growing rapidly. In particular, Southeast Asia—particularly China and India—is poised for significant near-term growth; many companies and designers describe this next wave of new internet users as being the 'next billion users' (or 'NBUs').

III. Differences of *Intelligence*

In addition to differences of *boundedness*, digital advertising departs from previous forms of advertising due to its increased *intelligence*. One place this is readily apparent is in the extremely sophisticated complex of capabilities known as 'analytics,' which encompasses tools and processes for measurement, experimentation, and prediction. A second, related, area is that of automation. On one level, advertising provides automation with a wide variety of socio-technical operations that can be made faster and more efficient. On a deeper level, though, advertising offers

automation a unique incentive structure, with a unique competitive urgency, that now serves as one of the primary, if not the foremost, drivers of innovation in machine learning and AI as a whole. Finally, the third area increased intelligence is apparent is in digital advertising's vastly increased capacity for customization, which influences not only the design of advertising messages but also the design of the medium itself.

a. Analytics

Most advertising throughout history has been faith-based. Without a comprehensive, reliable measurement infrastructure, it was simply impossible to study the effectiveness of one's advertising efforts, or to know how to make them better. As John Wanamaker, a department store owner near the turn of the twentieth century, is reported to have said, 'Half the money I spend on advertising is wasted; the trouble is I don't know which half.' Digital measurement finally gave advertisers visibility into 'which half' they had been wasting. To be sure, getting there was a process: as Turow (2012) notes, the potential for computing to revolutionize advertising was actually recognized as early as the 1960's, when advertising agencies began experimenting with 'media optimization' with large mainframe computers. Later, companies such as Nielsen began using diary and survey panel methods to understand media audiences' makeup and media consumption behaviors, which marginally improved intelligence by enabling demographic data. However, these methods were laborious and expensive, and their aggregate data was only useful directionally. Measuring the actual effectiveness of ads was still largely infeasible.

Then came the internet, a Cambrian Explosion of advertising measurement. It was now possible to measure—at the level of individual users—people's behaviors (e.g. page views), intentions (e.g. search queries), contexts (e.g. physical locations), interests (e.g. inferences from users' browsing

behavior), unique identifiers (e.g. device IDs or emails of logged-in users), and more. Also, vastly improved 'benchmarking' data—information about the advertising efforts of one's competitors—became available via market intelligence services like comScore and Hitwise. Turow (2012) notes how the emergence of web browsers was a 'crucial technical development' that enabled this sea change of advertising measurement, not only because of its actual technical affordances of measurement, but also in the precedent that it set for subsequent measurement capabilities in other contexts.

In particular, the browser 'cookie'—a small file delivered imperceptibly via website code to track user behavior across pages—played an essential role. It is useful to take a moment to look closely at the story of the cookie, as it embodies well the nature of the changes that occurred in the shift to digital advertising. In Turow's view, the cookie did 'more to shape advertising—and social attention—on the Web than any other invention apart from the browser itself.' (Turow 2012) Initially, cookies were created to enable 'shopping cart' functionality on retail websites: they were a way for the site keep track of a user as he or she moved from page-to-page. Soon, however, cookies were used to track people *between* sites, and indeed all across the web. Many groups raised privacy concerns about these scope-creeping cookies, and it soon became commonplace to speak of two broad types of cookies: 'first-party' cookies (cookies created by the site itself) and 'third-party' cookies (cookies created by someone else). In 1997, the Internet Engineering Task Force (IETF) proposed taking away third-party cookies, which sent the online advertising industry into a frenzy. For years after that, though, Google disallowed—for explicitly ethical reasons—the use of third-party cookies on AdWords, their advertising system. However, a few years later, Google—and with them, seemingly, the rest of the web—ultimately relented.

The cookie proved important because, as a unique identifier at the level of the browser session, it paved the way for the emergence of identifiers at higher levels of aggregation, such as that of the device (e.g. IDFA) and even that of the user (e.g. Google Analytics' User ID). This steadily increasing ability to tie together signals about a given user across a wide array of contexts has recently given rise to the measurement paradigm known as 'attribution.' Attribution refers to the vision of comprehensive user tracking across all relevant 'touch-points' a person has with a company—including across different ads, websites, devices, apps, and even interactions such as phone calls and store visits—from the very beginning of the customer 'funnel' or 'journey,' to the end (and even beyond, if an advertiser measures e.g. customer affect or loyalty with a view to driving repeat purchases). The ultimate purpose of this unified tracking is, for advertisers, to 'attribute' credit for a purchase to the one or more specific advertising interactions that led to it. The holistic nature of the view to which attribution aspires is particularly important at present, when many advertisers feel they have captured the so-called 'low-hanging fruit' of direct-response advertising (e.g. search or referral advertising, where user intentions are well-formed), and thus are shifting their focus to the upstream (or 'upper-funnel,' as is often said) challenges of demand *generation*, rather than only demand *fulfillment*. Attribution also has increasingly important interactions with automation, as the variety and complexity of multiple-touchpoint data make manual analysis infeasible and even undesirable, and virtually require an algorithmic analytical approach.

Importantly, the presence of a user ID also enabled the rise of web analytics. The wide availability of powerful web analytics platforms—such as Google Analytics, Omniture, and Coremetrics—fostered a culture of measurement and experimentation among technology designers generally. While the degree of emphasis a given company put in running experiments on design changes varied—for an extreme case, consider Google's 'hundred shades of blue' test that

prompted a lead designer to quit—the prevalence of, at the very least, A/B and multivariate testing represented the adoption of advertising design methods by interface and technology designers generally.

It is important to point out, however, that digital measurement has taken a fork in the road that it will likely have to backtrack on sooner or later. Rather than conceiving of the 'customer journey' in terms of user *intentions*—an approach that enabled the success of 'direct response' digital advertising—measurement is increasingly replicating the pre-internet model of targeting mere *attention*. In an intent-based measurement approach, digital advertising would begin with the atomic units of intention already measured (e.g. task-level intentions, such as search queries) and aggregate upward to higher-order intentions (e.g. user goals), which would require new mechanisms of explicitly capturing them—inference would not be sufficient. such an intent-based approach would enable advertising to be just as effective while also enhancing the overall navigability of people's lives. Instead, by merely inferring insights from user behavior, and targeting their mere *attention* (e.g. seeking to maximize engagement metrics such as 'number of pageviews' or 'time on site'), measurement increasingly creates perverse design incentives for players throughout the digital ecosystem—design incentives that ultimately result in negative externalities at both individual and societal levels.

Looking ahead, as devices and applications proliferate amid the rise of such trends as wearable computing, the 'quantified self' movement, and the 'internet of things,' the vision of attribution—and the methods of advertising measurement generally speaking—will assume an even more important role. Though the promise of 'big data' is only now coming to be realized, the proliferation of networked digital sensors and interfaces are dramatically increasing the aspects of human life that can be measured.

b. Automation

Before digital media, most processes involved in creating, buying, and running ads were very manual. Now, however, the application of machine-learning and algorithmic logic has brought far greater efficiency and speed to advertising processes. For example, ad creation, agility in ad bidding/purchasing (e.g. real-time bidding, or RTB), and performance analysis, just to name a few, have been transformed by—and continue to rapidly evolve in—their applications of automation. Automation significantly influences, if not underlies, many of the other digital advertising advances I have described here, especially insofar as they involve new affordances of scale. Increasingly, aspects of digital advertising that were previously manual are being automated, such as requests for proposals (RFPs), insertion orders, and price negotiations. This paradigm has been called 'programmatic' advertising. (Marshall 2014) Looking ahead, artificial intelligence (AI) is poised to have an even greater impact on digital media analysis as well as experience design.

However, the relationship between advertising and automation is not just notable for the latter's impact on the former. The influence flows both ways. One obvious point to make here is the organizational one, i.e. that some of the most advanced work in artificial intelligence is currently being done at companies whose primary business model is advertising—and so, having an existing profit motive to tend to, it is only natural that their first priority would be to apply their innovations toward growing their business. Yet the natural (so to speak) affinity between advertising and automation goes much deeper than that.

The *type* of outcome at which advertising aims is uniquely fit for the application of artificial intelligence. The ultimate goal, i.e. the purchase (or 'conversion,' as it is often called), is a clear

binary signal: either the person buys the pair of tennis shoes or he does not. Similarly clear are the myriad behavioral signals throughout the purchase funnel that inform inference and prediction: ad views, ad clicks, and so on. In other words, the *persuasive* nature of advertising makes it an uniquely appropriate application of artificial intelligence—in particular, the combination of the mind-boggling multiplicity of its inputs with the laser-like singularity of its ultimate goal.

Perhaps this is why games have been the other major domain where artificial intelligence has been tested and innovated. Consider, for example, Google DeepMind's training of a convolutional neural network to play Atari 2600 games (a system that soon thereafter beat the world champion in the board game Go). On a conceptual level, training an algorithm to play a computer game well is extremely similar to training an algorithm to advertise well. Both involve training an agent that interacts with its environment to grapple with an enormous amount of unstructured data, and to take actions based on that data to maximize expected rewards as represented by a single variable. Perhaps an intuition about this affinity between advertising and algorithmic automation lay behind that almost mystic comment of McLuhan's in *Understanding Media*:

> To put the matter abruptly, the advertising industry is a crude attempt to extend the principles of automation to every aspect of society. Ideally, advertising aims at the goal of a programmed harmony among all human impulses and aspirations and endeavors. Using handicraft methods, it stretches out toward the ultimate electronic goal of a collective consciousness. When all production and all consumption are brought into a pre-established harmony with all desire and all effort, then advertising will have liquidated itself by its own success. (McLuhan 1964)

It is probably not useful, or even possible, to ask what McLuhan got 'right' or 'wrong' here; in

keeping with his style, the observation is best read as a 'probe.' Regardless, it seems clear that he errs in two of his assumptions about advertising: (1) the assumption that the advertising system, or any of its elements, have 'harmony' as a goal, and (2) the assumption that human desire is a finite quantity merely to be balanced against other system dynamics. On the contrary, since the inception of modern advertising we have continually seen it seek not only to fulfill existing desires, but also to generate new ones; not only to meet people's needs and demands, but to produce more where none previously existed. McLuhan seems to view advertising as a closed system which, upon reaching a certain threshold of automation, settles into a kind of socio-economic homeostasis, reaching a plateau of sufficiency via the (apparently unregulated) means of efficiency. Of course, as long as advertising remains aimed at the ends of continual growth, its tools of efficiency are unlikely to optimize for anything like sufficiency or systemic harmony. Similarly, as long as some portion of human life manages to confound advertising's tools of prediction—which I suggest will always be the case—it is unlikely to *be able* to optimize for a total systemic harmony. This is a very good thing, because it lets us dispense at the outset with imagined, abstracted visions of 'automation' as a generalized type of force (or, even more broadly, 'algorithms'), and focus instead on the particular instances of automation that actually present themselves to us, the most advanced implementations of which we currently find on the battlefield of digital advertising.

c. Customization

Before digital media, there were very few mechanisms to allow the customization of advertising messages for specific contexts or groups, let alone for individual people. Furthermore, most mass media (such as television or radio) were only used at specific times or in specific places, and were in any event not anywhere nearly as ubiquitous as the mobile phone is today. This meant that the

timeliness of ads was often quite low: an advertiser was rarely able to reach someone at the 'moments that matter,' i.e. to intervene at the moment of consumer decision-making.

Digital media dramatically increased advertising's level of personalization. In particular, the emergence of search engine advertising—and especially Google—was a profound shift. 'For the first time, advertisers could reach out to huge numbers of *individuals* as they took consumer decision journeys online' (Turow 2012, emphasis original). In their 1998 paper describing the PageRank algorithm behind their nascent search engine, Larry Page and Sergey Brin wrote that search engine advertising would be unlikely to succeed because advertising, in their view, would ultimately corrupt the integrity of the platform: 'we expect that advertising funded search engines will be inherently biased towards the advertisers and away from the needs of the consumers' (Brin & Page 1998). However, the subsequent innovation of the AdWords 'quality score,' an ads quality metric used in conjunction with the advertiser's bid to determine how high their ad would appear on the page, proved to be a key addition that largely mitigated the platform corruption that Page and Brin feared. As a result, by 2010 search advertising comprised half of all revenue on the web (Turow 2012).

The success of Google's search advertising offering illustrated the value and potential of using signals of *intention* as the primary targeting mechanisms for advertising. This is an important point with significant ethical implications that I will return to in the next section. Informally, Google's search engine was sometimes referred to by employees as a 'database of intentions.' Search queries serve as strong signals of user intention, often at extremely granular levels of specificity. As a result, they enable an unprecedented degree of personalization in ad targeting at precisely the moment that a user's intention is being expressed. This means that an advertiser—a baking supply company, for instance—could now target advertisements not only to broad demographic segments (e.g.

'women aged 40-50') or to relevant interest groups (e.g. 'people who enjoy baking'), but to anyone looking for a 'train shaped baking pan' at that very moment—and even to personalize the post-ad 'landing page' experience on their website to show all the train-shaped baking pans they had for sale. Soon after the success of Google's search advertising offering, they launched a system that enabled dynamic contextual targeting on web pages, which similarly enabled monetization of the so-called 'long tail' of websites.

The capability to personalize users' experiences has continued to increase as more inputs about users' behaviors, interests, preferences, and intentions have become available. In an advertising context, this amounts to more effective (from an advertiser's standpoint) tracking, user modeling, segmentation, targeting, message delivery, and management of the 'customer journey' over time. These advances have informed recent advertising developments such as 'retargeting' (showing a user an ad based on something they put in a website shopping cart but did not buy) and 'attribution' (user tracking across multiple touchpoints, e.g. devices or browser sessions). Furthermore, the ubiquity of internet-connected 'smartphones' enables context-aware personalization as well: for instance, an advertiser can offer a user a better price on a product that they are about to buy in a competitor's store, or offer the user a coupon when walking by their own store in order to incentivize them to come in.

Finally, we can see a similar level of personalization occurring in 'traditional' media that have become digitized, such as television or radio, where new methods have been developed to enable greater specificity of messaging (e.g., the addition of dynamic, user-specific product placements in television shows). Of course, these digitized versions of traditional media also serve as mechanisms for the collection of user data that feeds back into the system as a whole, as do the devices currently beginning to emerge under the 'internet of things' (IoT) paradigm.

IV. From 'Underwriting' to 'Overwriting'

Digital advertising has profoundly influenced the design of our information technologies, and thereby the character of our attentional lives. Its effects come in both direct and systemic forms. The *direct* effect of primary importance is the increase in the prevalence and power of persuasive advertising in our lives. This power is amplified by our lack of awareness about the nature, and often even the presence, of advertising across various contexts of digital technology use. However, while these direct effects present important ethical challenges, it is ultimately digital advertising's *systemic* effects that warrant the greater concern. Broadly, these systemic effects relate to the way in which advertising has come, via the evolutions in boundedness and intelligence described above, to occupy a place of dominance over the design logic of digital technologies generally. This dominance complicates the question of whether digital advertising should, in fact, be viewed as a form of advertising at all.

In the digital attention economy, it is often remarked that 'the user is the product.' One might add: 'and the product is the ad.' Digital content is now routinely created and served in order to serve the sole purpose of capturing user attention and selling it to advertisers in as efficient a manner as possible. In the context of web articles, one common manifestation of this is the phenomenon of 'clickbait,' or articles whose headlines are written in a gimmicky way that maximizes users' impulsive click-throughs to the article. Even if a user manages to avoid the ads on the clickbait article's page (e.g. by using an ad blocker), they still see the 'ad for the ad,' so to speak: they still get the flypaper even if they manage to avoid the swatter. Turow writes that this process 'is only beginning, but its logic and trajectory are clear. Increasingly, these publishers are channeling their masters' voices in ways that will culminate in customized content. In other words, the principle of

personalization will no longer be applied just to advertisements but will shape news, information, and entertainment as well.'

That shift has in fact already taken place. Increasingly, advertisers are directly involved in the creation of the content, which has come to be known as 'native advertising.' This ad 'format' has rapidly become standard on the web in recent years. As Einstein (2016) writes, 'as media companies have become increasingly desperate for revenues, the wall between church and state (editorial and advertising) has come crumbling down, enabling advertising to invade the editorial realm.' This has led, among other things, to the phenomenon of 'content confusion' (i.e. where you don't know whether something is an ad or not).[28] This 'content confusion' is widely encountered when companies give free products to 'influencers' on 'social media' sites, who then create a YouTube video or Snapchat post where they draw attention to, and talk about, the product or brand. Of course, at the level of the user, this practice manifests as the provision of 'free' products or services. Advertising companies in particular have made liberal use of 'free' products, not only to capture more pieces of the existing attentional pie, as with free digital services (Facebook's sign-in screen currently reassures the user that 'It's free and always will be'), but also to enhance competition for those pieces of the pie (Google Analytics is offered for free[29] because advertisers that use it tend to spend more on their Google ads, via systems which are of course tightly integrated with GA).

Web analytics systems, and especially Google Analytics, played a major role in establishing advertising's 'engagement' metrics (e.g. number of clicks, impressions, or time on site) as default operational metrics for websites themselves. In doing so, it extended the design logic of

[28] Einstein (2016) broadly describes this type of practice as 'black ops advertising,' i.e. 'creating content that grabs our attention while hiding its corporate sales pitch.'
[29] Paid versions of Google Analytics aimed at enterprise users also exist.

advertising—and particularly *a*ttention-oriented advertising (as opposed to advertising that serves users' clearly expressed *in*tentions)—to the design of the entire user experience. Principles of A/B and multivariate testing, which had for some time already been used to identify high-performing ads, became widely used to test the 'effectiveness' of many different website elements such as colors, images, navigation components, button placements, etc. Web analytics systems also became the default home for user IDs that could give advertisers visibility into user behavior across multiple 'touchpoints,' such as ad views or website visits, in order to better orchestrate their messaging strategy to users. In this way, advertising effectively set the roadmap—and defined the underlying goals and logic—for web publishers in general, and ultimately for the design of user experiences in other digital contexts as well, e.g. in the design of mobile apps or services for other types of internet-connected devices. As Turow (2012) writes, 'Advertisers have been way ahead of publishers when it comes to the interest and ability to personalize material for audiences.' Of course, this also meant that some advertisers would *become* publishers and begin offering users content *for the sole purpose* of selling ads—hence clickbait, auto-playing videos, 'arbitrage' sites, 'content farms,' and the like. Because existing content providers had to compete with these new 'content' providers—the lowest common denominators of the infosphere, in a sense—they had little choice but to stoop to this new level of pettiness and adopt similar techniques of baiting and sensationalizing in order to survive. Thus, not only the techniques and metrics of advertising took over in this new digital context, but the *goals* of advertising were now also 'overwriting' the goals of the medium itself. The 'new media-buying logic' ... 'encourages [publishers] to further retreat from longtime professional norms in the interest of packaging personalized advertising with personalized soft news or entertainment.' As a result, even those who tried to resist these systemic effects could not help being dragged down into the mud by them as well.

On an even broader level, advertising now influences the design of products that seemingly have nothing to do with advertising at all. Much forward-looking R&D at such companies has advertising as its root design incentive. For example, Alphabet's self-driving cars will free up some non-trivial amount of time in a day for the average driver, which they can use to spend even more time looking at advertising-monetized screens. Similarly, Facebook's Internet.org project aims to produce incremental gains in monetizable attention in lower- and middle-income countries by adding millions of new internet users (whose attention, conveniently enough, due to the design of the 'Free Basics' implementation of the Internet.org app is limited to the Facebook platform itself and only a handful of other websites (Facebook 2017)). However, the most important R&D effort with advertising as a dominant concern is undoubtedly artificial intelligence[30]. Neither the existing links, nor the potential interactions, between advertising and AI have been sufficiently analyzed in the light of day. Advertising is in fact a key incentive driving the development of artificial intelligence. Google's DeepMind AI recently passed a major milestone by beating the world champion in the game Go, and at the time of writing two of the world-leading companies in AI research, Alphabet/Google and Facebook, are advertising companies—both of which have the working vision of an intelligent AI assistant as the next major platform, as embodied in the phrase 'AI is the new UI,' and as seen in recent products such as Google Home and Facebook chat 'bots.'

What emerges here is a picture of 'advertising' that has drifted very far indeed from its initial definition as a proactive appeal for some resource of value made as an *exception* to a medium's dominant design goals. Here, as a result of its unboundedness and intelligence, digital advertising

[30] Neither the existing links, nor the potential interactions, between advertising and AI have been sufficiently analyzed in the light of day. Advertising is in fact a key incentive driving the development of artificial intelligence. Alphabet/Google's DeepMind AI, which recently passed an AI milestone by beating the world champion in the game Go, is already being used to increase video watch rates on YouTube. At the time of writing, two of the world-leading companies in AI research, Alphabet/Google and Facebook, are advertising companies. And both have the working high-level design vision of an intelligent AI assistant as the next major platform: 'AI is the new UI.'

cannibalized the character of the media itself. In doing so, it moved from being an exception to being *the rule*. It now *is* the default logic of information delivery which other logics, with their competing design goals, must now override.

In the past, advertising was commonly said to be 'underwriting' content already present in a medium as the result of some higher design goal (such as artistic vision or editorial judgment). In the U.S., the term 'underwriting' is still used to describe advertisements on public radio or television stations, where law proscribes stricter rules about their delivery. These rules constrain not only the placement of the so-called 'underwriting spots' (which must occur in specific sixty-second periods before or after a program) but also their design and content (which must 'mirror the production values of the program,' 'flow smoothly with program content and other packaging elements,' and be free of any 'calls to action'). The notion of 'underwriting' is thus similar to that of 'sponsorship': it is support for the delivery of a particular piece of content that is constrained from overtaking the existing design goals of that medium.

Yet now, digital advertising is not underwriting but *overwriting* the design goals of our media. That is to say, advertising is no longer subservient to the existing dominant design goals of a given medium, but rather overtakes and replaces those goals with its own. While some ethical attention has been given to the increased pervasiveness of advertising's presence across all areas of human life (Spence & Van Heereken 2005, Drumwright & Murphy 2009), very little has emphasized the deeper point about the pervasiveness of advertising's 'overwriting' of the design goals of the media themselves.

In light of this dramatic shift from being an exception to being the rule, it is reasonable to ask whether digital advertising should be considered a type of advertising at all. It certainly does not fit

the definition of advertising I have given above. Instead, digital advertising now functions as the emergent, dominant design logic of our information environment. Other design logics, which optimize to other goals or values, now play the role advertising played prior to the internet, i.e. as exceptions to the rule. In order to compete with the design logic of online advertising for control over the structure of digital media, these other design logics will have to apply the same techniques of persuasion, measurement, and optimization.

I will not attempt to render judgment on this definitional question here, but will simply note that the more intentional we can be in engineering and aligning our terminologies for advertising—and, indeed, our terminologies for persuasive design generally—the more effective we will be at seeing, discussing, and ultimately reforming the ethical design of persuasive systems.

V. Some Ethical Implications

As I have described, digital advertising radically departs from advertising as it has historically been understood, especially in ways that relate to its boundedness and its intelligence. As a result, a corresponding shift in ethical focus is needed to properly consider the ethical implications, considerations, obligations, and opportunities this view brings into the foreground. Drumwright & Murphy (2009) write of digital media broadly that 'the ethical issues presented in new media and nontraditional media are different in kind,' and this is even more the case in the context of digital advertising due to the new role these differences of boundedness and intelligence allow it to play in setting the agenda for the design goals of digital media more broadly.

The Simonian Inversion: A Corrective

It is difficult to overstate the degree to which ethical analysis of digital advertising, like that of digital technology design generally, has been hindered by the widespread failure to take into account Herbert Simon's observation that information abundance results in a scarcity of attention. To encounter a project such as digital advertising that is ultimately concerned with persuasion, and then to engage it primarily as a project of data collection and management, which it only incidentally is, requires extraordinary force of either will or habit.

Yet in order to fully apply the lessons of the Simonian Inversion to the ethics of digital advertising, we must clarify what it is actually telling us. In particular, it is essential to note that information abundance is in fact a ratio concept. Abundance can only be abundant relative to some threshold. Here, the relevant threshold for abundance is not *historical*: i.e., we are not primarily interested in the amount of information available today relative to the amount that was available at some point in the past. Rather, the relevant threshold is *functional*: it is the amount of information that can be well processed given existing limitations. The fact that this threshold takes the form of a process means that we must understand information abundance as having a temporal component in addition to its quantity component. To illustrate, consider the video game Tetris, where the rain of blocks that waits off-screen for you to stack them is infinite—but their infinitude is not the problem. What really does you in is their increasing speed. Information quantity *as such* is only important insofar as it enables its velocity. And, ultimately, 'there is no competition against instantaneousness.' (Wiman 1899, qtd. in Marvin 1990)

Thus the necessary rider to Simon's observation about the relationship between information and attention is that its casting of attention as a substantive, rather than a procedural, element proves to

be too metaphorical, and potentially too problematic, to adopt by default. This is because viewing information abundance in process terms suggests that the main sort of risk it poses is not that one's attention will be *occupied* or *used up* by information, but rather that one will *lose control over* one's attentional processes. In other words, to return to the Tetris metaphor, the problems arise not when you stack a brick in the wrong place (though this can contribute to self-regulatory problems down the line), but rather when you lose control of the ability to direct, rotate, and stack the bricks altogether. The important risks here are not substantive, but procedural; they pertain to the management not of resources but of capacities. Framed in terms of its ethical implications, this means that information abundance does not lessen attention *per se* but rather one's *control over* their attention. The relevant ethical challenges here are thus primarily challenges of self-determination.

Self-Determination and the Freedom of Attention

Setting our ethical radar to detect issues of attention management, rather than issues of information management, does not require us to redesign our ethics from the ground up. Nor does it present insoluble challenges for the freedom of expression. Quite the opposite, in fact: it creates a space for us to reaffirm that, insofar as self-determination is our concern, freedom of speech *depends on* freedom of attention. In *On Liberty* (1859), Mill writes that the 'appropriate region of human liberty … comprises, *first*, the inward domain of consciousness,' a domain where relevant liberties include 'liberty of thought and feeling; absolute freedom of opinion and sentiment on all subjects, practical or speculative' (emphasis mine). For Mill, the external freedom of information management, such as 'the liberty of expressing and publishing opinions … [rests] in great part on the same reasons, and is practically inseparable from it.' This point—that the freedom of expression depends on the freedom of attention—is particularly important here because many objections to proposed restrictions or regulations on advertising have been based in appeals to free

speech. Also, this liberty of the 'inward domain of consciousness' is, importantly, not only a freedom of beliefs but also a freedom of intentions; it pertains not only to one's immediate perceptions but also one's long-term goals and aspirations. Mill continues: 'this principle requires liberty of tastes and pursuits; of framing the plan of our life to suit our own character.' Here Mill is articulating a liberty of life navigation that applies across all scales of the human experience.

We could read Mill here as articulating something akin to a 'freedom of attention'—a necessary complement to the freedom of expression, and a crucial aspect of self-determination. This view of attention, as one's life-navigation capacities across the domains of both thought and action, aligns with Hegel's (1820) view of the will: 'Those who treat thinking and willing as two special, peculiar, and separate faculties … show from the very start that they know nothing of the nature of willing.' It also resonates well with an active-externalist view of experience, as seen in Noë (2004): 'Experience is a dynamic process of navigating the pathways of ... possibilities. *Experience depends on the skills needed to make one's way*' (emphasis mine).

VI. Digital Advertising and Self-determination: Challenges, Opportunities, Obligations

As a framework for describing the ethically salient effects of digital advertising on attention, I will draw on the 'three types of distraction' that I outlined in the previous paper, which take their inspiration from Frankfurt's structure of the will. They are: (a) Functional Distraction, influences that frustrate desired *action* and hinder the user from doing what they want to do; (b) Existential Distraction, influences that frustrate desired *identity* and hinder the user from being who they want to be; and (c) Epistemic Distraction, influences that frustrate essential *capabilities* and hinder the user from wanting what they want to want.

The aspects of digital advertising I have described above raise ethical challenges, and present new obligations, in the following areas:

Choice & Capacity

Advertising plays a valuable role in society when it enables people to make better choices. The International Advertising Association (IAA)'s 'Case For Advertising' campaign has sought to associate consumer perceptions of advertising with such effects: 'What advertising provides - over and above any specific product or service – is choice. To express that core concept the IAA developed the central theme for the campaign: Advertising. Your Right To Choose.' (IAA 2014)

Yet advertising does not always lead to better choices. On one level, this is because what counts as 'better' is almost always defined and measured from the perspective of the advertiser, not that of the viewer. On a deeper level, though, is the fact that having *more* choices does not necessarily mean one is able to make a *better* choice. In information-scarce environments, we have generally viewed more options as meaning better choices. However, according to the phenomenon known as the 'paradox of choice' (Schwartz 2004), there is an optimal number of options we can consider before the cognitive-effort cost of considering additional options becomes greater than the incremental benefit of choosing a different option. Selecting from five types of cereal at the supermarket is a choice; selecting from five hundred is a burden. For Raz (1986), diversity is an essential consideration in personal autonomy: 'Clearly not number [of options] but variety matters.'

Willpower is a finite capacity of particular importance here. The so-called ego-depletion hypothesis is 'the idea that self-control is a limited resource, which is depletable in the short-term and which only gradually returns to its initial level' (Levy 2007). We have a finite number of decisions we can

make in a given day, after which we become more akratic and susceptible to persuasive appeals. Advertisers routinely exploit this by, e.g. 'requiring potential purchasers to engage in self-control tasks, which deplete their resources, before they are presented with the option of purchasing' (Levy 2007). Often, even regulatory interventions conceived from a perspective that is information-centric, rather than attention-centric, contribute to willpower depletion and thus have the opposite of their intended effect.

For example, in the EU, website owners are required to get users' consent to use tracking cookies, a law intended to protect user privacy and increase transparency of data tracking. Yet from an attentional perspective, this intervention asks a user to make, say, thirty more decisions per day (assuming they access thirty websites per day), which amounts to a non-trivial strain on their cognitive load. Furthermore, there is the fact that the 'cookie consent' notifications on websites can be (and have been) *designed to maximize compliance*: website owners have simply treated the request for consent as another persuasive interaction, and have deployed methods of measurement and experimentation used to optimize ad performance in order to manufacture users' consent. In fact, according to the user-centered definition of advertising I provided above, these EU cookie-consent notifications *are* advertisements.

There are other ways in which giving someone a choice can serve to undermine their autonomy. Walton Hamilton, in his introduction to Rorty (1934), observed, 'Business succeeds rather better than the state in imposing its restraints upon individuals, because its imperatives are disguised as choices.' Today, when governments are getting somewhat better at doing so, we understand this idea in terms of 'choice architecture' (Thaler & Sunstein 2008). In some cases, undermining can occur when the choice is *not* actually a choice at all, as in 'placebo' crosswalk buttons that illuminate the 'WAIT' sign but do not have an influence on when the stoplights change color. At other times,

there *is* a choice, but it has been architected in such a way that you are more likely to choose the route the designer wants, such as in the way users on YouTube are more likely to watch and engage with video ads they can skip than ones they cannot skip. (Ipsos/Innerscope 2011)

The essential point I am making in these examples is that, from an attentional perspective, it is not the *substance* of the choices that advertising enables which matters most, but rather how the *structure* of the choice architecture affects one's choice-making capacities. In an information-abundant world, choices are plentiful: millions of options are only a click away. Thus advertising's unique value proposition can no longer be that it brings us new choices; the world now does that for us.

In *The Morality of Freedom*, Raz (1986) uses the scenario of 'The Hounded Woman' to describe how capacity depletion threatens autonomy:

> A person finds herself on a small desert island. She shares the island with a fierce carnivorous animal which perpetually hunts for her. Her mental stamina, her intellectual ingenuity, her will power and her physical resources are taxed to their limits by her struggle to remain alive. She never has a chance to do or even to think of anything other than how to escape from the beast.

Raz writes that the Hounded Woman lacks personal autonomy because she lacks 'an adequate range of options to choose from.' Importantly, the 'lacking' here does not mean the *absence* of a range of options, but rather the *inability to give attention* to them. This is analogous to, say, having all the world's information at your fingertips, but then being 'hounded' around the internet by stimuli that appeal to your baser, more impulsive motives.

Navigability

Attention is a flexible construct that can be described via different metaphors depending on the particular purpose or context. One useful way of talking about attention is as the set of innate human navigational capacities over both the short and long-term (as I have done throughout this thesis, and most explicitly in the introduction). Such a view broadly aligns with an active externalist view of experience, on which navigability may be seen as a sort of 'knowing-how' (Noë 2004). It also enables us to draw useful comparisons between questions of physical and informational navigation.

Take, for example, the 2006 decision of the city of Sao Paulo, Brazil to ban outdoor advertisements via a measure called the 'Clean City Act.' Many of the reasons given to justify this regulation were to be expected: to increase aesthetic well-being, 'environmental comfort,' the cultural/historical value of the city, and quality of life. Yet another justification for banning ads was to improve the perception of *reference points* in the city's landscape of the city in order to aid people's navigation (Pires 2013).

Viewing the question of attention in navigational terms reminds us that more information does not necessarily help people navigate their world, and that the concept of informational 'relevance' contains several different sub-considerations. For instance, even if we were to grant that all advertising were purely informational (rather than persuasive) in nature—as well as that in all cases advertisements do meaningfully increase the number of choices a person could consider—there would remain the further questions of (a) whether the person wants to consider *this* particular choice in the first place, and (b) whether they want to expand their field of options for that choice in *this* particular way. In other words, the question of ad 'relevance' includes not only the relevance

of the ad to a particular goal, but also the relevance of the goal to the person's higher goals, the timeliness or frequency of the communication, the way in which it inhibits pursuit of other immediate tasks, and so on.

But the Sao Paulo example is not quite yet an appropriate comparison for digital advertising. An appropriate comparison would be not merely if ads appeared on billboards, but if the ads reshaped the streets to take you past a particular store, if they determined how much traffic was on the road around you, etc.—and could do this in a fluid, automated way in real-time.

Goal Transparency

In the context of digital advertising ethics, the concept of transparency has of course come into play primarily in questions of privacy and consent vis-a-vis user data collection and management. Beyond that, however, it has been an important concept in discussions about where and how one ought to signal the ad/non-ad boundary to the user. While most ads running on the major digital ad networks usually *technically* satisfy some common standard of transparency in ad labeling (e.g. by noting that a particular piece of content is 'Sponsored' or an 'Ad'), these signals are almost always designed to minimize the transparency they exist to produce. Notices that are technically 'transparent' (i.e. available in some form for viewing) themselves become inputs for experimentation, continually 'optimized' until the particular form of 'transparency' that maximizes user compliance with the persuasive design goal is found. (This effect also be seen in websites' implementations of the EU 'cookie consent' notification.)

A major implication of this opacity-via-distraction is that without a signal about where the ad/non-ad boundary lies, there are no (or very few) signals about the design logic—about the

why—of the products or services in use. This also means, in a more general sense, that there are *no/few bases for trust*. In *Meditations* (8.14), Marcus Aurelius says, 'Whatever man you meet, say to yourself at once: 'what are the principles this man entertains as goods and ills?'' This is good advice for when we 'meet' new technologies as well. What is Facebook's persuasive goal for me? What metric does Twitter aim to maximize with my time use? Why *did* Amazon build Alexa, after all? Do the goals that my trusted systems set for me align with the goals I have set for myself? As the Russian saying goes, the answer is to find a way to 'trust, but verify.'

However, as regards advertising transparency, there is an even more fundamental question lurking here. Most users fail to realize that advertising is the business model of the digital products and services they use to begin with. (Most people do not even realize this about *newspapers*, which are arguably the oldest ad-supported medium of all.) What are the obligations alert users to the basic nature of the economic tradeoff they're making—to the fact that they are not the user, but the product? Furthermore, who bears these obligations—advertisers, publishers, platforms, or some other party within society? There is an important meta-issue of transparency here regarding the nature, purpose, and mechanisms of advertising in society broadly speaking—and a conversation that urgently needs to be advanced.

Measurement and Care

Ethical discussions about digital advertising often assume that limiting user measurement is axiomatically desirable due to considerations such as privacy or data protection. These are indeed important ethical considerations, and if we conceive of the user-technology interaction in informational terms then such conclusions may very well follow. Yet if we take an attention-centric

perspective, as I have described above, there are ways in which limiting user measurement may complicate the ethics of a situation, and possibly even actively hinder it.

The vision of 'attribution' in advertising measurement that I described above ultimately aims at producing a 'full view' of the user that can inform advertising actions: a view across e.g. devices, applications, and contexts of use. In principle, unifying one's model of a user so that it more faithfully reflects who they are ought to be a *good* thing. Understanding someone better can be the basis not only for acting more wrongly toward them, but also for acting more rightly. One reason stereotyping is seen as as objectionable, for instance, is because it subjugates a person's individual characteristics to potentially incorrect assumptions about some group of which they may be a part. Similarly, if a system has an impoverished set of measurements about a given user, then it must infer their relevant attributes on the basis of demographic or similar aggregations. However, if improved measurement could help an advertiser seeing a user less as a 'cookie' and more as a person, would that not at core be a process of humanization and increased understanding? Is not greater measurement—of the right things—in essence the process of ceasing to see someone as an *it* and beginning to see them as a *thou*?

If so, the key ethical question we should be asking with respect to user measurement in digital advertising is not *merely* 'Is it ethical to collect more information about a user?' (though of course in some situations that *is* the relevant question), but rather, What information about the user are we not measuring, that we have a moral obligation to measure?' If the vision of 'attribution' truly aims to get a full view of the user, it is precisely the role of advertising ethics to ask, 'What does it *mean* to get a full view of the user?' At present, this question is being asked and answered by advertising practitioners but is being virtually ignored by ethicists.

Greater measurement (of the right things) is in principle a good thing. Measurement is an advertiser's primary method of attending to the user, and as such can serve as the ground on which conversations, and if necessary interventions, pertaining to their moral responsibilities may take place. The paradigm of 'attribution' serves as a natural conceptual waypoint where ethicists and advertising practitioners can meet and enter into this conversation about 'measurement as care.' As one industry leader in the study undertaken by Drumwright & Murphy (2009) said, 'When Ray Kroc invented McDonald's, he didn't know how bad high-fat food is for children. With the accumulation of knowledge comes the responsibility to respond to that knowledge.'

What are the right things to measure? One is potential vulnerabilities on the part of users. This includes not only signals that a user might be part of some vulnerable *group* (e.g. children or the mentally disabled), but also signals that a user might have particularly vulnerable *mechanisms*. (For example, a user may be more susceptible to stimuli that draw them into addictive or akratic behavior.) If we deem it appropriate to regulate advertising to children, it is worth asking why we should not similarly regulate advertising that is targeted to 'the child within us,' so to speak.

In general, however, the most ethically important thing for advertisers to measure is probably user intent. The way in which search queries function as signals of user intent, for instance, has played a major role in the success of search engine advertising. Broadly, signals of intent can be measured in forward-looking forms (e.g. explicitly expressed in search queries or inferred from user behavior) as well as backward-looking forms (e.g. measures of regret, such as web page 'bounce rates').

As an example of how this can enhance the ethical operation of advertising, take the example of 'retargeting.' Retargeting is a form of digital advertising in which certain non-converting user actions near the point of purchase (such as a user placing a pair of hiking boots in a website's

shopping cart without purchasing them) are used to inform subsequent ad serving (e.g. the user subsequently sees an ad for that specific pair of hiking boots on their favorite camping website). Retargeting often crosses what Eric Schmidt has called the 'creepy line,' and has prompted objections that largely pertain to data collection (Jerome 2010). In fact, when retargeting is properly deployed, it is potentially a much more ethical way to advertise from an attention-centric perspective. To be properly deployed, retargeting would need to measure and respond on the basis of the user's true intentions. The problems with retargeting emerge when it fails to measure and respond to such intentions in an accurate enough manner: it too often infers intentions from users' behavioral data, rather than asking them directly. In such cases, the ads end up in an 'uncanny valley' of relevance from the user's perspective: they are relevant enough to make the user aware of them (and aware that the system is aware of their behavior), but not relevant enough to provide value.

From an information-centric perspective, digital advertising measurement is a risk to be mitigated and minimized wherever possible. This is today the standard perspective, and it is not incorrect. It is, however, incomplete. The information-centric view must be balanced by an attention-centric one; indeed, advertising systems only manage users' information with a view to ultimately managing their attention. If we grant that advertising can influence users' attention in ethical (and not just unethical) ways, and if we acknowledge that user measurement serves as the basis for the user modeling that informs the (increasingly algorithmic) decisions made about whether, how, and toward what ends they may be persuaded, then it is the responsibility of ethics to show not just where advertising measurement must be curtailed, but also where it must be advanced. 'The nature of moral judgments,' Susan Sontag wrote, 'depends on our capacity for paying attention.' (Sontag 2007) This is equally true for human and artificial agents. Helping advertising make moral

judgments requires not only diminishing its capacity for 'focusing on the user' in some places, but also enhancing it in others.

VII. Conclusion

In *Common Sense* (1776), Thomas Paine wrote, 'a long habit of not thinking a thing wrong, gives it a superficial appearance of being right.' My contention is that this has been the case with the practice of advertising during its rise to cultural prominence over the past century—but that now, in the era of digital technology, its 'superficial appearance' of rightness is in dire need of reevaluation. On the basis of the user-centered definition of advertising I advanced in the previous chapter, I have here considered the nature and dynamics of digital advertising and showed that by virtue of its radically altered boundedness and intelligence, it does not function as an 'exception to the rule,' as advertising has historically done. Rather than 'underwriting' media design, digital advertising now 'overwrites' it and installs its own aims as primary design goals. It is therefore arguable whether 'advertising' is an appropriate description of this media dynamic now running rampant across the infosphere. Finally, I discussed several direct as well as systemic effects of these dynamics and the implications they have for user self-determination—implications which have received far too little attention to date, and which therefore merit a sustained project of ethical scrutiny, debate, and intervention.

Conclusion:

Stand Out of Our Light

'It is disgraceful to be unable to use our good things.'

—*Aristotle,* Politics *(VII. XIII. 19)*

I. Diogenes and Alexander

On a sunny day in Corinth in the fourth century BC, a fabled meeting took place between two giants of the age: Diogenes of Sinope, philosopher and founder of the Cynics, and Alexander the Great. Diogenes, having vowed himself to poverty, lived in a ceramic barrel in his local marketplace. He was notorious for his rude, outrageous, and offensive behavior: today we would no doubt call him a 'troll.' Yet Diogenes was deeply admired by Alexander, who was arguably the most powerful person in the world at the time. According to various sources (e.g. Diogenes Laertius vi. 38; Arrian VII.2), on the day in question Alexander, flanked by his retinue, fawningly approached Diogenes in the market and offered to grant him any wish he desired. Diogenes, who was reclining in the sun at the time, looked up at Alexander and replied, 'Stand out of my light!'

Alexander's offer to Diogenes expresses a certain imperial optimism that is familiar to us from the way digital technologies have entered our lives in recent decades and offered to fulfill all manner of our needs and wishes. Of course, in many ways they have done so—extremely well. Our digital technologies have profoundly enhanced our ability to inform ourselves, to communicate with one another, and to understand our world. And yet, as they have come to envelop our day-to-day lives, we have begun to realize that they, like Alexander, have been 'standing in our light,' so to speak.

For all their *informational* benefits, they have imposed serious *attentional* costs by diminishing certain capacities that enable us to navigate our lives—capacities so essential for human flourishing, that without them technology's other benefits stand to do us little good.

Alexander did not know he was standing in Diogenes' light because it did not occur to him to ask. He was focused on *his* offer and *his* goals, not Diogenes' goals or what was being obscured by his offer. In the same way, the creators of our digital technologies do not know they are standing in our light because it does not occur to them to ask: they have focused on *their* goals and *their* desired effects, rather than our goals or the important 'lights' in our lives they may be obscuring.

We still have no good single name for this 'light,' this set of capacities, that is increasingly being obscured. The word 'attention,' as far as I can tell, is the best we can do for now. In the sense I have meant it here, 'attention' refers to a wider set of capacities than its day-to-day usage typically allows. In this wider sense, attention extends well beyond what cognitive scientists have called the 'spotlight' of attention, or our immediate perceptual capacities for managing awareness and action in the task domain: it encompasses the broader capacities we use to navigate our life over longer timeframes and in the light of of our higher goals and values. In fact, attention in this wide sense even includes the deeper set of foundational capacities, such as reflection or metacognition, that enable us to define our goals and values to begin with. Ultimately, this expanded sense of 'attention' converges on conceptions of the human will. In doing so it resonates with one of the earliest psychological treatments of attention, that of William James (1890), who even described 'effort of attention' as 'the essential phenomenon of will.' However, I am not arguing here for a view of attention that is coextensive with the will, nor has my task here been to explore the boundaries of its definition. For our present purposes, it is enough to think of this wider view of attention as the 'full stack' of navigational capacities across all levels of human life.

For all its informational benefits, the rapid proliferation of digital technologies has compromised attention, in this wide sense, and produced a suite of cognitive-behavioral externalities that we are still only beginning to understand and mitigate. The enveloping of human life by information technologies has resulted in an informational environment whose dynamics the global persuasion industry has quickly come to dominate, and, in a virtually unbounded manner, has harnessed to engineer unprecedented advances in techniques of measurement, testing, automation, and persuasive design. The process continues apace, yet already we find ourselves entrusting enormous portions of our waking lives to technologies that compete with one another to maximize their share of our lives (and, indeed, to grow the stock of life that is available for them to capture).

In this thesis I have sought to identify and describe certain 'lights' which, though newly imperiled by the dynamics of the digital attention economy, have so far remained in the periphery of ethical analysis and societal conversation. This project was originally motivated by the question of whether the increasing 'persuasiveness' of digital technologies—owing in large part to the 'attention economy,' or the environment in which design is incentivized to capture and exploit human attention as effectively as possible—poses a serious threat to people's abilities to make their lives go well. Because of the cross-disciplinary nature of this problem space, each chapter has approached this question via different, yet complementary, angles of attack. In Chapter One, I took as my angle of attack the common notion of 'persuasive technology.' There I clarified the nature and role of intentionality, and also untied the knot of descriptive and normative considerations that have thus far complicated the dominant definitions of 'persuasive technology.' In Chapter Two I shifted to a perspective rooted in questions of 'attention' and 'distraction.' This effort led me to expand the language of these concepts in a way that gives them applicability across

a wider range of freedom and autonomy considerations—across the 'full stack' of the human will, so to speak. This involved, in particular, drawing new distinctions between three different types of 'distraction,' as well as between ethical objections rooted in autonomy concerns and objections that appear to concern autonomy but in fact pertain to issues of dignity. Then, in Chapter Three, I turned to the perspective of advertising, a force which of course plays a pivotal role in the attention economy. Here I argued for the need to begin giving advertising the ethical attention it has long deserved—attention which ought to adopt a patient-centered, rather than an agent-centered, ethical perspective. Furthermore, because a patient-centered ethical perspective demands a patient-centered (i.e., user-centered) definition of advertising, I developed what is, to my knowledge, the first such definition. Then, on the basis of that definition, in Chapter Four I turned to specifically *digital* advertising and argued that it departs in several ethically salient ways from advertising as it has historically existed. In particular, we may identify in digital advertising unprecedented degrees of boundlessness and intelligence, such that it has moved from a position of 'underwriting' the design goals of a medium—i.e. serving and supporting them—to one of 'overwriting' them, and replacing the medium's higher purposes with its own. As a result, the digital systems that carry the name 'advertising' have evolved into something qualitatively very different from 'advertising' as it has been historically understood.

The picture that emerges here is that of a pervasive, persuasive media environment that is, by and large, adversarial with its users: a navigation system for our lives that is not on our side. For all the informational benefits it has brought us, the rapid proliferation of digital technologies has compromised our attention, in this wide sense, and produced a suite of cognitive-behavioral externalities that we are still only beginning to understand and mitigate. The enveloping of human life by information technologies has resulted in an informational environment whose dynamics the global persuasion industry has quickly come to dominate, and, in a virtually unbounded manner,

has harnessed in order to engineer unprecedented advances in techniques of measurement, testing, automation, and persuasive design. Already we entrust enormous amounts of our waking lives to these technologies that compete to maximize their share of it. There is still no single term that satisfactorily captures the nature, scope, and importance of what is at stake here; ultimately, it is nothing less than the continued integrity and success of the human will. The term 'attention economy' will have to suffice for now, at least in its shorthand usage which refers broadly to the total environment of attention competition (as opposed to solely its economic aspects *per se*). Similarly, the cognitive-behavioral externalities of the attention economy lack a good name. Heretofore erroneously minimized as 'distractions,' these externalities have broad consequences for all aspects of the human will. Additionally, they are ultimately symptoms of a large and complex *infraethical* challenge, i.e. they result from the ethics of the infrastructure or environment as a whole, as opposed to the ethics of individual agents' actions (Floridi 2013).

Doing justice to the solutions here, i.e. covering the many 'infraethical' interventions that would be necessary to improve or even transcend the attention economy—would require an entire thesis of its own. However, I will briefly sketch a high-level outline of how we might begin to think about the nature and structure of such a project, as there are many areas in dire need of further research that even the flash of a quick glance could usefully help illuminate.

II. Asserting and Defending the Freedom of Attention

Ultimately, responding to the moral challenge of the attention economy requires us to assert and defend our freedom of attention. *Asserting* the freedom of attention means developing its linguistic and conceptual foundations. While doing so fully is beyond my scope here, broadly speaking this effort would involve developing a broadened conception of 'attention' which could serve as a hub

uniting the spokes of existing literatures that address these questions in terms of freedom, autonomy, dignity, etc. (a glimmer of which one may discern in my brief discussion of Mill, Raz, et al. in the previous chapter). This conceptual groundwork would be essential for *defending* the freedom of attention, i.e. embarking on a project that aims to reform, and perhaps even to transcend, the attention economy, in order to render our informational environment fit for human habitation.

What would such a project aimed at reforming the attention economy look like? Again, there is much to say here, and I can only sketch a brief, broad outline. However, first it is important to note what such an effort must take care to *avoid*. There are many pitfalls that could easily complicate such a project. First, it is essential that we reject the impulse to ask users to 'just adapt' to the status quo of the attention economy. Throwing more self-regulatory burdens onto the backs of users whose self-regulation is already under siege can only amplify the problem. 'Media literacy' is one such imagined adaptation that is particularly worth moving past as quickly as a possible, at least as a possible systemic solution to these problems. Second, we cannot reply that if someone does not like the choices on technology's menu, their only option is to 'unplug' or 'detox'—this is a pessimistic and unsustainable view of technology. Third, we must take care not to unnecessarily constrain the ethical conversations about attention and persuasion within the overly narrow boundaries of of 'addiction,' nor unconsciously revert back into informational, as opposed to attentional, framings of this problem. Finally, we must of course also avoid the false hope that the attention economy might somehow correct itself.

Because the major ethical challenges of the attention economy are *infraethical* in nature, any project of reform would do well to adopt a systemic view. Such a view could, for example, consist of a simplified version of Meadows' (1997) framework of 'Places to Intervene in a System,' which helps

model and prioritize potential intervention points likely to have the greatest influence on the system as a whole. In the following table, for instance, Meadows' nine types of leverage points have been compressed into four categories (in order of decreasing importance). Listed alongside each are potential infraethical interventions specific to the attention economy:

1. Paradigms *(Advance conceptual, normative, and linguistic toolsets)*	**(a) Technology and Design** — Transcend view of technology as 'neutral'; challenge *efficiency* as default design value (identify 'useful latencies'); up-level User-Centered Design from task to life domain; advance 'infraethics' perspectives; clarify advertising's nature and purpose in information-abundant world; re-invigorate advertising ethics **(b) Attention** — Clarify and advance philosophy/ethics of attention; further develop the relationship between attention and autonomy; advance/standardize the language of attention and persuasion; accelerate shift to perspective of attention, not information, scarcity **(c) Ethics** — Transcend the 'branding problem' of ethics in industry; shift from agent- to patient-centered ethical perspective; move beyond taking 'choice' to be the primary consideration in the design ethics of attention (e.g. as in the inordinate focus on 'addiction') **(d) Humanity** — Move beyond the human-as-computer metaphor; develop new ways of talking about the impulsive/non-rational self; de-euphemize the user (use human words for human beings)
2. Goals and Metrics	**(a) User intent** — Measurement of users' tasks, goals, and values; signals of user regret, context awareness

(Measure what we value, rather than simply valuing what we measure)	**(b) Wellbeing/Fulfillment** — Measurement/prioritization of overall benefit *given*, not just the benefit *received* **(c) Externalities** — Monitor/minimize other unwanted behavioral 'pollution' (for use as anti-goals) **(d) 'Measure the Mission'** — Set expectation that companies openly measure success toward their mission statements & use these metrics as upstream determinants of design
3. Rules *(Align design incentives w/ human goals, values, & other ethical priorities)*	**(a) Incentives** — Develop and test post-advertising business models (e.g. microtransactions); incentivize advertising that targets *in*tention rather than *a*ttention; investor incentives, internal company incentives (e.g. rewarding, rather than punishing, raising design ethics concerns); **(b) Accountability Mechanisms** — Policy & ethics bodies (e.g. governmental bodies, NGOs, company ethics boards); ratings/certification bodies (e.g. reputation systems for persuasion); ethical oaths/codes (e.g. 'Hippocratic Oath' for designers) **(c) Ethical/Legal Constraints** — New types of corporate structures (e.g. B-corps), duty of care for designers, pro-ethical design patterns, principles, and 'best practices' (e.g. respect for user attention/goals, reflection by design, 'moral imagination,' participatory design) **(d) Counter-Technologies** — Advance mechanisms for asserting attentional control at the user level (e.g. do-not-contact registries, ad blockers)
4. Information Flows (including *feedback loops*)	**(a) Transparency** — Expect/mandate transparency of persuasive design goals and values (especially in automated systems); challenge when divergent from marketing messages; develop 'organic' label for technologies that align with user values

(Catalyze conversation throughout society and across silos)	**(b) Conversation** — Elevate technology criticism to place of literary/art criticism; bridge industry/academia design ethics 'silos'; enable & foster design ethics advocacy campaigns; promote/incentivize *grounded* science fiction **(c) Design** — Conversational design patterns (explicitly asking, and not just inferring, user intent); design not merely 'smart' but 'wise' technologies; context & intention awareness; measurement as care

Having a common framework of this sort will be necessary, it seems to me, for any effort aimed at reforming the attention economy to proceed with any clarity. It is also worth pointing out that the interventions here are not limited to designers, or even companies, but indeed span all corners of society: everyone has a stake in the shape of our informational environment. Finally, I would suggest that an early, in-depth project of risk identification ought to be one of the highest priorities for anyone embarking on such a system-wide project of reform of the attention economy. Among those risks, undoubtedly one of the greatest—if not *the* greatest—is that such a project aiming to reform the attention economy, insofar as it requires persuading others via digital media, could become utterly subjugated to the very media dynamics it seeks to change. How can one effectively speak out against the attention economy without themselves becoming a piece of clickbait? There is no 'escape hatch' on the attention economy, no neutral position from which to analyze it as a coherent whole. Perhaps persuading or criticizing via *form*, rather than via *content*—i.e., opting for avenues of parody or satire, rather than direct argumentation—may prove to be the more fruitful route. But the answers here are still far from clear.

III. The Benefit of Competence

Churchill famously remarked that 'the empires of the future are the empires of the mind.' Historically, the forces that have had the greatest power to shape our lives—and, thus, the forces against which we have most often struggled—have been forces in the foreground, forces whose activities and effects we are most likely to notice. This attentional foreground has been the implicit domain of politics, the visible stage upon which corporeal empires of borders, bodies, buildings, and bombs have emerged and engaged and expired. Yet for all their power, these political forces have been limited in the depth and degree of skill with which they can operate directly upon our attentional faculties and alter the course of our will. Today, however, our media systems *do* shape our attentional processes at this unprecedented degree of intimacy. Having effectively done an 'end-run' around all other political systems, our digital technologies now serve as total companion systems for our lives of an unprecedented sort. There is no real analogue in human history for the monopoly of the mind they now enjoy—especially on the scale of billions of human beings. Perhaps we could view the totality of influence these technologies have over their users as akin to that of religious initiates who carry their holy books everywhere they go, or the memorization of complete Homeric epics in the Greek oral tradition, or the day-long recitations of Buddhist mantras under one's breath, or the isolation and brainwashing practices of cults, or the experience of working at a technology company and 'drinking the kool-aid,' as they say, of banal corporate jargon and value systems, or the relentless propaganda machines of totalitarian states. In any event, we must dip into the religious, the mythic, the totalitarian—into environments where, for better or worse, the individual will is 'overwritten' by some great ineluctable power, whether by coercion or persuasion or some amalgam of the two—to find any remotely appropriate comparisons. However, in the end, even these comparisons fail to capture the heights of these new offices of political power occupied by those who now grip the reins of all the world's attention in their hands.

In fact, the only way these forces could have greater domain over human attentional processes at scale is if they were to intervene physiologically or chemically to shape our experience. (Of course, this dream is already under development. Remember: 'there is no competition against instantaneousness.') Needless to say, we have not been primed, either by nature or habit, to notice, much less struggle against, these new forces that have installed themselves as the new formal cause of our thought and action, the new stage on which we act out the stories of our lives, newly written in a room somewhere we cannot go, and by people we cannot petition. Yet these effects, these stories, are already with us, already being lived by us, and thus this new stage of technological influence over our lives is now the ground of first struggle for our political future. Thus the mere existence of this thesis, even if it were only a dim spark of light in the dark, would serve a purpose, and open the possibility of a catalyst to more light—light that so far has been eerily absent, attention to the situation that has so far been pillaged wholesale by the forces of the situation itself.

The unstated corollary of Churchill's maxim is that the *freedoms* of the future are the *freedoms* of the mind. His future was the present we now struggle to see. However, when the light falls on it just right, we may clearly discern the clear and urgent danger it poses to our freedom of attention.

Yet even amidst the general urgency of this struggle, we can identify a task whose urgency ought to make it the highest-priority question for anyone working in this domain. That task is to determine whether there exists a 'point of no return' for human attention in the face of technological distraction. Is there a point at which our essential capacities—such as reflection, metacognition, reason, or intelligence—might be so undermined that we would become unable to reconstitute them, and thereby become unable to bootstrap ourselves back into possessing our freedom of attention? If so, it is imperative that we understand what that point of no return would be, and then take all available steps to ensure that we do not pass it. In identifying this threshold of

minimally necessary capacities worth protecting, we find a fitting analogue in Roman law, in the 'benefit of competence' (*beneficium competentiae*), which designated the set of belongings that an insolvent debtor could not have confiscated from him as payment for his debts: e.g. his tools, personal effects, and other items necessary to enable a minimally acceptable standard of living, and potentially even to bootstrap himself into thriving. Absent the 'benefit of competence,' a Roman debtor could find himself ruined, financially destitute. In the same way, if a point of no return for human attention does in fact exist, then absent a 'benefit of competence' we could also find ourselves ruined, attentionally destitute. And we are not even debtors: we are *donors* of the attention that fuels the financial engines of our digital technologies. They are in *our* debt. And they owe us, at absolute minimum, the benefit of competence.

IV. The Brightest Heaven of Invention

'O for a Muse of fire, that would ascend / The brightest heaven of invention'

– Shakespeare, *Henry V* (Prologue)

The minimum is not enough, of course. While we must preserve a 'minimum viable product' with respect to our attention—a 'minimum viable mind,' if you will—we cannot settle for it. To do so would be to adopt far too low a view of technology. It would be to grant the assumption that our technologies must be adversarial, that they cannot truly be on our side. Such a view would be unbearably pessimistic, even dystopian—and ultimately unsustainable. A better path would be to take seriously the claims that technology can 'make the world a better place,' and to ask that it 'first, do no harm.' In the case of the attention economy, I believe this would mean adopting a stance toward it that is akin to Diogenes' response to Alexander: it would mean asking our technologies and their designers, before anything else, to 'stand out of our light.'

Churchill also said, 'We shape our buildings; thereafter they shape us.' McLuhan said much the same thing about 'our tools,' and I could certainly say the same thing about 'my thesis.' The word 'thesis,' before it ever meant 'proposition' or 'dissertation,' referred to a musical downbeat, and, more generally, 'a setting down, a placing, an arranging; position, situation.' The present work is a thesis in all these senses. In addition to its functional role as the central requirement for the D.Phil. degree, this dissertation serves, far more than I had ever anticipated, as the 'downbeat' to a longer-term rhythm of research in the problem domain I have addressed here, i.e. the philosophy and ethics of attention and persuasion as they relate to digital technology design and deployment. When I began this thesis, I had expected to build on existing foundations. Instead, I have found it necessary to build pieces of the foundation itself: to gather conceptual and linguistic bricks, to place and arrange them, and to join them in hopes that the structure will ultimately hold. It is my hope that this thesis can serve as a guide for the attention of others—for those who, like me, find themselves motivated by a deep concern about the vast infrastructure of technological persuasion we have inherited—but who, also like me, take solace in encountering others on this road who see the same problems, and respond to them with the same vigor of inquiry that I have been fortunate enough to enjoy throughout the course of this D.Phil research. In order to do anything that matters, we must first give attention to the things that matter. It is my firm conviction, now more than ever, that the degree to which we are able and willing to struggle for ownership of our attention is the degree to which we are free.

Bibliography

'Internet.org by Facebook.' (n.d.). Retrieved September 22, 2017, from https://www.facebook.com/Internetdotorg/

'Ten things we know to be true' (n.d). Google. Retrieved September 22, 2017, from: https://www.google.com/about/philosophy.html

'What are acceptable ads?' (n.d). Retrieved September 22, 2017, from https://acceptableads.com/

Abu-Saud, Z. (2013). The Dogma of advertising and consumerism. The Huffington Post.

Ainslie, G. (1992). Picoeconomics: The strategic interaction of successive motivational states within the person. Cambridge University Press.

Ajzen, I. (1991). The theory of planned behavior. Organizational behavior and human decision processes, 50(2), 179-211.

Basheer, A. A. A., & Ibrahim, A. A. (2010). Mobile marketing: Examining the impact of trust, privacy concern and consumers' attitudes on intention to purchase. International Journal of Business and Management, 5(3), 28.

Allsopp M. (1994). G. M. Hopkins, Narrative, and the Heart of Morality: Exposition & Critique. Irish Theological Quarterly, 60: 287.

Appiah, A. (2008). Experiments in ethics. Harvard University Press.

Appiah, K.A. (2008) Experimental philosophy. Proc. Am. Phil. Ass. 82, 7–22

Aristotle (1998). Metaphysics. Penguin.

Aristotle (2008). Physics (R. Waterfield, Trans.). Oxford.

Aristotle. (2004). Rhetoric. Kessinger Publishing.

Aurelius, M., & Staniforth, M. (1964). Meditations. Translated with an Introduction by Maxwell Staniforth. Harmondsworth.

Austin, J. T., & Vancouver, J. B. (1996). Goal constructs in psychology: Structure, process, and content. Psychological bulletin, 120(3), 338.

Awad, N. F., & Krishnan, M. S. (2006). The personalization privacy paradox: an empirical evaluation of information transparency and the willingness to be profiled online for personalization. MIS quarterly, 13-28.t.

Awad, N. F., & Krishnan, M. S. (2006). The personalization privacy paradox: an empirical evaluation of information transparency and the willingness to be profiled online for personalization. MIS quarterly, 13-28.

Bagozzi, R. P., Bergami, M., & Leone, L. (2003). Hierarchical representation of motives in goal setting. Journal of Applied Psychology, 88(5), 915.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being.American psychologist, 54(7), 462.

Barilan, YM Weintraub, M. (2001). 'Persuasion as Respect for Persons: An Alternative View of Autonomy and the Limits of Discourse'. J Med Philos (2001) 26 (1): 13-33.

Barilan, YM Weintraub, M. (2001). 'Persuasion as Respect for Persons: An Alternative View of Autonomy and the Limits of Discourse'. J Med Philos. 26 (1): 13-33.

Becerra, E.P., Korgaonkar, P.K . (2011). "Effects of Trust Beliefs on Consumers' Online Intentions", European Journal of Marketing, Vol. 45 Iss: 6.

Benn, S. I. (1967). Freedom and persuasion. Australasian Journal of Philosophy, 45(3), 259-275.

Berdichevsky, D., & Neuenschwander, E. (1999). Toward an ethics of persuasive technology. Communications of the ACM, 42(5), 51-58.

Berlin, I. (1969). Four Essays on Liberty. Oxford: Oxford University Press.

Bernays, E. (1928). Propaganda. Liveright. New York, 159.

Bobonich, C. (1991). Persuasion, Compulsion and Freedom in Plato's Laws. The Classical Quarterly (New Series), 41 : pp 365-388.

Booth, R. (2014, June 29). Facebook reveals news feed experiment to control emotions. Retrieved from https://www.theguardian.com/technology/2014/jun/29/facebook-users-emotions-news-feeds

Borgmann, A. (1987). Technology and the character of contemporary life: A philosophical inquiry. University of Chicago Press.

Brehm, Jack W. (1966). A Theory of Psychological Reactance.

Brey, P. (2010). Values in technology and disclosive computer ethics. The Cambridge handbook of information and computer ethics, 41-58.

Brin, S., & Page, L. (2012). Reprint of: The anatomy of a large-scale hypertextual web search engine (1998). Computer networks, 56(18), 3825-3833.

Broadbent, S., & Lobet-Maris, C. (2015). Towards a Grey Ecology. In The Onlife Manifesto (pp. 111-124). Springer International Publishing.

Brown M., Muchira R., (2004). Investigating the relationship between Internet privacy concerns and online purchase behavior. Journal of Electronic Commerce Research, Vol. 5, No. 1: 62-70.

Buchanan, A., Brock, D. W., Daniels, N., & Wikler, D. (2001). From chance to choice: Genetics and justice. Cambridge University Press.

Cai, X. & Zhao, X. (2010). Click here kids! Online advertising practices on children's websites. Journal of Children and Media, 4(2), 135-154. DOI: 10.1080/17482791003629610.

Calvert, SL. (2008). "Children As Consumers: Advertising and Marketing." The Future of Children / Center for the Future of Children, the David and Lucile Packard Foundation. 18.1: 205-34. Print.

Cameron, K. A. (2009). A practitioner's guide to persuasion: An overview of 15 selected persuasion theories, models and frameworks. Patient Education and Counseling, 74(3), 309-317.

Campbell, D. E., & Wright, R. T. (2008). Shut-up I don't care: Understanding the role of relevance and interactivity on customer attitudes toward repetitive online advertising. Journal of Electronic Commerce Research, 9(1), 62.

Campbell, J. (2004). Pathways to bliss: Mythology and personal transformation (Vol. 16). New World Library.

Carpenter, C. J. (2013). A meta-analysis of the effectiveness of the "but you are free" compliance-gaining technique. Communication Studies, 64(1), 6-17.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. Journal of personality and social psychology, 39(5), 752.

Choi, S. M., & Rifon, N. J. (2002). Antecedents and consequences of Web advertising credibility: A study of consumer response to banner ads." Journal of Interactive Advertising, . 3(1), Available online at: http://jiad.org

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. Annual review of psychology, 62, 73-101.

Cialdini, R. B. (2006). Influence: The psychology of persuasion. HarperBusiness.

Cooper, J., & Cooper, G. (2006). Subliminal motivation: A story revisited. Journal of Applied Social Psychology, 32(11), 2213-2227.

Costa C., Damásio JM. (2010). "How media literate are we? The voices of 9 years old children about brands, ads and their online community practices." Observatorio Journal, vol.4.4.: 093-115.

Cova, F., & Naar, H. (2012). Testing Sripada's deep self model. Philosophical Psychology, 25(5), 647-659.

Crawford M. (2015). The World Beyond Your Head: How to Flourish in an Age of Distraction. Penguin.

Crisp, R. (1987). Persuasive advertising, autonomy, and the creation of desire. Journal of Business Ethics, 6(5), 413-418.

Csikszentmihalyi, M. (2008). Flow: The Psychology of Optimal Experience. Harper Perennial.

Curtis, V. "A natural history of hygiene." The Canadian Journal of Infectious Diseases & Medical Microbiology 18.1 (2007): 11.

Cushman, F., & Young, L. (2011). Patterns of moral judgment derive from nonmoral psychological representations. Cognitive science, 35(6), 1052-1075.

Cushman, F., Young, L., Greene, J.D. (2010) Our multi-system moral psychology: Towards a consensus view, in The Oxford Handbook of Moral Psychology, J. Doris, G. Harman, S. Nichols, J. Prinz, W. Sinnott-Armstrong, S. Stich, Eds. Oxford University Press.

Dahlberg, L.(2005). "The corporate colonization of online attention and the marginalization of critical communication?" Journal of Communication Inquiry. vol. 29 no. 2: 160-180.

Daniels, N. (1996). Justice and justification: Reflective equilibrium in theory and practice. Cambridge University Press.

Darwall, S. (2006). The Value of Autonomy and Autonomy of the Will*. Ethics, 116(2), 263-284.

Davis, F. D. (1985). A technology acceptance model for empirically testing new end-user information systems: Theory and results (Doctoral dissertation, Massachusetts Institute of Technology, Sloan School of Management).

Davis, J. (2009, April). Design methods for ethical persuasive computing. In Proceedings of the 4th International Conference on Persuasive Technology (p. 6). ACM.

Davis, J. (2009, April). Early experiences with participatory design of ambient persuasive technology. In CHI'09: Workshop on defining the role of HCI in the challenges of sustainability (pp. 119-128).

de Ruyter, B. & Pelgrim, E. (2007). Ambient assisted-living research in carelab. interactions, 14(4), 30-33.

Deacon, T. W. (2011). Incomplete nature: How mind emerged from matter. WW Norton & Company.

Dolan, P. (2014). Happiness by Design: Change What You Do, Not How You Think. Penguin.

Drumwright, M. (1993). Ethical Issues in Advertising and Sales Promotion. In Ethics in Marketing, N. Craig Smith and John A. Quelch, eds., Homewood, IL: Irwin, 607–625.

Drumwright, M. E., & Murphy, P. E. (2009). The current state of advertising ethics: Industry and academic perspectives. Journal of Advertising, 38(1), 83-108.

Dworkin, G. (1988). The theory and practice of autonomy. Cambridge University Press.

Einstein, M. (2016). Black Ops Advertising: Native Ads, Content Marketing and the Covert World of the Digital Sell. OR Books.

Elliot, A. J., Sheldon, K. M., & Church, M. A. (1997). Avoidance personal goals and subjective well-being. Personality and Social Psychology Bulletin, 23(9), 915-927.

Ellul, J. (1965). Propaganda: The formation of men's attitudes. New York: Knopf.

Evans, David S. The Online Advertising Industry: Economics, Evolution, and Privacy. Journal of Economic Perspectives. 23.3 (2009): 37. Print.

Eyal, N. (2013). Hooked: A Guide to Building Habit-Forming Products. CreateSpace.

Falk, E. B., Berkman, E. T., Mann, T., Harrison, B., & Lieberman, M. D. (2010). Predicting persuasion-induced behavior change from the brain. The Journal of Neuroscience, 30(25), 8421-8424.

Falk, E. B., Rameson, L., Berkman, E. T., Liao, B., Kang, Y., Inagaki, T. K., & Lieberman, M. D. (2010). The neural correlates of persuasion: a common network across cultures and media. Journal of cognitive neuroscience, 22(11), 2447-2459.

Fallman, D. (2007). Persuade into what? why human-computer interaction needs a philosophy of technology. Persuasive Technology, 295-306.

Feltz, A., & Cokely, E. T. (2009). Do judgments about freedom and responsibility depend on who you are? Personality differences in intuitions about compatibilism and incompatibilism. Consciousness and Cognition, 18(1), 342-350.

Ferster, C. B. and Skinner. BF (1957). Schedules of reinforcement. Retrieved from: http://psycnet.apa.org/record/2004-21805-000

Finch, J. (1987). The Vignette Technique in Survey Research. Sociology, 21: 105.

Fisher, L. (2014) 'Digital Advertising Trends—Programmatic, Big Data, Native, Viewability. (2014, May 22). eMarketer. Available 22 September 2017, from: https://www.slideshare.net/eMarketerInc/e-marketer-webinardigitaladvertisingtrendsprogrammaticbigdatanativeviewability

Floridi, L. (1999). Information ethics: On the philosophical foundation of computer ethics. Ethics and information technology, 1(1), 33-52.

Floridi, L. (2012). Big data and their epistemological challenge. Philosophy & Technology, 25(4), 435-437.

Floridi, L. (2013). Distributed morality in an information society. Science and engineering ethics, 19(3), 727-743.

Floridi, L. (2014). The Fourth Revolution: How the Infosphere Is Reshaping Human Reality. Oxford: Oxford University Press.

Floridi, L., & Sanders, J. W. (2001). Artificial evil and the foundation of computer ethics. Ethics and Information Technology, 3(1), 55-66.

Floridi, L., & Sanders, J. W. (2004). On the morality of artificial agents. Minds and Machines, 14(3), 349-379.

Floyd, C., Mehl, W. M., Resin, F. M., Schmidt, G., & Wolf, G. (1989). Out of Scandinavia: Alternative approaches to software design and system development. Human–computer interaction, 4(4), 253-350.

Fogg, B. J. (2002). Persuasive technology: using computers to change what we think and do. Ubiquity, 2002(December), 5.

Fogg, B. J., Cuellar, G., & Danielson, D. (2003). Motivating, influencing, and persuading users. The human-computer interaction handbook, 358-370.

Fogg, B.J. (2003). Persuasive Technology: Using Computers to Change What We Think and Do. Morgan Kaufmann.

Fogg, B. J., Soohoo, C., Danielson, D. R., Marable, L., Stanford, J., & Tauber, E. R. (2003, June). How do users evaluate the credibility of Web sites?: a study with over 2,500 participants. In Proceedings of the 2003 conference on Designing for user experiences (pp. 1-15). ACM..

Frankfurt, H. G. (1988). The importance of what we care about: Philosophical essays. Cambridge University Press.

Fricker, M. (2007). Epistemic injustice: Power and the ethics of knowing. Oxford University Press.

Friedman, B. (1996). Value-sensitive design. Interactions, 3(6), 16-23.

Friedman, B., Kahn Jr, P. H., & Borning, A. (2006). Value sensitive design and information systems. Human-computer interaction in management information systems: Foundations, 4.

Ganong, L. H., Coleman, M. (2006). Multiple Segment Factorial Vignette Designs. Journal of Marriage and Family, 68: 455–468.

Garrett, J. J. (2010). The elements of user experience: user-centered design for the Web and beyond. New Riders Pub.

Gasché, R. (2008). On Seeing Away: Attention and Abstraction in Kant. CR: The New Centennial Review, 8(3), 1-28.

Goel, V., (2014, June 29). New York Times. Retrieved from http://www.nytimes.com/2014/06/30/technology/facebook-tinkers-with-users-emotions-in-news-feed-experiment-stirring-outcry.html?_r=0

Goldfarb, A. (2014). What is different about online advertising?. Review of Industrial Organization, 44(2), 115-129.

Goss, J. (1993). The "magic of the mall": an analysis of form, function, and meaning in the contemporary retail built environment. Annals of the Association of American Geographers, 83(1), 18-47.

Grant, I. C. (2005). Young peoples' relationships with online marketing practices: An intrusion too far?. Journal of Marketing Management, 21(5-6), 607-623.

Greene, J. (2014). Moral tribes: Emotion, reason, and the gap between us and them. Penguin.

Guglielmo, S., Monroe, A., Malle, B., (2009). At the Heart of Morality Lies Folk Psychology. Inquiry: An Interdisciplinary Journal of Philosophy. Volume 52, Issue 5.

Guillermina J. (2006). Factorial Survey Methods for Studying Beliefs and Judgments. Sociological Methods and Research. Vol. 34 no. 3 334-423

Ha, L. (2008). Online advertising research in advertising journals: A review. Journal of Current Issues & Research in Advertising, 30(1), 31-48.

Haidt, J. (2003). The moral emotions. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), Handbook of affective sciences. Oxford: Oxford University Press.(pp. 852-870).

Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog?. Journal of Personality and Social Psychology; Journal of Personality and Social Psychology, 65(4), 613.

Harjumaa, M., & Oinas-Kukkonen, H. (2007). Persuasion theories and IT design. Persuasive Technology, 311-314.

Haynes, A. W. (2006). Online privacy policies: contracting away control over personal information. Penn St. L. Rev., 111, 587.

Hébert P., Meslin E.M., Dunn, E.V., Byrne, N. (1990). Evaluating ethical sensitivity in medical students: using vignettes as an instrument. Journal of medical ethics, 16,141-145.

Hebert, P., Dunn, E.V. (1992). Measuring the ethical sensitivity of medical students: a study at the University of Toronto. J Med Ethics, 18:142-147.

Hegel, G. W. F. (1996). The Philosophy of Right (1820), translated by SW Dyde, originally published in English in 1896.

Hemp, P. (2009). Death by information overload. Harvard business review, 87(9), 83-89.

Herman, E. S., & Chomsky, N. (1988). Manufacturing consent: A propaganda model. Manufacturing Consent.

Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). Communication and persuasion; psychological studies of opinion change.

Hyman, M. R., Tansey, R., & Clark, J. W. (1994). Research on advertising ethics: Past, present, and future. Journal of Advertising, 23(3), 5-15.

Information Commissioner's Office. (2012.). New EU Cookie Law. In Information Commissioner's Office. Retrieved 28 October 2012, Retrived from http://www.ico.gov.uk/for_organisations/privacy_and_electronic_communications/the_guide/cookies.aspx.

International Advertising Association. (2014). The Case for Advertising. Retrieved from http://www.iaaglobal.org/YourRightToChoose.aspx on 22 September 2017.

Introna, L. D. (2005). Disclosive ethics and information technology: disclosing facial recognition systems. Ethics and Information Technology, 7(2), 75-86.

Ipsos MediaCT and Innerscope Research Inc. (2011) "TrueView Skippable PreRolls: Evaluating the Effect of Consumer Choice on Pre-roll Effectiveness. Retrieved from http://www.slideshare.net/antonovn/youtube-trueview-skippable-prerolls-study

Jelsma, J. (2006). Designing 'moralized' products. User Behavior and Technology Development, 221-231.

Jelsma, J., & Knot, M. (2002). Designing environmentally efficient services; a 'script' approach. The Journal of Sustainable Product Design, 2(3), 119-130.

Jenkins, N., Bloor, M., et al. (2010). "Putting it in Context: The Use of Vignettes in Qualitative Interviewing." Qualitative Research 10(2):175-198B

Jerome, S. (2010). Schmidt: Google gets 'right up to the creepy line.' The Hill. Retrieved from http://thehill.com/policy/technology/122121-schmidt-google-gets-right-up-to-the-creepy-line.

Johnson, L. (2017) 'U.S. Digital Advertisers Will Make $83 Billion This Year, Says EMarketer. (2017, March 14). Adweek. Retrieved September 22, 2017, from http://www.adweek.com/digital/u-s-digital-advertising-will-make-83-billion-this-year-says-emarketer/

Joines, J. L., Scherer, C. W., & Scheufele, D. A. (2003). Exploring motivations for consumer Web use and their implications for e-commerce. Journal of consumer marketing, 20(2), 90-108.

Kagan, S. (1998). Normative ethics. Oxford: Westview Press.

Kahneman, D. (2011). Thinking, fast and slow. Farrar, Straus and Giroux.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction.Psychological review, 80(4), 237.

Kane, R. (1996). The significance of free will. Oxford University Press, USA.

Kaptein, M. C., Markopoulos, P., de Ruyter, B., & Aarts, E. (2010). Persuasion in ambient intelligence. Journal of Ambient Intelligence and Humanized Computing, 1(1), 43-56.

Kaptein, M., & Eckles, D. (2010). Selecting effective means to any end: Futures and ethics of persuasion profiling. Persuasive technology, 82-93.

Kierkegaard, S. (1980). The concept of anxiety. Trans. Reidar Thomte, Princeton: Princeton University Press.

Knobe, J., & Burra, A. (2006). Experimental philosophy and folk concepts: Methodological considerations. Journal of Cognition and Culture, 6(1-2), 331-342.

Knobe, J., (2004): 'Intention, Intentional Action and Moral Considerations', Analysis 64, 181–187.

Knobe, J., Nichols, S., (eds) (2008). Experimental Philosophy, Oxford University Press.

Knutson, K., et al (2010). Behavioral norms for condensed moral vignettes. Soc Cogn Affect Neurosci. 5(4): 378-384.

Kosta, E., Pitkänen, O., Niemelä, M., & Kaasinen, E. (2010). Mobile-Centric Ambient Intelligence in Health-and Homecare—Anticipating Ethical and Legal Challenges. Science and engineering ethics, 16(2), 303-323.

Kotler, P. (1973). Atmospherics as a marketing tool. Journal of retailing, 49(4), 48-64.

Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. Proceedings of the National Academy of Sciences, 111(24), 8788-8790.

Kuhl, B. A., & Chun, M. (2014). Memory and attention. In The Oxford handbook of attention.

Landa, R. (2005). Designing brand experience: creating powerful integrated brand solutions. Cengage Learning.

Langheinrich, M., Nakamura, A., Abe, N., Kamba, T., & Koseki, Y. (1999). Unintrusive customization techniques for Web advertising. Computer Networks, 31(11), 1259-1272.

Lanham, R. A. (2006). The economics of attention: Style and substance in the age of information. University of Chicago Press.

Latour, B. (1992). Where are the missing masses? The sociology of a few mundane artifacts. Shaping technology/building society: Studies in sociotechnical change, 225-258.

Légal, J. B., Chappé, J., Coiffard, V., & Villard-Forest, A. (2012). Don't you know that you want to trust me? Subliminal goal priming and persuasion. Journal of Experimental Social Psychology, 48(1), 358-360.

Lerbinger, O. (1972). Designs for persuasive communication. Prentice-Hall.

Lessig, L. (2006). Code: And Other Laws of Cyberspace, Version 2.0. Basic Books.

Levy, N. (2007). Neuroethics: Challenges for the 21st Century. Cambridge University Press.

Livengood, J., Sytsma, J., Feltz, A., Scheines, R., & Machery, E. (2010). Philosophical temperament. Philosophical Psychology, 23(3), 313-330.

Llamas et al. "Smartphone OS Market Share, Q4 2014." (n.d.) Retrieved from http://www.idc.com/prodserv/smartphone-os-market-share.jsp

Lobo, P., Romão, T., Dias, A., & Danado, J. (2009). A framework to develop persuasive smart environments. Ambient Intelligence, 225-234.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. American psychologist, 57, 705.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. American psychologist,57(9), 705.

Lockton, D., Harrison, D., & Stanton, N. (2008). Design with intent: Persuasive technology in a wider context. Persuasive Technology, 274-278.

Loo, R., (2002). Tackling ethical dilemmas in project management using vignettes. International Journal of Project Management, 20: 489–495.

Machery, E. (2008). The folk concept of intentional action: Philosophical and experimental issues. Mind & Language, 23(2), 165-189.

MacManus, T. F. (1928). The Nadir of Nothingness. Atlantic Monthly, 141, 594.

Maier, M. A., Barchfeld, P., Elliot, A. J., & Pekrun, R. (2009). Context specificity of implicit preferences: the case of human preference for red. Emotion, 9(5), 734.

Mandelli, A. (2005). "Banners, E-Mail, Advertainment and Sponsored Search: Proposing a Value Perspective for Online Advertising." International Journal of Internet Marketing and Advertising. 2: 92-108. Print.

Mark, G., Gudith, D., & Klocke, U. (2008, April). The cost of interrupted work: more speed and stress. In Proceedings of the SIGCHI conference on Human Factors in Computing Systems (pp. 107-110). ACM.

Marshall, J. (2014, February 20). 'WTF is programmatic advertising?' Digiday. Retrieved from https://digiday.com/media/what-is-programmatic-advertising/

Marvin, C. (1990). When old technologies were new: Thinking about electric communication in the late nineteenth century. Oxford University Press.

Max-Neef, M. A. (2005). Foundations of transdisciplinarity. Ecological economics, 53(1), 5-16.

McDonald A.M., Cranor L.F. (2010). "Beliefs and Behaviors: Internet Users' Understanding of Behavioral Advertising." Telecommunications Policy Research Conference, 2010.

McDonald, P., Mohebbi, M. and Slatkin, B. (2012). Comparing Google Consumer Surveys to Existing Probability and Non-Probability Based Internet Surveys. In Google Consumer Surveys. Retrieved 28 October 2012, from http://www.google.com/insights/consumersurveys.

McGonigal, J. (2011). Reality is broken: Why games make us better and how they can change the world. Penguin.

McGuire, W. J. (1964). Inducing resistance to persuasion. Adv. Experimental Social Psychology, 1, 191.

McLuhan, M. (1964). Understanding media: the extensions of man. McGraw-Hill.

McLuhan, M., & Gordon, W. T. (2006). The classical trivium: The place of Thomas Nashe in the learning of his time. Gingko Pr Inc.

McLuhan, M., & McLuhan, E. (1998). Laws of Media: The New Science.

McNair, C.(2017, April 12). Worldwide Ad Spending: The eMarketer Forecast for 2017. eMarketer. Retrieved from https://www.emarketer.com/Report/Worldwide-Ad-Spending-eMarketer-Forecast-2017/2002019

Meadows, D. (1997). Places to Intervene in a System. Whole Earth, 91(1), 78-84.

Mehrabian, A. (1971). Silent messages (Vol. 8). Belmont, CA: Wadsworth.

Meyer, Robinson. (2014). "Everything We Know About Facebook's Secret Mood Manipulation Experiment." The Atlantic, 2014. Retrieved from http://www.theatlantic.com/technology/archive/2014/06/everything-we-know-about-facebooks-secret-mood-manipulation-experiment/373648/

Mill, J. S. (1989). JS Mill: 'On Liberty' and Other Writings. Cambridge University Press. Originally published 1859.

Miller, E. and Buschman, T. (2014). Neural Mechanisms for the Selective Control of Attention. In The Oxford Handbook of Attention, 777-805. Oxford: Oxford University Press.

Mori, M. (1970). The uncanny valley. Energy, 7(4), 33-35.

Mundfrom, DJ, Shaw, DG, & Tian, LK (2005) Minimum sample size recommendations for conducting factor analysis. International Journal of Testing 5:2:p 159-168

Nadelhoffer, T., & Nahmias, E. (2007). The past and future of experimental philosophy. Philosophical Explorations, 10(2), 123-149.

Nairn, A., & Dew, A. (2007). Pop-ups, pop-unders, banners and buttons: The ethics of online advertising to primary school children. Journal of Direct, Data and Digital Marketing Practice, 9(1), 30-46.

Nakajima, T., Yamabe, T., & Sakamoto, M. (2011). Proactive ambient social media for supporting human decision making. Ubiquitous Intelligence and Computing, 25-39.

Newman, K. (2002). Poisons, Potions, and Profits: Radio Rebels and the Origins of the Consumer Movement. In Radio Reader: Essays in the Cultural History of Radio, 157-181. Psychology Press.

Nichols, S. (2011). Experimental Philosophy and the Problem of Free Will. Science. Vol. 331 no. 6023 pp. 1401-1403

Nielsen, (2015). 'Online advertising spend is expected to surpass offline advertising for the first time ever' (2017, February 2015). Retrieved September 22, 2017, from http://www.nielsen.com/hk/en/press-room/2017/online-advertising-spend-is-expected-to-surpass-offline-advertising-for-the-first-time-ever.html

Nobre, A. and Kastner, S. (2014). The Oxford Handbook of Attention. Oxford University Press.

Noë, A. (2004). Action in perception. MIT Press.

Noë, A. (2011). Ideology and the Third Realm (Or, a Short Essay on Knowing How to Philosophize). Knowing How: Essays on Knowledge, Mind, and Action, 196-2

Offer, A. (2006). The challenge of affluence: Self-control and well-being in the United States and Britain since 1950. Oxford University Press.

Oinas-Kukkonen, H. (2010). Behavior change support systems: A research model and agenda. Persuasive Technology, 4-14.

Oinas-Kukkonen, H., & Harjumaa, M. (2008). A systematic framework for designing and evaluating persuasive systems. Persuasive Technology, 164-176.

Öncel, L. (2014). Career adapt-abilities scale: Convergent validity of subscale scores. Journal of Vocational Behavior, 85(1), 13-17.

Packard, V. (2007). The Hidden Persuaders. 1957. Intro. Mark Crispin Miller. Brooklyn, NY: Ig.

Paine, T. (2001). Common Sense: 1776. Infomotions, Incorporated.

Parr, Ben. (2015). "Captivology: The Science of Capturing People's Attention."

Pennock, J. R. (1972). Coercion: an overview. Coercion, 1-15.

Pereboom, D. (2001). Living without free will. Cambridge University Press.

Pettit, D., & Knobe, J. (2009). The pervasive impact of moral judgment. Mind & Language, 24(5), 586-604.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. Advances in experimental social psychology, 19(1), 123-205.

Phillips J, Knobe J. (2009). Moral Judgments and Intuitions about Freedom. Psychological Inquiry: An International Journal for the Advancement of Psychological Theory. 20, 1.

Pires, H. (2013). Uses of space, freedoms and constraints: São Paulo, Clean City: a case study. Lusophone Journal of Cultural Studies, 1(2), 249-263.

Posner, M. I., Rafal, R. D., Choate, L. S., & Vaughan, J. (1985). Inhibition of return: Neural basis and function. Cognitive neuropsychology, 2(3), 211-228.

Postman, N. (1985) Amusing Ourselves to Death: Public Discourse in the Age of Show Business. Penguin.

Postman, N. (1992). Technopoly: The Surrender of Culture to Technology. Vintage.

Powers, P. (2007, June). Persuasion and coercion: a critical review of philosophical and empirical approaches. In HEC Forum (Vol. 19, No. 2, pp. 125-143). Springer Netherlands.

Prehn K, et al. (2008). Individual differences in moral judgment competence influence neural correlates of socio-normative judgments. Soc Cogn Affect Neurosci. 3:33–46.

PricewaterhouseCoopers (2016). Internet advertising revenue report. PricewaterhouseCoopers. Retrieved September 22, 2017, from https://www.iab.com/wp-content/uploads/2016/04/IAB_Internet_Advertising_Revenue_Report_FY_2016.pdf

Prinz, J. (2008). "Empirical Philosophy and Experimental Philosophy." In Knobe and Nichols, eds. Experimental Philosophy. Oxford UP, 189-208.

Rader, E. and Gray, R. (2015). "Understanding User Beliefs About Algorithmic Curation in the Facebook News Feed." CHI 2015.

Rawls, J. (1971). A Theory of Justice. Cambridge: Harvard University Press.

Raz, J. (1986). The morality of freedom. Clarendon Press.

Redström, J. (2006). Persuasive design: Fringes and foundations. Persuasive Technology, 112-122.

Reeves, B., Nash, C. (1998). The media equation. New York, NY: CSLI Publications.

Richards, J. I., & Curran, C. M. (2002). Oracles on "advertising": Searching for a definition. Journal of Advertising, 31(2), 63-77.

Rideout, V.J., Foehr, U.G., & Roberts, D.F. (2010). Generation m2: Media in the lives of 8- to 18-year-olds. Menlo Park: Kaiser Family Foundation.

Rivlin, G. (2007). Slot machines for the young and active. New York Times. Retrieved September 22, 2017 from http://www.nytimes.com/2007/12/10/business/10slots.html

Rodgers, S., & Thorson, E. (Eds.). (2012). Advertising theory. Routledge.

Rogers, K. (2014). The Attention Complex: Media, Archeology, Method. Palgrave Macmillan.

Rogers, R. W. (1975). A Protection Motivation Theory of Fear Appeals and Attitude Change. The Journal of Psychology, 91(1), 93-114.

Rorty, J. (1934). Our Master's Voice: Advertising.

Rossi, P. H., & Anderson, A. B. (1982). The factorial survey approach: An introduction. Measuring social judgments: The factorial survey approach, 15-67.

Rossi, P., Sampson, W., Bose, C., Jasso, G., Passel, J., (1974). "Measuring Household Social Standing." Social Science Research 3:169-90.

Ruyter, B. D., & Pelgrim, E. (2007). Ambient assisted-living research in carelab. interactions, 14(4), 30-33.

Ryan, Richard M., and Edward L. Deci. (2001). "On happiness and human potentials: A review of research on hedonic and eudaimonic well-being." Annual review of psychology 52.1: 141-166.

Samuels, R., Stich, S., & Bishop, M. (2002). Ending the rationality wars: How to make disputes about human rationality disappear.

Savulescu, J. (2007). Autonomy, the good life, and controversial choices. The blackwell guide to medical ethics, 17-37.

Savulescu, J. (2011). Human liberation: Removing biological and psychological barriers to freedom. Monash bioethics review, 29(1), 04-1.

Schechter, S., & Bravo-Lillo, C. (2014). Using ethical-response surveys to identify sources of disapproval and concern with facebook's emotional contagion experiment and other controversial studies. Technical Report MSR-TR-2014-97.

Schroeder, R. (2013). An age of limits: social theory for the 21st century. Palgrave Macmillan.

Schroeder, R., & Cowls, J. (2014). Big data, ethics, and the social implications of knowledge production. GeoJournal (accessed 24.01.15).

Schüll, N.D. (2012). Addiction by Design: Machine Gambling in Las Vegas. Princeton University Press.

Shawyer RJ, bin Gani AS, Punufimana AN, Seuseu NK. (1996). The role of clinical vignettes in rapid ethnographic research: a folk taxonomy of diarrhoea in Thailand. Soc Sci Med. 42(1):111-23.

Sheehan, Kim B, and Mariea G. Hoy. (1999). "Flaming, Complaining, Abstaining: How Online Users Respond to Privacy Concerns." Journal of Advertising. 28.3: 37-51. Print.

Shelton, C. M., & McAdams, D. P. (1990). In search of an everyday morality: The development of a measure. Adolescence.

Sheppard, B. H., Hartwick, J., & Warshaw, P. R. (1988). The theory of reasoned action: A meta-analysis of past research with recommendations for modifications and future research. Journal of Consumer research, 325-343.

Sherif, C. W., Sherif, M., & Nebergall, R. E. (1965). Attitude and attitude change: The social judgment-involvement approach. Philadelphia: Saunders.

Simon, H. (1971). Designing organizations for an information-rich world. In M. Greenberger (Ed.), Computers, communications, and the public interest. Baltimore, MD: The Johns Hopkins Press.

Singer, P. (1993). Practical ethics. Cambridge University Press. Originally published 1973.

Skipper, R., Hyman, M., (1993). On measuring ethical judgments. Journal of Business Ethics. Volume 12, No 7.

Sloterdijk, Peter. (2013). "You must change your life: On Anthropotechnics, trans. Wieland Hoban:" 12.

Smids, J. (2012). The voluntariness of persuasive technology. Persuasive Technology. Design for Health and Safety, 123-132.

Smith, A; Rogers, V. (2000). Ethics-related responses to specific situation vignettes: Evidence of Gender-Based Differences and Occupational Socialization. Journal of Business Ethics. 28, 1.

Sontag, S. (2007). At the same time: Essays and speeches. Farrar, Straus and Giroux.

Spahn, A. (2011). And Lead Us (Not) into Persuasion…? Persuasive Technology and the Ethics of Communication. Science and Engineering Ethics, 1-18.

Spence, E. and Van Heereken, B. (2005). Advertising Ethics. Upper Saddle River, NJ: Pearson Prentice Hall.

Sripada, C. S. (2011). What Makes a Manipulated Agent Unfree?. Philosophy and Phenomenological Research.

Stone, L. (2008). Just Breathe: Building the Case for E-mail Apnea. The Huffington Post.. Retrieved from http://www.huffingtonpost.com/linda-stone/just-breathe-building-the_b_85651.html (2008).

Strahan, E. J., Spencer, S. J., & Zanna, M. P. (2002). Subliminal priming and persuasion: Striking while the iron is hot. Journal of Experimental Social Psychology, 38(6), 556-568.

Strauss, D. (1991). Persuasion, Autonomy, and Freedom of Expression. Columbia Law Review. Vol. 91, No. 2, pp. 334-371.

Sunstein, C. R. (2015). Nudging and choice architecture: ethical considerations. Harvard John M. Olin Discussion Paper Series Discussion Paper No. 809, Jan. 2015, Yale. J. Reg.

Taylor, B., (2006). Factorial Surveys: Using Vignettes to Study Professional Judgement. British Journal of Social Work, 36, 1187–1207.

Thaler, R. and Sunstein, C. (2008). Nudge: Improving decisions about health, wealth, and happiness. Yale University Press.

Toubiana, V., Narayanan, A., Boneh, D., Nissenbaum, H., & Barocas, S. (2010). Adnostic: Privacy preserving targeted advertising.

Townsend, G. F. (2004). Aesop's Fables. Kessinger Publishing.

Triandis, H. C. (1977). Interpersonal behavior. Monterey, CA: Brooks/Cole Publishing Company.

Turow, J. (2012). The daily you: How the new advertising industry is defining your identity and your worth. Yale University Press.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. Cognitive psychology, 5(2), 207-232.

van Tuinen, H. K. (2011). The Ignored Manipulation of the Market: Commercial Advertising and Consumerism Require New Economic Theories and Policies. Review of Political Economy, 23(2), 213-231.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. MIS quarterly, 425-478.

Verbeek, P. (2006). Materializing Morality. Science Technology Human Values. vol. 31 no. 3 361-380.

Verbeek, P. P. (2009). Ambient intelligence and persuasive technology: the blurring boundaries between human and technology. Nanoethics, 3(3), 231-242.

Verbeek, P. P. (2011). Moralizing technology: understanding and designing the morality of things. University of Chicago Press.

Verplaetse, J. (2008). Measuring the moral sense: morality tests in continental Europe between 1910 and 1930. Paedagogica Historica: International Journal of the History of Education. Volume 44, Issue 3.

Virilio, P. (2009). Grey Ecology (D. Burk, Trans.).

Von Foerster, H. (2003). Ethics and second-order cybernetics. In Understanding understanding (pp. 287-304). Springer New York.

Wainwright, P., et al. (2010). The use of vignettes within a Delphi exercise: a useful approach in empirical ethics? J Med Ethics, 36:656-660.

Wall, M. (2014). Big Data: Are you ready for blast-off?. BBC News, March.

Wallander, L. (2009). 25 years of factorial surveys in sociology: A review. Social Science Research 3, 505–520.

Walsh, M. F. (2010). New insights into what drives Internet advertising avoidance behaviour: The role of locus of control. International Journal of Internet Marketing and Advertising, 6(2), 127-141.

Weinberg J, Nichols S, Stich S. (2001). Normativity and epistemic intuitions. Philos. Topics 29:429–60

Wilkinson, T. M. (2013). Nudging and manipulation. Political Studies, 61(2), 341-355.

Williams, B. (2001). From freedom to liberty: The construction of a political value. Philosophy & public affairs, 30(1), 3-26.

Williams, J. (2012). Telling Stories in the Uncanny Valley: Exploring the Role of Narrative in Moral Vignettes. Unpublished paper.

Williams, J. (2016). Why It's OK to Block Ads. In Philosophers Take on the World. Oxford: Oxford University Press.

Winner, L. (1980). Do artifacts have politics?. Daedalus, 109(1), 121-136.

Witte, K. (1992). Putting the fear back into fear appeals: The extended parallel process model. Communications Monographs, 59(4), 329-349.

Wittgenstein, L. (1998). Tractatus Logico-Philosophicus (C.K. Ogden, Trans.). Dover.

Wolin, Lori, Korgaonkar, Pradeep. "Web advertising: gender differences in beliefs, attitudes and behavior", Internet Research, Vol. 13 Iss: 5, (2003): pp.375 - 385.

Woolfolk, R., Doris, J., Darley, J.M. (2007). Identification, Situational Constraint, and Social Cognition : Studies in the Attribution of Moral Responsibility. In Joshua Knobe (ed.), Experimental Philosophy. Oxford University Press.

Wu, T. (2016). The attention merchants: The epic scramble to get inside our heads. Vintage.

Yang, H., & Oliver, M. B. (2004). Exploring the effects of online advertising on readers' perceptions of online news. Journalism & Mass Communication Quarterly, 81(4), 733-749.

Zalta, E. N., Nodelman, U., Allen, C., & Perry, J. (2012). Personal autonomy. In Stanford encyclopedia of philosophy. Retrieved August 1, 2012.