Jeremy Rose & Oskar MacGregor

The Architecture of Algorithm-driven Persuasion

Abstract

Persuasion is a process that aims to utilize (true or false) information to change people's attitudes in relation to something, usually as a precursor to behavioural change. Its use is prevalent in democratic societies, which do not, in principle, permit censorship of information or the use of force to enact power. The transition of information to the internet, particularly with the rise of social media, together with the capacity to capture, store and process big data, and advances in machine learning, have transformed the way modern persuasion is conducted. This has led to new opportunities for persuaders, but also to well-documented instances of abuse: fake news, Cambridge Analytica, foreign interference in elections, etc. We investigate large-scale technology-based persuasion, with the help of three case studies derived from secondary sources, in order to identify and describe the underlying technology architecture and propose issues for future research, including a number of ethical concerns.

Introduction

Practically everyone with a computing device and an internet connection benefits from the last decade's rapid expansion in what is often referred to as the 'volume, velocity and variety' of digital data sets (Mcafee, A., and Brynjolfsson 2012). These days, most of us use digital platforms such as Google, Facebook, YouTube, Twitter, or Instagram daily. And each time we use such services (for internet search, social media, e-commerce, user help, recommendation systems, operating system optimisation, etc.), they collect, curate, and analyse the data that we — many times unknowingly — actively contribute to them. The primary techniques for making use of this data are algorithmic in nature (machine learning, deep learning, etc.) — techniques loosely collected under the heading of artificial intelligence — and their existence facilitates numerous opportunities for the persuasion of individual users.

Persuasion is defined here as ‘ï»¿human communication designed to influence the autonomous judgments and actions of others’ (Simon 2001). In other words, it is a process aimed at changing a person or group’s attitude in relation to some event, idea, object, or other person(s), by using written and/or spoken words or visual tools to convey combinations of information, feelings, and reasoning, usually as a precursor to behavioural change. More specifically for the current context, the existence of large-scale digital platforms has transformed the possibilities of using one-way mass communication for persuasion, by adding a variety of networking, multi-dimensional communication, and user-generated content effects to the more traditional domains of mass media (news, television, billboards, etc.). In other words, the way in which the digital platforms are structured — their architecture — enables novel means of persuasion (cf. Trottier 2012). In order to more fully understand how the digital platforms facilitate such persuasion, it is therefore critical to investigate the architectural structures of the platforms themselves.

As Lessig (1999) has argued, architecture inevitably governs human actions, including persuasion, regardless of whether the architecture is embedded in bricks and mortar, or the code structures of digital platforms. In this context, therefore, architecture describes the manner in which the components of a computer or computer system are organised and integrated, i.e. how digital structures create user environments and contexts (Hinton 2015). An architecture description is then ‘a conceptual design or planâ€¦ often expressed in terms of drawings (blueprints) showing the general composition and layout of all of the parts’ (Dumas 2006); i.e., a formal description and representation of a system, organised in a way that supports reasoning about the structures and behaviours of the system.

Bossetta (2018), for instance, discusses digital platform architecture in terms of network structure, functionality, algorithmic filtering, and datafication models. Information is a primary component of persuasion, so it is also reasonable to understand the mechanics of modern persuasion as an information architecture — ‘the intentional composition of nodes and links as organised structures that facilitate understanding’ (Arango 2011). Thus, a persuasion architecture is an arrangement of technologies and information that structures and filters human capacity for interpretation, decision-making, and action, with the intention of presenting the most relevant information to the most relevant individual(s) in order to achieve some desired (e.g. behavioural) outcome. Since persuasion involves cognition, the primary human actors are also included in our understanding of architecture — a socio-technical system. Persuasion architectures affect behaviour by ‘modifying the content, context, and conditions of choice-making’ (Pascal 2018).

More specifically, a persuasion architecture presupposes two groups of humans: persuaders, and their target audience (the 'persuadees'). The intention of persuaders is to reinforce or alter the attitudes, values and beliefs of the target audience in such a way that their behaviour is changed, to help enact the desired goals and outcomes of the persuaders. Most users who provide content to the digital sphere act as persuaders from time to time, but in this context we mean groups who consciously and knowledgeably use the full resources of the architecture to persuade — a task well beyond the competence and resources of most ordinary users. Thus, persuaders use the architecture to send messages to their audience, where the message concept is governed not by the form it takes (which may be textual, visual, or symbolic, encapsulated as an email, tweet, text, video, image, etc. — any kind of information that can be digitalised), but by its intention: to persuade. The function of persuasion content is to reinforce the message, though the primary persuasion message may not be explicit throughout this content. It should be noted that the truth value of persuasion content is inconsequential; what is important is whether the content is on-message, and can be integrated into the existing worldview of the target audience. Contemporary digital platforms, with their inherently algorithmic nature, facilitate this.

It should be noted, that we use a fairly wide applied (rather than computer science theoretical) definition of algorithm to mean a sequences of steps, enacted in programming code, that solve a computational problem or perform a computational task. Algorithm-driven persuasion therefore describes persuasion that is primarily enabled by programming code executed by computerised platforms. Pascal (2018) explains that algorithmic persuasion techniques differ from earlier broadcast media in at least four ways. First, web tracking technologies (beacons, cookies, pixel tags, etc.) enable the large-scale collection of behavioural data about individual users — which sites they visit, how they navigate through them, what they buy, their click behaviour ('likes' and 'dislikes'), and so on. Second, algorithm-driven architectures can target individuals or groups based on demographic, locational, psychometric, and behavioural characteristics. Third, algorithmic persuasion architectures can be designed to be flexible and adaptive. Finally, the logic of how and why a certain persuasion message is delivered to particular users cannot be precisely specified or recovered, because of the scale and complexity of the data sources and algorithmic learning and distribution techniques involved.

This form of persuasion — across digital platforms and driven by software algorithms — is both fairly recent, and (in technologically advanced societies) ubiquitous. It may therefore be sensible to speak, with Howard (2005) of 'managed' (digital) citizens, whose choices and behaviours are conditioned and influenced to some extent by the architectural spaces they inhabit when online. Thus, 'communication on social media is mediated by a platform's digital architectureâ€"the technical protocols that enable, constrain, and shape user behaviour in a virtual space' (Bossetta 2018). In this vein, Arango (2011) notes that there are 'few means of societal control as powerful as the ability to define the boundaries of discussion and the language used for the exchange'.

There are, in other words, several reasons for investigating the architecture of digital algorithmic persuasion. First, persuasion that is organised algorithmically may have considerably larger reach and pervasiveness than traditional mass persuasion. Second, the providers of algorithmically-driven technical architectures, such as Google and Facebook, have commercial reasons for keeping the details of how they operate secret, since they generate much of their revenue from advertising (Zuboff 2019). Third, persuasion architectures are only partially visible to the scrutiny of their users, instead mostly characterised by a 'ï»¿lack of visibility, information asymmetry and hidden influence' (Tufekci 2015). Even informed users are unable to precisely understand why and how the information that populates their information spaces is presented to them. Finally, the resources necessary to operate or exploit such an architecture for mass persuasion are considerable; Richterich (2018), from a critical theory perspective, suggests that the operators of persuasion architectures are primarily the already-rich and powerful, which raises a number of ethical considerations.

In research terms, the topic is partially investigated in a number of rather diverse literatures: political scientists investigate the effects of social media on election campaigns, marketers discuss how to influence customers through the new media computer scientists devise and test algorithms that analyse, sort, and distribute various types of data, and information architects discuss how to structure web platforms to inform users. There are very few attempts to integrate understandings across these disciplines, with the notable exception of Zuboff (2019), who developed the term 'surveillance capitalism' to describe the way that big tech companies use the data they collect to support targeted marketing.

The objective of this article is to further the exploration of the common features and structural properties of digital persuasion through the mechanism of architectural thinking. The research question here is therefore:

> *What digital architectures underpin contemporary large-scale algorithm-driven persuasion?*

The literature underpinning the topic is rather heterogeneous, so we investigate and organise many theoretical sources to create an initial map of architectural components. We then refine this component map through the analysis of multiple case studies. Difficulties with commercial secrecy and ethical sensitivity lead us to choose case studies compiled from secondary sources: the Facebook voting experiment, Cambridge Analytica's involvement with the 2016 American presidential election, and McDonald's' and Starbucks' use of Pokémon Go.

The next section introduces the research approach for the study, followed by the literature analysis, featuring six important architectural components. Three case studies with their architectural analyses follow, and the generalised map of persuasion architecture resulting from the literature study and case analyses is then explained. The final section draws implications for future research, poses ethical dilemmas, and specifies conclusions.

Research Approach

This study is exploratory in nature, addressing 'what' questions concerning the dynamics present within a particular contemporary context (Eisenhardt 1989), in this case algorithm-driven persuasion, with the objective of developing initial understandings. The research approach (Figure 1) combines cross-disciplinary literature study with a multiple case study approach.

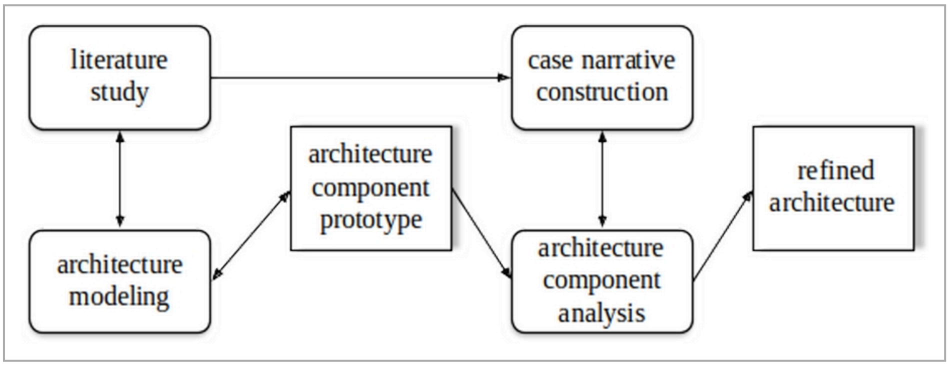

literature study
case narrative construction
architecture component prototype
refined architecture
architecture modeling
architecture component analysis

Figure1. Research approach

A scoping literature review (Munn et al 2018) was used both to identify architectural components and to identify sources for the case studies. Scoping reviews are considered appropriate for identifying key characteristics, factors or components related to a concept. The recommendations of Webster & Watson (2002) (concerning selection of sources from multiple databases and analysis by concept) were followed to develop the material. All combinations of the search words 'persuasion', 'algorithm', 'architecture', and 'social media' (excluding the combination 'algorithm and architecture' which yielded a massive technical literature with little relevance) were used with the primary search engines Web of Science and Google Scholar. This process generated in excess of 1,246,000 hits at Google Scholar, and 2,181 hits in topic search at Web of Science core databases. Experimentation showed that the relevancy was a problem at Google Scholar because of the genericity of the search words, and that the first ten pages contained the majority of the relevant articles.

A further selection process involved downloading abstracts of the candidate articles and manually inspecting them. Backward chaining by reviewing the citations in the articles identified in the search was operated in several iterations. Articles addressing aspects of algorithmically facilitated persuasion in social media, including architectural considerations, were included in the study. This process yielded a Mendeley database of 66 books and articles for which the full texts were obtained. Content analysis (Krippendorff 2004) was used to identify important concepts, in this case representing primary components of persuasion architectures, and an initial prototype architecture developed from this. The scoping literature study, organised by architectural component, is given in the following section.

A multiple case study approach was used to provide a strong theoretical base for theory building (Benbasat, Goldstein & Mead 1987; Yin 2009). Cross-case comparisons clarify whether findings are idiosyncratic to a single case, or replicated through several cases. Multiple case studies facilitate establishing patterns of relationships between constructs within and across cases with their underlying logical arguments (Eisenhardt & Graebner 2007), by recursive cycling among the case data. The method consists of selecting multiple cases, triangulating data, and analysing the data both within, and across, cases (Yin 2009). The data is investigated in various ways, including the construction of a unifying narrative around the subject of investigation from multiple written accounts, as well as comparing this iterative analysis against a developing architecture model.

Whereas single case studies may richly describe phenomena, multiple case studies provide comparisons across varied empirical settings (Eisenhardt & Graebner 2007), where developing propositions can be more deeply grounded. Because of the difficulty of obtaining reliable first hand empirical data in areas which combine commercial secrecy with the stretching of ethical and legal boundaries, the case study narratives used here are drawn together from published secondary sources. This limits the choice of cases to those that are sufficiently well-known to be the subject of multiple research articles, and sufficiently far back in time for these articles to be developed and published. However, such known and previously researched cases suit the societal level of analysis taken here.

For this research, we selected three cases: an insider account of persuasion on voting behaviour from Facebook, a well-known case from the political arena in which a company (Cambridge Analytica) claimed to be able to use the Facebook platform to influence the 2016 American presidential election, and a less well-known commercial example where major retailers McDonald's and Starbucks used the Pokémon Go platform to persuade people to visit their outlets. The construction of a relevant case narrative from the various document sources is intended to 'capture the essential features of the primary narrative in an ordered context which will allow its relevance to be easily perceived and understood â€¦ essential if one is to establish a clear theoretical conjecture or set of conjectures that one hopes to test against the collected data' (Remenyi & Williams 1996).

The first (Facebook voting) case is intended to illuminate the internal activity of a major digital platform — the digital stage for persuasion — since these are controlled by big technology and not normally accessible for scrutiny by researchers in any open way. The other cases are chosen to illustrate the twin arenas of political and commercial persuasion, in order to examine similarities and differences in the structure of algorithmically delivered persuasion. In the Facebook voting experiment case, the primary source is the peer-reviewed article Bond et al (2012), written by a combination of Facebook insiders and academics; additional sources are given in Appendix A. For the other cases, conventional literature search procedures were used to identify sources from reputable publishers that cover various aspects of the cases. The sources identified for the Cambridge Analytica case are given in appendix B. Three journalist and insider sources were included for completeness, but used carefully with consideration of the authors' personal and political agendas. There is a great deal of research on Pokémon Go, but we limited our search to articles explicitly naming McDonald's or Starbucks, since these are the best documented commercial sponsors of the game. The sources for this case are given in appendix C.

Algorithm-driven Persuasion in the Literature

This section presents six important components of an algorithm-driven persuasion architecture, as derived from our literature study. The six components are:

- the ubiquitous collection of data concerning the target audience,

- algorithmic analysis of that data, in order to understand the behavioural characteristics of audience segments,
- personalised content generation to respond to those characteristics,
- algorithmic logistics for the delivery of messages to audience segments,
- message amplification reinforcing the persuasion message, and
- behavioural effect measurements designed to understand the results of the persuasion exercise.

Ubiquitous Data Collection

The previous two decades have been marked by the arrival of big data and the technologies which store and analyse it (Chen, Chiang, & Storey 2012; Kosinski, Wang, Lakkaraju, & Leskovec 2016). The digitisation of many of our communication forms and their distribution via the internet, together with the rise of social media and the ubiquity of computing devices, has generated unprecedented volumes of data (structured and unstructured, mobile and sensor) from a variety of old and new sources. The development is accompanied by a wide commercial understanding of the value of data, and new industries that collect and analyse it (Manokha 2018). Much of this data is identified or personal — data collected that can be associated with individuals, either by the various markers that identify us (names, id numbers, email addresses, etc.) or by simple extrapolation and data combination (locational data about where we live and work, digital face recognition, car registration numbers, etc.).

A significant amount of the data just described is supplied voluntarily, often in return for various internet services, or as part of our online activity, while a further level of data is collected by the digital platforms we use (search histories, click patterns, site usage, likes and dislikes, retweets, etc). This is supplemented by extensive monitoring and tracking, facilitated by cookies, web bugs, and other trackers (Zuboff 2019). Extended data collection about individuals for commercial purposes extends beyond the internet into public spaces (e.g. Google Street View), further into the home, and even to the monitoring of individuals' biometrics. For instance, Internet of Things home devices relay information to their vendors. A smart TV's camera may monitor a family's reaction to television advertisements, while an intelligent speaker relays their questions to Google's servers for language parsing. And the ubiquitous smartphone collects numerous forms of data about its owner (Zuboff 2019), stationing up to 30 sensors in close bodily proximity, and sending data to service operators and app providers, often without the knowledge or informed consent of the phone's owner. The extent, normalisation and routine character of this commodification of individuals' extended data leads Zuboff (2019) to characterise it as 'surveillance'.

Algorithmic Analysis of Data

Ubiquitous extended data collection is of little value without the ability to analyse it. Since the volume of data is large, analysis depends upon algorithmic and data fusion techniques. Conventional market segmentation (by age, gender, location, etc.) of digital platform users does not usually require sophisticated algorithmic techniques. Instead, advanced analytics, such as machine learning, are used to inductively infer various further characteristics of individuals or groups (Tene & Polonetsky 2013). More specifically, classification algorithms assign users to predefined categories on the basis of their data characteristics. Common methods include naÃ¯ve Bayes, support vector machines, k-nearest neighbour, decision trees, expectation maximisation, and so on. In addition, clustering analysis uncovers unanticipated trends, correlations or patterns in the data, using techniques such as decision tree construction, rule induction, clustering, logic programming, and so on.

These algorithmic techniques enable a variety of behavioural characteristic inferences, including trait modelling, and psychological and lifestyle profiling. Kosinski, Stillwell, and Graepel (2013) describe the modelling of the latent traits of 58,000 volunteers from their Facebook likes, including sexual orientation, ethnicity, religious and political views, personality traits, intelligence, happiness, use of addictive substances, parental separation, age, and gender. Similarly, Matz et al (2017) demonstrate that individuals can be psychologically profiled from their Facebook likes or other digital footprints. In fact, algorithmic profiling is shown, in many respects, to outperform conventional questionnaire profiling (Youyou, Kosinski, & Stillwell 2015), demonstrating higher external validity when predicting life outcomes, including substance use, political attitudes, and physical health.

Psychological profiling can also be used together with interest, lifestyle, and demographic profiles (Baldwin-Philippi 2017). In the health domain, these inferences are characterised as 'category leaps' by Horvitz and Mulligan (2015), defined by the authors as 'the use of machine learning to make leaps across informational and social contexts to infer health conditions and risks from non-medical data'. In the educational domain, O'Neil (2016) describes how for-profit university recruiters identify social media users suffering life crises from their data traces in order to sell them 'redemptive' education.

Algorithmic categorisation is also used to identify 'lookalike' audiences of users with similar characteristics to a known individual (for instance a supporter of a particular political party; Baldwin-Philippi 2017). A typical use of such algorithmic analysis techniques might be to identify potential buyers of a new product, or persuadable voters in swing states in an election (Howard & Bradshaw 2017). Persuaders therefore can work with the data their audience voluntarily supplies, extended data collected during their online activity, and algorithmically-drawn inferences derived from analysis, while additional data fusion techniques allow the linking of many different data sources.

â€‹Personalised Content

When they fall into a more obvious advertising or political campaign genre, persuasion messages are relatively easily recognised by their audience. However, direct advocacy on social media is often accompanied by persuasion content that is, to some extent, disguised. For instance, sponsored search and keyword auctions at Google can result both in the placement of obvious sidebar and in-feed advertisements, as well as manipulation of the order of the links returned for searches on popular consumer items and services — the sponsor effectively pays to have their link further up in the ranking (Jansen et al 2009). The latter is a disguised form of promotion that may not be understood by the service user.

On social media, most advertising is targeted, although users may not understand that they have been targeted, or why. While exposure to news can have a positive effect on democratic discourse (Diehl, Weeks, & Gil de ZÃºÃ±iga 2016), a large volume of social media postings are junk news — extremist, sensationalist, or fake news, or photos with misleading captions and doctored videos masquerading as news — often deliberately using the genre conventions of legitimate news services to disguise their origin (Bradshaw & Howard 2018). When 'junk news is backed by automation â€¦ through dissemination algorithms â€¦ political actors have a powerful set of tools for computational propaganda' (Howard & Bradshaw 2017; Neudert, Kollanyi, & Howard 2017).

Algorithmic analysis at digital platforms facilitates accurate message targeting, and persuaders make increasing use of micro-targeting; 'creating finely honed messages targeted at narrow categories of [voters] â€¦ based on data analysis garnered from individuals' demographic characteristics and consumer and lifestyle habits' (Borgesius et al 2018). For instance, Auger (2013) reports that 'advocacy and fundraising messages employed different rhetoric with advocacy messages designed to inspire logical decision-making and fundraising to appeal to readers' emotional decision-making'. Behavioural micro-targeting uses automated psychological profiling as the basis for generating tailored persuasion content (Wilson 2017). Tailoring involves adapting the message and/or its framing and imagery to suit the profile (e.g. OCEAN personality score) of the receiver. Modern algorithms enable digital platforms to 'micro-analyse and micro-serve content to increasingly specialised segments of the population down to the individual' (Wilson 2017), allowing individually tailored and personalised persuasion content. Personalised persuasion content may benefit from a resonance effect (Wilson 2017), in that the subject is less aware of the manipulative intent of the message. A junk news posting, for example, is taken as factual because it aligns with the reader's pre-existing sympathies.

Algorithmic Delivery Logistics

Persuasion content must be delivered to its audience, often in rather precise ways. Digital platforms use algorithmic gatekeeping (Bossetta 2018; Bucher 2012; Cotter, Cho, & Rader 2017; Tufekci 2015) to feed content to users, e.g. through Facebook's Newsfeed (EdgeRank) algorithm or the Google web-search algorithm based on PageRank. Algorithms sort and rank content for display to individual users based on information supplied by them: a search word or phrase in the case of Google's search algorithm, or a previous history of likes, clicks, and preferences in the case of Facebook's news feed. More specifically, algorithmic gatekeeping is:

> *the process by which such non-transparent algorithmic computational tools dynamically filter, highlight, suppress, or otherwise play an editorial role — fully or partially — in determining information flows through online platforms â€¦ Gatekeepers acting with computational agency are able to tweak the content viewers receive on an individualised basis, without being visible â€¦ This functionality is often largely unknown to the users of given services â€¦ Algorithmic filtering refers to how developers prioritise the selection, sequence, and visibility of posts. (Tufekci 2015).*

For instance, in a controversial experiment, Facebook researchers Kramer, Guillory, and Hancock (2014) used their algorithm to manipulate the amount of positive and negative content on a large number of users' feeds. Similarly, algorithms provide the delivery logistics for persuasion messages by, for instance, embedding targeted advertisements in a Facebook newsfeed. In this sort of context, Bucher (2012) defines reach as 'how far a post cascades across a broadcast feed or set of networks', where algorithmic filtering can either promote or limit a post's reach. Many digital platforms offer pay-to-promote services (such as 'boosting' on Facebook), which allow persuaders to adjust or override algorithmic filtering and further the reach of their content.

Baldwin-Philippi (2017) reports that Facebook has added to its advertising platform to make micro-targeting easier, allowing persuaders to target more precise interests based on keywords or categories, geographical data, and algorithmically created lifestyle profiles.

Message Amplification

Individual persuasion messages may have little effect, however well targeted and tailored, if they are not reinforced by amplification. Zhang et al (2018) define amplification as 'the contribution of â€¦ publics to the attention paid to a particular object (person, message, idea) by elevating other actors' (citizens, journalists, media platforms) perceptions of the object's worthiness or significance'. Message amplification often involves the frequent repetition of the message in variant forms, combined with surrounding a primary message with supporting persuasion content.

For instance, marketers use viral marketing and retargeting [1] based on cookies to serve adverts across and between e-commerce and social media sites, which remind potential customers of sites and products they have already visited. Social media platforms provide many vehicles for repeating and increasing the reach of content (likes, shares, retweets) and for amplifying the message across platforms. The nature of the content may affect its reach, with social media distribution algorithms reportedly favouring sensationalist content (Diehl et al 2016). For instance, the Facebook algorithm is thought to promote virtual echo chambers/filter bubbles — i.e. spaces where limited sets of ideas are constantly reinforced — by feeding users types of content to which they have previously responded favourably (Wilson 2017). This creates selective exposure to news (Messing & Westwood 2014) and echo chamber amplification effects (Hameleers & Schmuck 2017).

Moreover, the persuasion effects of social engagement (such as Facebook interest groups), and endorsement networking effects (likes, shares, retweets, etc.) are thought to considerably enhance the penetration of content (Diehl et al 2016; Messing & Westwood 2014). Social media opinion leaders and prosumers (Weeks, ArdÃ¨vol-Abreu, & De ZÃºÃ±iga 2017) concentrate and exploit these amplification effects. The power of these effects is demonstrated by Facebook researchers in their emotional contagion experiment (Kramer et al 2014) — they showed that the mood of users could be affected by the (algorithmically manipulated) content they were shown, and then spread from affected users to their networks of friends.

Selective exposure in social media is also promoted by partisan sharing (An, Quercia, & Crowcroft 2014), leading to distorted understandings of the reliability of sources, skewed exposure to content, and issue polarisation. Powerful actors use both human and automated means to exploit message amplification effects. Commercial actors sponsor social media influencers to promote their products and services, while political actors use cyber troops and troll farms (Badawy, Addawood, Lerman, & Ferrara 2019; Bradshaw & Howard 2017).

Bradshaw and Howard (2017) report that all countries use these techniques, whether orchestrated by military units or strategic communication firms. Authoritarian regimes largely target their own populations and democracies target foreign audiences, while political parties target potential voters. Techniques include positive discursive interaction, negative abuse and harassment (trolling), and diversion manoeuvres (hashtag poisoning), which divert attention from embarrassing trends — a form of de-amplification deemed necessary because of the power of viral amplification.

Message amplification is also automated through bots. For instance, Neudert, Kollanyi and Howard (2017) found that 7.4% of the traffic in their sample of political messages in the 2017 German election were generated by bots. Likewise, Badawy et al (2019), analysing a large sample of political tweets from the 2016 American presidential election, determined that 5% of the liberal tweeters and 11% of the conservative ones were bots. In fact, another study (Howard & Bradshaw 2017) found the following, also in relation to the 2016 election:

> *The number of links to professionally produced content was less than the number of links to polarising and conspiratorial junk news (…) A worryingly large proportion of all the successfully catalogued content provides links to polarising content from Russian, WikiLeaks, and junk news sources (…) This content uses divisive and inflammatory rhetoric, and presents faulty reasoning or misleading information to manipulate the reader's understanding of public issues and feed conspiracy theories (…) Fully 32% of (…) political content was polarising, conspiracy driven, and of an untrustworthy provenance.*

Behavioural Effect Measurement

A final necessary component in algorithmic persuasion is the ability to measure the behavioural effect of digital persuasion. The major digital platforms offer automated tools: Google Trends, Facebook Insight and Ad analytics, Twitter analytics, Instagram analytics, YouTube analytics, and so on. YouTube analytics, for instance provide metrics on estimated traffic sources, watch time, views, earnings, ad performance, audience retention and subscribers — covering reach, engagement, and audience. These are complemented by an exhaustive variety of third party add-ons and professional research tools for marketers and researchers. Besides allowing persuaders easy access to understanding the impact of their message, they enable randomised experiments which better craft and hone messages for persuasive effect (Bossetta 2018). Measurement data collected in this way can be matched with other datasets such as voter registers or purchase histories for greater insight.

Taken together, these six components - ubiquitous data collection, algorithmic analysis of data, personalised content, algorithmic delivery logistics, message amplification, and behavioural effect measurement — constitute the essential elements of an algorithm-driven persuasion architecture.

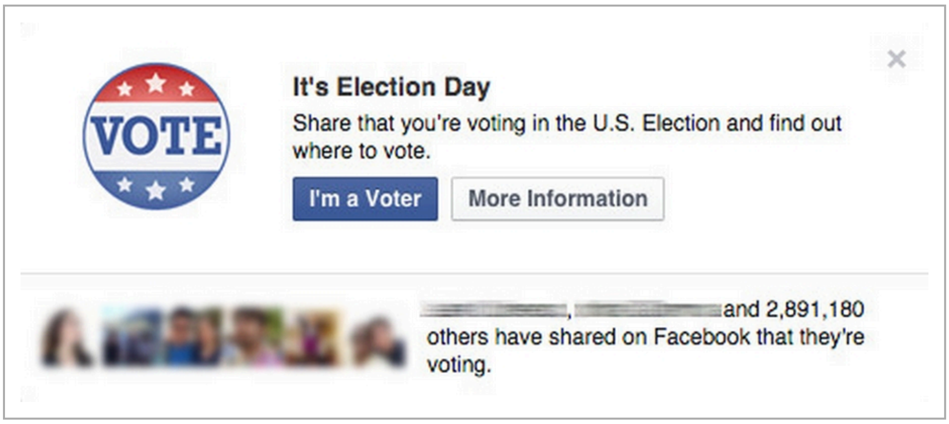
Case Study Analysis

In this section, the architecture components are further explored and refined through the analysis of three case studies.

The Facebook Voting Experiment

A group of Facebook data scientists and University of California researchers published an article in Nature that documented the results of an experiment designed to establish whether Facebook's 'I'm a voter' button was influential in persuading Americans to vote in the 2010 US Congressional elections (Bond et al 2012). For the experiment, 61 million adult Facebook users accessing the social media site on Election Day were randomly assigned to a social message group, an informational message group, or a control group. The social message group (60m) was shown a link to local polling places, a clickable 'I'm a voter' button, a counter for users who had previously clicked, and six small profile pictures of friends who had already clicked (Figure 2). The informational message group (0.6m) were fed the same items without the pictures of friends. A control group (0.6m) received an unaltered NewsFeed. 6.3m users were matched to publicly available voting records to study their actual voting behaviour.

The messages in the treatment groups were designed to encourage voting, with and without the social network effect of identifiable social connections (the pictures of friends), which personalise the messages in such a way that no two users' NewsFeeds are likely to be identical, and add a message endorsement to promote the penetration of the message — an amplification effect referred to as 'social contagion' by the researchers, and 'social pressure' by Haenschen (2016).



It's Election Day

Share that you're voting in the U.S. Election and find out where to vote.

[redacted], [redacted] and 2,891,180 others have shared on Facebook that they're voting.

Figure2. Simulated anonymized news feed header

The researchers showed small but significant effects of the messages: 'online political mobilisation can have a direct effect on political self-expression, information seeking and real-world voting behaviour, and â€¦ messages including cues from an individual's social network are more effective than information-only appeals' (Bond et al 2012). They could also demonstrate small contagion effects — users were more likely to vote if their close friends (i.e. friends they interacted regularly with) also received the message. They estimated that 'the Facebook social message increased turnout directly by about 60,000 voters and indirectly through social contagion by another 280,000 voters, for a total of 340,000 additional votes' (Bond et al 2012). This effect appears small, but finely balanced elections are decided by small margins - about 80,000 votes in this case, according to the Washington Post [2]. In Facebook's experimentation strategy, 'economies of action are discovered, honed and ultimately institutionalised in software programs and their algorithms that function automatically, continuously, ubiquitously and pervasively â€¦ to modify your behaviour' (Zuboff 2019). Zittrain (2014) speculated whether such techniques could be used to engineer an election result — without the public ever being aware that it was being influenced.

Table 1 gives the architecture analysis for the Facebook voting experiment. It provides details on both the general case specifics (the persuader, the intended target audience, the persuasion message — whether explicit or implicit, and the digital platform on which the persuasion attempt took place), as well as a brief description of how the case — as summarized above — pertains to the six architectural components gleaned from our literature review (i.e. the means of data collection, algorithmic data analysis, personalised content generation, algorithmic delivery logistics, message amplification, and behavioural effect measurements).

Table 1. Architecture analysis for Facebook voting experiment

Architecture component	Facebook voting experiment
Persuaders	Facebook executives, carried out by their data scientists and programmers (represented by Kramer and Marlow for the Nature article), assisted by University of California researchers
Target audience	The American electorate (represented by the segment with Facebook accounts)
Persuasion message	Vote today
Digital platform	Facebook social media platform
Target audience data collection	User profiles already available in Facebook's databases, publicly available digital state voter records
Audience data analytics	Sorting of Facebook users to distinguish American users of voting age; random assignment to two treatment groups (social message and informational message) and a control group
Personalised content generation	Different messages for the two treatment groups as described above; friend endorsement effects
Delivery logistics	Facebook newsfeed algorithm that determines the order in which users are shown content based on a ranking score; commercially secret but thought to be based on the Vickreyâ€“Clarkeâ€“Groves auction algorithm
Message amplification	Amplification built in through a voting counter, and for the social message through pictures of voting friends
Behavioural effect measurements	Clickthrough to the two buttons (understood as intention to vote); matching against state voter files (accounting for about 40% of voters) to correlate actual voting behaviour; friends analysis

Cambridge Analytica and the 2016 US Presidential Election

Cambridge Analytica were employed by the Trump presidential campaign to manage its online campaigning — designated as 'Project Alamo'. According to one insider report, the company used the following procedure. They amassed large quantities of data about voters, segmented them into groups, used predictive algorithms to further refine and characterise the groups, identified the interactive media where particular voter groups could be reached, devised microtargeted advertising content for groups and individuals, and then distributed it through a variety of digital platforms, while refining the approach through real-time behavioural metrics (Kaiser 2019). This account corresponds well with many earlier descriptions in the scientific literature. The firm claimed to have collected up to 5,000 data points on over 220 million Americans, which included 'Facebook likes, retweets and other data gleaned from social media â€¦ commercially available personal information: land registries, automotive data, shopping data, bonus cards, club memberships, what magazines you read, what churches you attend â€¦ [supplied by] data brokers such as Acxiom' (Grassegger & Krogerus 2017).

The company's signature audience analysis was based on psychographic techniques incorporating the big five personality traits: openness, conscientiousness, extroversion, agreeableness and neuroticism (OCEAN). These were developed by researchers — in particular Cambridge University psychology researcher Aleksandr Kogan — working to infer various forms of sensitive personal information (such as sexual preferences) from seemingly trivial social media interactions or 'data exhaust' (cf. Kosinski et al 2016). Kogan's team created Facebook 'personality' tests harvesting various data points from the individuals taking the tests, but also, through a Facebook loophole, all their Facebook friends. This led to the collection and processing of personal data from as many as 87 million Facebook users (Manokha 2018).

At least part of this trove of data was transferred to Cambridge Analytica's servers in order to develop similar functionality (Berghel 2018; Cadwalladr 2018; Manokha 2018; Tarran 2018). Cambridge Analytica's methods combine OCEAN profiles with information about personal preferences, consumption patterns, reading and viewing habits, and other data mined from a range of public and private sources, to reportedly sort voters into 32 different personality types (Grassegger & Krogerus 2017). This framework was then used to identify what was presumed to be 20 million persuadable voters in key battleground states. Voter analysis was paired with a large number of tailored messages, some created for the campaign, some leveraging existing content on the social media platforms.

Cambridge Analytica could thus micro-target different clusters of US voters; serving them with ads that were seen to cater to their particular intersection of interests and concerns with their personality types and demographics (Isaak & Hanna 2018; Ward 2018). One insider claimed that there were 'many different types of an ad, all tailored to different groupsâ€¦ hundreds or thousands of versions of the same ad concept' with personalised delivery so that 'most of the population didn't see what their neighbour saw' (Kaiser 2019). For example, dark posts (ads or updates not open to public scrutiny) targeted African-American voters in crucial states, reminding them of Clinton's earlier characterisation of African-American men as 'super predators', aiming to discourage them from voting (Green & Issenberg 2016). Wilson (2017) reports a complex automated advertisement administration — 'based on the ads selected by users, content was added to their feed in posts personalised for them, determined by their behaviour profile â€¦ automatically selecting from the thousands of ad variants available, these rules targeted specific individuals and seem to have created the same echo chambers as described in Twitter'. In other words, a combination of real news and misinformation with unmoderated Internet content was used to target voters with reinforcing content across platforms, without them realising that they were receiving personalised content, and without any warning that these were political campaign messages.

Amplification effects for Trump's twitter campaign were studied by Zhang et al (2018) who concluded that 'the far-right, Trump supporters, and Alt-Right â€¦ 11% of our sample â€¦ accounted for a full 60% of all of Trump's retweets'. Here partisan supporters consciously provided amplification effects through retweeting. The involvement of trolls and bots is also documented (Badawy et al 2019). Real-time monitoring of ad responses, including real-time substitution to find clickbait that worked, enabled the campaign to both maximise its impact and detect trends not visible at the macro scale. Isaak and Hanna (2018) estimate that 'tipping the scale in a few states with as few as 100,000 voters, using individualised, high-impact messages is sufficient to impact election results'. Cambridge Analytica's executives claimed they were able to carry the Electoral College for Trump by manipulating only 40,000 voters in three states (Berghel 2018).

In reality, it is uncertain whether Cambridge Analytica was able to provide any significant competitive edge for the Trump campaign (as suggested by Berghel 2018; Cadwalladr 2018; Isaak and Hanna 2018; Persily 2017), or if this was mostly marketing bluster from the company (the view of Baldwin-Philippi 2017; GonzÃ¡lez 2017). Some commentators (e.g. Baldwin-Philippi 2017) claim that the primary campaign effect was based on Facebook's inherent micro-targeting capacities, rather than Cambridge Analytica's psychographic analysis. However, the effects were ostensibly the same: targeted ads crafted to cater to very specific groups (Borgesius et al 2018). Although the impact of this strategy is still under debate, many commentators have concluded that it constituted a small, but nevertheless significant reason for Trump's victory (Berghel 2018; Cadwalladr 2018).

As for the previous case, Table 2 gives the architecture analysis for the

Cambridge Analytica case, on the basis of the case summary above.

Table 2. Architecture analysis for Trump presidential campaign on Facebook

Architectural component	Trump presidential campaign
Persuaders	The Trump campaign, carried out by some combination of campaign personnel and/or Cambridge Analytica, the latter assisted by Cambridge University researcher Aleksandr Kogan
Target audience	The American electorate (represented in particular by the segment with Facebook accounts)
Persuasion message	Vote for Trump and/or don't vote for Clinton
Digital platform	Facebook social media platform, among others
Target audience data collection	Facebook user profiles, combined with Facebook's own marketing categorisations and/or results from previous â€˜personality' tests delivered through Facebook, combined with other bought-in third party data
Audience data analytics	The means by which Facebook develop their own marketing categorisations is uncertain. Cambridge Analytica based its user information on OCEAN personality tests, on the basis of previous "proof-of-concept" work by Kosinski & Stilwell (2013)
Personalised content generation	Variant political messages for different user groups, designed to activate individual users' particular concerns or personality, including so-called â€˜dark posts', and reinforcing content without an obvious election campaign message
Delivery logistics	Driven by Facebook's algorithms for ad distribution, e.g. in the sidebar or in user feeds
Message amplification	Amplification effected primarily through social media re-posting and re-tweeting, with cross-channel effects
Behavioural effect measurements	Facebook analytics combined with swing state voting patterns, used for refining and optimising targeting

Pokémon Go and McDonald's/Starbucks

The Pokémon Go case provides an example of a commercial use of a persuasion architecture. Pokémon Go is a geocaching/augmented-reality spinoff of the well-known children's game Pokémon, in which players catch and battle virtual Pokémon (displayed on the games' digital maps and through augmented reality using players' smartphone cameras). The game was created by Niantic Labs, a spinoff of Google currently owned by Google's parent company Alphabet, and many of the early developers came from Google Maps. Niantic's founder was John Hanke, product vice-president of Google Maps, and director of StreetView. The primary feature of the game is that players move through physical space, guided by the game's maps and their smartphone GPS systems, in order to compete. Niantic's algorithms govern when and where Pokémon appear and can be caught; typically for short periods of time so that players are encouraged to move through physical terrain.

The game app was launched in 2016 and was downloaded more than 500 million times by the end of the year, creating a global craze (Pokémania) which subsided as quickly as it arrived, though the game is still widely played. A variety of third-party apps complement the game, including several in-game chat and messaging apps. Community sites are located on Facebook and Reddit, and the game evolved a strong community focus with groups of players organising 'raid battles' and 'lure parties' (i.e. different forms of multi-player events). The combination of these community effects with the changing locations of Pokémon led to large groups of players temporarily occupying public (and occasionally private) spaces (Colley et al 2017).

Commercial involvement concentrates on so-called 'Pokéstops' (locations where players could acquire powerful game aids) and 'gyms' (places where battles could occur). These operate as 'spawn points', at which Niantic's algorithms determine which wild Pokémon are available and when they are visible. In addition, Niantic offer these to corporations as sponsored locations, with a cash-per-visit payment model. Gyms serve as the locations for raids — where teams of multiple players battle, requiring coordination within the player community. These locations have the dual advantage of acting as nodal points that many players would visit, and involving activities with somewhat longer durations, so that players were encouraged to stay for at least some minutes. Niantic retains historical location data (together with a variety of other personal data collected from its players' smartphones) on its servers, generating a comprehensive picture of where the players are active and where many of them congregate.

As a result, Bloomberg reported that Niantic had achieved 'the retailers' elusive dream of using location tracking to drive foot traffic' (Zuboff 2019), a conclusion reinforced by Colley et al, (2017), who describe the game as 'a rare catalyst for large-scale destination choice change'. McDonald's became an official game sponsor, and by February 28 2017, Pokémon Go had around 3,500 sponsored locations with 2,000 Pokéstops and 500 Gyms at McDonald's outlets. 12,000 US Starbucks also became Pokéstops or gyms. Both franchises developed complementary products to attract players: McDonald's gave away Pokémon toys with their Happy Meals, and Starbucks offered a Pokémon Go Frappuccino.

Although the primary message of these retailers is to consume at their outlets (as with all marketing efforts), this message is never directly available to the target audience, who may not understand that their gaming behaviour has been deliberately modified for commercial purposes. Instead, the game incentivises (nudges) its players to be physically present at the outlet locations, with the fairly reliable assumption that they will be tired, hungry and thirsty when they arrive. Colley et al, (2017) — in a five-country field study with 375 interview subjects — report that 'almost half of interviewees (46%) had purchased something at a venue they were near because of Pokémon GO-related movement'. Typically, these were foodstuffs (25% mentioned purchasing drinks and 23% food). Relatedly, researchers describe how retail outlets in the USA set lures that, for a small payment, increase the number of Pokémon at a Pokéstop, thus generating more foot traffic (Frith 2017; Kirkpatrick et al 2017). In McDonald's case, Calvo (2016) estimates that the community of 3.4 million players in Japan made 2 to 2.5 million daily visits across the 3,000 sponsored locations, with a sales increase of 22% and an increase in market capitalisation of 9.8%. Pamuru, Khernamnuai and Kannan (2017) investigated the effects of being located close to a PokéStop on Houston restaurants, concluding that these restaurants attracted significantly more customer traffic. The case is recounted in Zuboff (2019 pp. 309-19).

As above, Table 3 provides the architecture analysis for the Pokémon Go case.

Table 3. Architecture analysis of McDonald's/Starbucks Pokémon Go campaign

Architectural component	McDonald's/Starbucks Pokémon Go campaign
Persuaders	McDonald's, Starbucks
Target audience	Pokémon Go players
Persuasion message	Visit a relevant outlet (and possibly purchase something while there)
Digital platform	Pokémon Go
Target audience data collection	Player location data, game-related data, other personal data (combined with open maps data)
Audience data analytics	Location analytics
Personalised content generation	Individual gaming objectives incentivising sponsored locations (no openly available and understandable persuasion messages)
Delivery logistics	Algorithmic creation of desirable game objects and their locations, automated guidance through GPS systems to these locations (McDonald's/Starbucks outlets) on digital maps
Message amplification	Additional paid in-game incentives, internal player community communication
Behavioural effect measurements	Visit metrics (pay per visit)

â€‹

The Architecture of Large-scale Algorithm-driven Persuasion

The final architecture (fig. 3) is derived from the initial literature-generated components, refined through the case study analyses.

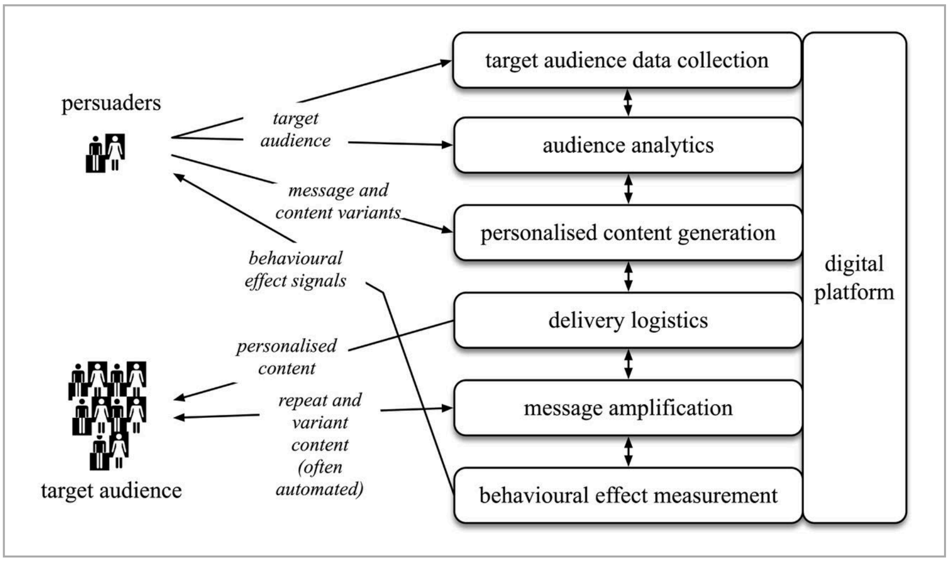

persuaders
target audience
message and content variants
behavioural effect signals
personalised content
repeat and variant content (often automated)
target audience
target audience data collection
audience analytics
personalised content generation
delivery logistics
message amplification
behavioural effect measurement
digital platform

Figure3. The architecture of large-scale algorithm-driven persuasion

The backbone of a persuasion architecture is an existing digital platform, which provides the computing infrastructure for collecting and storing data, analytical computing procedures, and the structured delivery of digital content to the members of the target audience. Examples are social media platforms — Facebook, Twitter, Instagram, Reddit, etc. — but in principle most large networked computing infrastructures can be, and are used, for persuasion. Google and Facebook, for example, use their infrastructure for targeted marketing, while Pokémon Go drives foot traffic to retail outlets in a bid to impact consumer behaviour (and in exchange for payment from the outlets in question). A large digital platform can address an audience of several billion people, and the algorithmic computational power of the platform enables persuasion on a scale, and in a form, not previously known, as both the Facebook voting experiment and Cambridge Analytica cases make clear. In a mass persuasion effort — such as the Cambridge Analytica case — several digital platforms may be targeted, and the majority of platforms have mechanisms for cross-platform sharing.

To summarize, the first component of the architecture collects data about the target audience members. Users voluntarily supply their personal data to digital platforms in return for services, such as for Facebook. However, data about users is also collected in many other ways, sometimes without the informed consent of users — through clicks, likes, and retweets; through cookies and tracking pixels; through the ubiquitous smartphone; and through the enormous array of devices connected to the Internet of Things. Cambridge Analytica's reliance on the Facebook loophole that made it possible for them to collect data on individual users' friends is a prime example of this. Datasets can also be algorithmically combined, as in Cambridge Analytica's reliance both on its own 'personality' test results and external (e.g. Acxiom) datasets. In this manner, the large datasets of the various digital media platforms can be incorporated into persuasion activities.

The second, audience analytics, component uses algorithmic techniques to determine key characteristics of the target audience. The analysis can be simple (which regions do the members live in? which age groups do they belong to?) or complex: personality profiling, or categorisation or clustering using machine learning algorithms. The purpose of the analysis is to break the target audience into segments according to key characteristics, as in Facebook's own built-in micro-targeting ad categories. The segments are then allotted personalised content that reflects their key characteristics, as part of the third architecture component. Personalised content can be understood as complementary subsets of the persuaders' primary message — designed to mobilise a particular audience segment with a particular attitudinal makeup towards the persuaders' desired attitudes and goals. Personalisation designs messages that are in tune with particular audience segments' known or projected attitudes, beliefs, values, and worldviews, or that respond to their needs, cater to their prejudices, or awake deep-seated anxieties. They are therefore typically believable, independent of their truth component. An illustrative example is the 2016 Trump campaign distributing 'dark posts' to individual users, with tailored and publically inaccessible content.

In the fourth component, delivery logistics send the messages to the correct audience segments, using the algorithms facilitating the platform (Facebook and Google content-ranking algorithms, for instance) or other programming devices, such as algorithmically creating in-game incentives to direct individual players to specific (sponsored) gaming locations in Pokémon Go. In these senses, personalised messaging on a large scale is scarcely possible without the resources and algorithmic structures of the digital platform. However, audiences are familiar with persuasion messages, and individual messages do not typically carry much weight without amplification. The fifth, message amplification, component of a persuasion architecture is therefore designed to reinforce the persuasion message, through repetition and exposure to complementary variant content, such as the design including pictures of Facebook friends who had already pledged to vote in the Facebook voting experiment case. Digital platforms also facilitate amplification through reposting, retweeting, trending, viral effects, and other content sharing and focusing mechanisms, and content ranking algorithms encourage echo chamber effects that deliver content to like-minded users. In addition to the sharing of messages, some users can be expected to devise variant messages — new or modified content that echoes the original primary message and is returned to the digital platform for distribution, such as internal player communication for organizing community events in Pokémon Go. If much of the amplification layer is performed by users, though facilitated by the digital architecture, an increasing proportion of amplification might also be organised algorithmically, through fake user accounts and bots, as in the case of the re-posting and retweeting of significant amounts of the 2016 Trump campaign material. In a persuasion architecture, the role of bots is normally the automation of amplification — the sharing of message content — but they may also be used for other tasks, such as variant message generation. These algorithmic devices are usually enabled by the application programming interfaces of the digital platforms.

The sixth and final architecture component is behavioural effect measurement. The platform provides mechanisms for monitoring the effects of persuasion — counting likes, shares and reposts; qualitative feedback from the audience that can be scraped and algorithmically analysed, whether through pre-programmed tools such as Google Trends, or through access to the application programming interface (API); etc. Behavioural effect measurements outside the confines of the platform is more complex and may involve the integration of additional datasets, such as in all three of the cases analysed above.

Conclusions and Implications

This article has examined the contemporary practice of large-scale algorithm-driven persuasion executed on digital platforms. Various sources within the disciplines of political science, computing science, information systems, and marketing describe various elements of this phenomenon. We used this, together with three literature-generated case studies, to describe the common architecture underpinning such persuasion efforts. More specifically, we used the architecture concept to describe a pervasive feature of digital communication, generalised across many digital platforms, rather than in its more common computer science role of providing a blueprint for development.

The architecture contains both technical and informational elements. Six architecture components facilitate algorithm-driven persuasion: target audience data collection, algorithmic audience data analytics, personalised content generation, algorithmic delivery logistics, message amplification, and behavioural effect measurements. Persuaders deliver personalised message variants and accompanying content to their audience, and use the amplification and measurement features of the platform to improve the effectiveness of their efforts. The use of multiple platforms is common, and the architecture components do not necessarily need to lie on the same platform; Cambridge Analytica, for instance, ported Facebook user data to their own servers for audience analytics, and encouraged amplification on other platforms, such as Twitter.

A common architecture serves both political campaigning and commercial purposes, as the case studies illustrate, and targeted advertising is therefore, and not surprisingly, ubiquitous across many digital platforms. This architecture describes features of personalised large-scale persuasion not previously available, before the emergence of large-scale digital platforms and a sophisticated use of algorithms. Many theorisations of various elements of algorithm-driven persuasion architectures are available in the literature, but integrating theoretical overviews are in short supply.

A technical architecture may be considered morally neutral, but its use is always subject to ethical evaluation; a variety of ethical considerations arise. Encouraging citizens to vote, as in the Facebook experiment, may be embraced by societies. However, persuasion architectures can also be used for a variety of ethically dubious purposes, which may come to resemble abuse. O'Neil (2016), for instance, reveals how for-profit universities use social media behavioural markers to target people with life crises (such as divorce or bankruptcy), and serve them personalised advertisements for 'life-changing' education, in the knowledge that these groups attract government loans. The resulting students have low completion rates, and the universities allegedly spend more on the consultancies recruiting them than on their education.

Similar technological architectures are used for social control in China (Creemers 2018). Ubiquitous data collection without adequate control leads to privacy dilemmas concerning extended personal data. It is generally accepted (for instance by the European Court of Human Rights), that the right to information privacy is not absolute but must be held in balance with other factors, including economic prosperity. This is expressed as the 'trade-off between data privacy and data utility' by Monreale et al (2014). However, Cambridge Analytica's exploitation of Facebook users' personal data went far beyond what users had knowingly given informed consent for. They also used algorithmic techniques to infer individual character traits from the data, which went well beyond the (voluntarily supplied) personal data. The resulting threat is called 'predictive privacy harm' by Crawford and Schultz (2014), and its commercialisation referred to as the 'behavioural futures market' by Zuboff (2019).

Persuasion may be overt, delivered in the form of advertisements or exhortations to vote for a particular candidate, but it may also be covert, as in the Pokémon Go case, raising more complex ethical issues. In addition, persuasion content may be factual, or it may be faked. Fake content may achieve the status of truth for many people through amplification — by being reinforced many times in many variations. Thus 'automation adds incredible efficiency to misinformation messaging' (Berghel 2018). Social media operators employ various strategies (including some algorithmic ones) to police their sites, but the global nature of digital platforms makes it difficult to oversee, supervise, and legislate against abuse. A further concern is power and knowledge asymmetry (Carbonell 2016); since individuals rarely have the expertise or resources necessary to utilise the large digital platforms for persuasion purposes, algorithmic persuasion tends to be the preserve of large companies and established political groupings with extensive financial support.

This article sets out a generalised descriptive architecture for a very common feature of digital society, with literature-generated case studies, which are appropriate for the societal level of analysis, although both of these features also act as limitations to the research: namely, an architecture that is not generated by inspection of a real-world technical system, and case studies lacking first-hand empirical evidence. In future research, the generalised architecture should be detailed and localised for particular situations, and supported, where possible, by direct empirical observation. Detailed similarities and differences between political persuasion and commercial persuasion (marketing) should also be examined and categorised. Various ethical considerations related to algorithm-driven persuasion are currently being explored, but researchers need to take a more active role in explaining these dilemmas and providing potential solutions for policy and law makers. There is also a need for greater public awareness of how online environments can be used to condition opinion and behaviour (beyond obvious advertisements and campaign messages), which researchers need to contribute to developing.

References

An, J., Quercia, D., & Crowcroft, J. (2014) Partisan sharing: Facebook evidence and societal consequences. COSN 2014 - Proceedings of the 2014 ACM Conference on Online Social Networks, 13–23. https://doi.org/10.1145/2660460.2660469

Arango, J. (2011) Architectures. Journal of Information Architecture, 3(1). https://doi.org/10.1016/0042-207x(68)92558-x

Auger, G. A. (2013) Fostering democracy through social media: Evaluating diametrically opposed nonprofit advocacy organizations' use of Facebook, Twitter, and YouTube. Public Relations Review, 39(4), 369–376. https://doi.org/10.1016/j.pubrev.2013.07.013

Badawy, A., Addawood, A., Lerman, K., & Ferrara, E. (2019) Characterizing the 2016 Russian IRA influence campaign. Social Network Analysis and Mining, 9(1), 1–11. https://doi.org/10.1007/s13278-019-0578-6

Baldwin-Philippi, J. (2017) The Myths of Data-Driven Campaigning. Political Communication, 34(4), 627–633. https://doi.org/10.1080/10584609.2017.1372999

Benbasat, I., Goldstein, D. K., & Mead, M. (1987) The case research strategy in studies of information systems. MIS Quarterly: Management Information Systems, 11, 369–386. https://doi.org/10.4135/9781849209687.n5

Berghel, H. (2018) Malice Domestic: The Cambridge Analytica Dystopia. Computer, 51, 84–89. https://doi.org/10.1109/MC.2018.2381135

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H. (2012) A 61-million-person experiment in social influence and political mobilization. Nature, 489(7415), 295–298. https://doi.org/10.1038/nature11421

Borgesius, F. J. Z., Möller, J., Kruikemeier, S., Fathaigh, R., Irion, K., Dobber, T., â€¦ de Vreese, C. (2018) Online political microtargeting: Promises and threats for democracy. Utrecht Law Review, 14(1), 82–96. https://doi.org/10.18352/ulr.420

Bossetta, M. (2018) The Digital Architectures of Social Media: Comparing Political Campaigning on Facebook, Twitter, Instagram, and Snapchat in the 2016 U.S. Election. Journalism and Mass Communication Quarterly, 95(2), 471–496. https://doi.org/10.1177/1077699018763307

Bradshaw, S., & Howard, P. N. (2017) Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation. University of Oxford (Vol. Computatio).

Bradshaw, S., & Howard, P. N. (2018) Challenging truth and trust: A global inventory of organized social media manipulation. Oxford Internet Institute, University of Oxford. http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/07/ct2018.pdf

Bucher, T. (2012) Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. New Media and Society, 14(7), 1164–1180. https://doi.org/10.1177/1461444812440159

Cadwalladr, C. (2018) The Cambridge Analytica Files. The Guardian.

Calvo, J. (2016) McDonald's Japan: AR and IoT Marketing Strategy with Pokemon GO. Journal of Global Economics, 7(2).

Carbonell, I. M. (2016) The ethics of big data in big agriculture. Internet Policy Review, 5(1). https://doi.org/10.14763/2016.1.405

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012) Business intelligence and analytics: From big data to big impact. MIS Quarterly: Management Information Systems. https://doi.org/10.2307/41703503

Colley, A., Thebault-Spieker, J., Lin, A. Y., Degraen, D., Fischman, B., Häkkilä, J., â€¦ Schöning, J. (2017) The geography of Pokémon GO: Beneficial and problematic effects on places and movement. Conference on Human Factors in Computing Systems - Proceedings 2017-May, 1179–1192. https://doi.org/10.1145/3025453.3025495

Cotter, K., Cho, J., & Rader, E. (2017) Explaining the News Feed algorithm: An analysis of the "News Feed FYI" blog. Conference on Human Factors in Computing Systems - Proceedings, Part F1276, 1553–1560. https://doi.org/10.1145/3027063.3053114

Crawford, K., & Schultz, J. (2014) Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms. Boston College Law Review. https://doi.org/10.1525/sp.2007.54.1.23.

Creemers, R. (2018) China's Social Credit System: An Evolving Practice of Control. SSRN Electronic Journal, 12, 59–71. https://doi.org/10.2139/ssrn.3175792

Diehl, T., Weeks, B. E., & Gil de ZÃºÃ±iga, H. (2016) Political persuasion on social media: Tracing direct and indirect effects of news use and social interaction. New Media and Society, 18(9), 1875–1895. https://doi.org/10.1177/1461444815616224

Dumas, J.D. (2006), Computer Architecture, Boca Raton; Taylor and Francis

Eisenhardt, K. M. (1989) Building Theories from Case Study Research. Academy of Management Review, 14, 532–550.

Eisenhardt, K. M., & Graebner, M. E. (2007) Theory building from cases: Opportunities and challenges. Academy of Management Journal, 50(1), 25–32.

Frith, J. (2017) The digital "lure": Small businesses and Pokémon Go. Mobile Media and Communication, 5(1), 51–54. https://doi.org/10.1177/2050157916677861

GonzÃ¡lez, R. J. (2017) Hacking the citizenry? Anthropology Today, 33(3), 9–12.

Green, J., & Issenberg, S. (2016) Inside the Trump bunker, with days to go. Bloomberg Businessweek, 27.

Hameleers, M., & Schmuck, D. (2017) It's us against them: a comparative experiment on the effects of populist messages communicated via social media. Information Communication and Society 20(9), 1425–1444. https://doi.org/10.1080/1369118X.2017.1328523

Horvitz, E., & Mulligan, D. (2015) Data, privacy, and the greater good. Science, 349(6245), 253–256.

Howard, P. N. (2005) New media campaigns and the managed citizen. New Media Campaigns and the Managed Citizen. https://doi.org/10.1017/CBO9780511615986

Howard, P. N., & Bradshaw, S. (2017) Social Media, News and Political Informatio n during the US Election: Was Polarizing Content Concentrated in Swing States?, COMPROP Data Memo, 1–6.

Isaak, J., & Hanna, M. J. (2018) User Data Privacy: Facebook, Cambridge Analytic a, and Privacy Protection. Computer, 51(8), 56–59. https://doi.org/10.1109/MC.201 8.3191268

Jansen, B. J., Flaherty, T. B., Baeza-Yates, R., Hunter, L., Kitts, B., & Murphy, J. (2 009) The components and impact of sponsored search. Computer, 42(5), 98–101. ht tps://doi.org/10.1109/MC.2009.164

Kaiser, B. (2019) Targeted. London: Harper Collins.

Kirkpatrick, M. G., Cruz, T. B., Goldenson, N., Allem, J.-P., Chu, K.-H., Pentz, M. A., & Unger, J. B. (2017) Electronic cigarette retailers use Pokémon Go to mark et products. Tobacco Control, 26(E2), e145–e147.

Kosinski, M., Stillwell, D., & Graepel, T. (2013) Private traits and attributes are pred ictable from digital records of human behavior. Proceedings of the National Acade my of Sciences of the United States of America, 110(15), 5802–5805. https://doi.or g/10.1073/pnas.1218772110

Kosinski, M., Wang, Y., Lakkaraju, H., & Leskovec, J. (2016) Mining big data to ex tract patterns and predict real-life outcomes. Psychological Methods, 21(4), 493–50 6. https://doi.org/10.1037/met0000105

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014) Experimental evidence of massive scale emotional contagion through social networks. Proceedings of the Nat ional Academy of Sciences of the United States of America, 111(29), 10779. http s://doi.org/10.1073/pnas.1412469111

Krippendorff, K. (2004) Content Analysis. Thousand Oaks: Sage.

Manokha, I. (2018) Surveillance: The DNA of platform capital â€" The case of Cam bridge Analytica put into perspective. Theory and Event, 21(4), 891–913.

Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017) Psychological targetin g as an effective approach to digital mass persuasion. Proceedings of the National A cademy of Sciences of the United States of America, 114(48), 12714–12719. http s://doi.org/10.1073/pnas.1710966114

Mcafee, A., & Brynjolfsson, E. (2012) Big Data; The Management Revolution. Har vard Business Review, (October), 1–9.

Messing, S., & Westwood, S. J. (2014) Selective Exposure in the Age of Social Medi a: Endorsements Trump Partisan Source Affiliation When Selecting News Online. Communication Research, 41(8), 1042–1063. https://doi.org/10.1177/00936502124 66406

Monreale, A., Rinzivillo, S., Pratesi, F., Giannotti, F., & Pedreschi, D. (2014) Privac y-by-design in big data analytics and social mining. Epj Data Science, 3(1) https://d oi.org/10.1140/epjds/s13688-014-0010-4

Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018) Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Medical Research Methodology, 18(1), 1–7. https://doi.org/10.1186/s12874-018-0611-x

Neudert, L.-M., Kollanyi, B., & Howard, P. N. (2017) Junk News and Bots during the German Federal Presidency Election: What Were German Voters Sharing Over Twitter? Data Memo, 2(September), 1–5.

O'Neil, C. (2016) Weapons of Math Destruction. New York: Penguin, Random House. https://doi.org/10.1057/s11369-017-0027-3

Pamuru, V., Khernamnuai, W., & Kannan, K. N. (2017) The Impact of an Augmented Reality Game on Local Businesses: A Study of Pokemon Go on Restaurants. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2968221

Pascal, U. (2018) Personalizing Persuasion Architecture: Privacy Harms and Algorithmic News Media. In AAAI.

Persily, N. (2017) Can democracy survive the internet? Journal of Democracy, 28(2), 63–76. https://doi.org/10.1353/jod.2017.0025

Remenyi, D., & Williams, B. (1996) The nature of research: Qualitative or quantitative, narrative or paradigmatic? Information Systems Journal, 6(2), 131–146. https://doi.org/10.1111/j.1365-2575.1996.tb00009.x

Richterich, A. (2018) The Big Data Agenda: Data Ethics and Critical Data Studies. The Big Data Agenda: Data Ethics and Critical Data Studies. London: University of Westminister Press. https://doi.org/10.16997/book14

Simons, H.W. (2001) Persuasion in society. Thousand Oaks, CA: Sage.

Tarran, B. (2018) What can we learn from the Facebookâ€"Cambridge Analytica scandal? Significance, 15(3), 4–5. https://doi.org/10.1111/j.1740-9713.2018.01139.x

Tene, O., & Polonetsky, J. (2013) Big Data for All: Privacy and User Control in the Age of Analytics. Northwestern Journal of Technology and Intellectual Property, 11(5), 239–273.

Tufekci, Z. (2015) Algorithmic Harms beyond Facebook and Google: Emergent Challenges of Computational Agency. Journal on Telecommunications & High Tech Law, 13(23) 203–216. https://doi.org/10.1525/sp.2007.54.1.23.

Ward, K. (2018) Social networks, the 2016 US presidential election, and Kantian ethics: applying the categorical imperative to Cambridge Analytica's behavioral microtargeting. Journal of Media Ethics: Exploring Questions of Media Morality, 33(3), 133–148. https://doi.org/10.1080/23736992.2018.1477047

Webster, J., & Watson, R. T. (2002) Editorial: Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Quarterly, 26(2), xiii–xxiii.

Weeks, B. E., ArdÃ¨vol-Abreu, A., & De ZÃºÃ±iga, H. G. (2017) Online influence? Social media use, opinion leadership, and political persuasion. International Journal of Public Opinion Research, 29(2), 214–239. https://doi.org/10.1093/ijpor/edv050

Wilson, D. G. (2017) The ethics of automated behavioral microtargeting. AI Matter s, 3(3), 56–64. https://doi.org/10.1145/3137574.3139451

Yin, R. K. (2009) Case study research: design and methods (4th ed). Thousand Oaks, CA: Sage.

Youyou, W., Kosinski, M., & Stillwell, D. (2015) Computer-based personality judgments are more accurate than those made by humans. Proceedings of the National Academy of Sciences of the United States of America, 112(4), 1036–1040. https://doi.org/10.1073/pnas.1418680112

Zhang, Y., Wells, C., Wang, S., & Rohe, K. (2018) Attention and amplification in the hybrid media system: The composition and activity of Donald Trump's Twitter following during the 2016 presidential election. New Media and Society 20(9), 3161–3182. https://doi.org/10.1177/1461444817744390

Zittrain, J. (2014) Engineering an election. Harvard Law Review, 127(8), 335–341.

Zuboff, S. (2019) The Age of Surveillance Capitalism; the fight for a human future at the new frontier of power. London: Profile Books Ltd.

Appendix A

Sources for the Facebook Voting Experiment Case

Bond, R.M., Fariss, C.J., Jones, J.J., Kramer, A.D.I., Marlow, C., Settle, J.E. and Fowler, J.H. (2012) A 61-million-person experiment in social influence and political mobilization. Nature. 489, 7415 (2012), 295–298. doi:https://doi.org/10.1038/nature11421.

Haenschen, K. (2016) Social Pressure on Social Media: Using Facebook Status Updates to Increase Voter Turnout. Journal of Communication. 66, 4 (2016), 542–563. doi:https://doi.org/10.1111/jcom.12236.

Jones, J.J., Bond, R.M., Bakshy, E., Eckles, D. and Fowler, J.H. (2017) Social influence and political mobilization: Further evidence from a randomized experiment in the 2012 U.S. presidential election. PLoS ONE. 12, 4 (2017), 1–9. doi:https://doi.org/10.1371/journal.pone.0173851.

Sifry, M. l. (2014) Facebook Wants You to Vote on Tuesday. Here's How it Messed With Your Feed in 2012. Mother Jones.

Tufekci, Z. (2015) Algorithmic Harms beyond Facebook and Google: Emergent Challenges of Computational Agency. Journal on Telecommunications & High Tech Law. 13, 23 (2015) 203–216. doi:https://doi.org/10.1525/sp.2007.54.1.23.

Zittrain, J. (2014) Engineering an election. Harvard Law Review. 127, 8 (2014), 335–341.

Zuboff, S. (2019) The Age of Surveillance Capitalism; the fight for a human future at the new frontier of power. Profile Books Ltd.

Appendix B

Sources for the Cambridge Analytica Trump Campaign Case

Amer, K., Barnett, E. and Kos, P. (2019) The Great Hack. Netflix.

Baldwin-Philippi, J. (2017) The Myths of Data-Driven Campaigning. Political Communication. 34, 4. Pp. 627–633. doi:https://doi.org/10.1080/10584609.2017.1372999.

Berghel, H. (2018) Malice Domestic: The Cambridge Analytica Dystopia. Computer. 51. Pp. 84–89. doi:https://doi.org/10.1109/MC.2018.2381135.

Bessi, A. and Ferrara, E. (2016) Social Bots Distort The 2016 U.S. Presidental Election. First Monday. 21, 11. Pp. 1–15.

Borgesius, F.J.Z., Möller, J., Kruikemeier, S., Fathaigh, R., Irion, K., Dobber, T., Bodo, B. and de Vreese, C. (2018) Online political microtargeting: Promises and threats for democracy. Utrecht Law Review. 14, 1. Pp. 82–96. doi:https://doi.org/10.18352/ulr.420.

Cadwalladr, C. (2018). The Cambridge Analytica Files. The Guardian.

Cotter, K., Cho, J. and Rader, E. (2017. Explaining the News Feed algorithm: An analysis of the "News Feed FYI" blog. Conference on Human Factors in Computing Systems - Proceedings. Part F1276. Pp. 1553–1560. doi:https://doi.org/10.1145/3027063.3053114.

GonzÃ¡lez, R.J. (2017) Hacking the citizenry? Anthropology today. 33, 3. Pp. 9–12.

Howard, P.N. and Bradshaw, S. (2017) Was Polarizing Content Concentrated in Swing Statesâ€¯? Pp. 1–6.

Isaak, J. and Hanna, M.J. (2018) User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection. Computer. 51, 8. Pp. 56–59. doi:https://doi.org/10.1109/MC.2018.3191268.

Kaiser, B. (2019) Targeted. HarperCollins Publishers.

Laterza, V. (2018) Cambridge Analytica, independent research and the national interest. Anthropology Today. 34, 3. Pp. 1–2. doi:https://doi.org/10.1111/1467-8322.12430.

Manokha, I. (2018) Surveillance: The DNA of platform capital â€" The case of Cambridge Analytica put into perspective. Theory and Event. 21, 4. Pp. 891–913.

Richterich, A. (2018) How data-driven research fuelled the Cambridge Analytica controversy. Partecipazione e Conflitto. 11, 2. Pp. 528–543. doi:https://doi.org/10.1285/i20356609v11i2p528.

Tarran, B. (2018) What can we learn from the Facebookâ€"Cambridge Analytica scandal? Significance. 15, 3. Pp. 4–5. doi:https://doi.org/10.1111/j.1740-9713.2018.01139.x.

Wilson, R. (2019) Cambridge Analytica, Facebook and Influence Operations: A Ca se study and Anticipatory Ethical Analysis. European Conference on Cyber Warfar e and Security. Pp. 587–595.

Appendix C

Sources for the Pokémon McDonald's / Starbucks Case

Bauder, M. and Hackenbroch, K. (2018) Public space as Code/Space: The geograph ies of Pokémon Go.

Calvo, J. (2016) McDonald's Japan: AR and IoT Marketing Strategy with Pokemon GO. Journal of Global Economics. 7, 2.

Colley, A., Thebault–Spieker, J., Lin, A. Y., Degraen, D., Fischman, B., Häkkilä, J., Kuehl, K., Nisi, V., Nunes, N. J., Wenig, N., Wenig, D., Hecht, B. and Schöning, J. (2017) The geography of Pokémon GO: Beneficial and problematic effects on pla ces and movement. Conference on Human Factors in Computing Systems – Procee dings 2017. Pp. 1179–1192. doi:https://doi.org/10.1145/3025453.3025495.

Evans, L. and Saker, M. (2019) The playeur and Pokémon Go: Examining the effect s of locative play on spatiality and sociability. Mobile Media and Communication. 7, 2. Pp. 232–247. doi:https://doi.org/10.1177/2050157918798866.

Frith, J. (2017) The digital "lure": Small businesses and Pokémon Go. Mobile Media and Communication. 5, 1. Pp. 51–54. doi:https://doi.org/10.1177/20501579166778 61.

Jin, D. Y. (2017) Critical interpretation of the Pokémon GO phenomenon: The int ensification of new capitalism and free labor. Mobile Media and Communication. 5, 1. Pp. 55–58. doi:https://doi.org/10.1177/2050157916677306.

Kirkpatrick, M. G., Cruz, T. B., Goldenson, N., Allem, J. P., Chu, K. H., Pentz, M. A. and Unger, J. B. (2017) Electronic cigarette retailers use Pokémon Go to market products. Tobacco Control. 26, E2. Pp. e145–e147.

Pamuru, V., Khernamnuai, W. and Kannan, K.N. (2017) The Impact of an Augme nted Reality Game on Local Businesses: A Study of Pokemon Go on Restaurants. S SRN Electronic Journal. doi:https://doi.org/10.2139/ssrn.2968221.

Zuboff, S. (2018) The Age of Surveillance Capitalism. Profile Books Ltd.

Footnotes

[1] ReTargeter (nd) What is Retargeting and How Does It Work. https://retargeter.com/what-is-retargeting-and-how-does-it-work/

[2] Bump, P. (2016) Donald Trump will be president thanks to 80,000 people in three states. Washington Post. https://www.washingtonpost.com/news/the-fix/wp/2016/12/01/donald-trump-will-be-president-thanks-to-80000-people-in-three-states/ (behind paywall)

Cite as

(2021) The Architecture of Algorithm-driven Persuasion. Journal of Information Architecture. Vol. 06. Iss. 01. Pp. xx-xx. http://journalofia.org/volume6/issue1/02-rose/.

Jeremy Rose

University of Skövde (retired)

Jeremy Rose is the (recently retired) Professor of Informatics at Skövde University, where he directed the PhD education.
He has worked with a variety of nationally and internationally funded research projects over the previous twenty years, primarily concerned with IT and organizational change, the management of IT, innovation, e-Government and big data. He publishes in leading eGovernment and information systems journals and served until recently as associate editor for four of them.

Oskar MacGregor

University of Skövde

Oskar MacGregor is a senior lecturer in cognitive neuroscience, with a background in philosophical and practical ethics.
His research interests range from empirical electrophysiology to the ethics and philosophy of privacy in the digital landscape.