The anti-influence engine
Pinder, Charlie

DOI: 10.1145/3027063.3052762
License: Other (please specify with Rights Statement)

Document Version
Peer reviewed version

Citation for published version (Harvard):

Link to publication on Research at Birmingham portal

Publisher Rights Statement:

General rights
Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

• Users may freely distribute the URL that is used to identify this publication.
• Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
• Users may use extracts from the document in line with the concept of ‘fair dealing’ under the Copyright, Designs and Patents Act 1988 (?)
• Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Download date: 03. Aug. 2019
The Anti-Influence Engine: Escaping the Diabolical Machine of Pervasive Advertising

Charlie Pinder
HCI Centre,
School of Computer Science,
University of Birmingham, UK
c.pinder@cs.bham.ac.uk

Abstract
This paper aims to stimulate discussion about the need for and possible incarnations of anti-advert technology. Advertisers are increasingly using pervasive and nonconscous routes to emotionally manipulate people. HCI researchers have yet to provide the tools to counter these unwanted influences. This paper outlines a design fiction solution, the Anti-Influence Engine: a distributed system that returns to users the power over their own associative memories. The Engine gathers advert-exposure information, and offers users multiple ways to counteract emotionally manipulative ads. Design and ethical issues are discussed.

Author Keywords
Anti-advertising technology; nonconscious technology; design fiction; pervasive advertising.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction
This paper is a deliberately provocative design fiction [60]. It draws on behavioural science to explore how technology might counter the ubiquitous conditioning of...
people by manipulative advertisers. It responds to calls for research into enabling people to protect themselves against advertisers seeking increasingly sophisticated routes to affect consumer choice beyond their conscious control [6,62].

The work is in the tradition of Walden 2, a utopian science fiction book by psychologist B F Skinner [58], which uses behavioural science to speculate how technology might improve society. It was also inspired by the CV Dazzle project [29] which explores how fashion can defeat face-detection algorithms, enabling people to protect their privacy against detection technology. The project prompted the question: how might technology itself provide protection against advertisers using increasingly sophisticated algorithms to target people’s associative memories, often without their knowledge?

Delivering large-scale, personalised ads is no longer prohibitively expensive. Advertisers use increasingly pervasive digital channels, such as mobiles [25], social networking sites (SNS) [11] and public displays [43], to deliver ads personalised through big-data-driven individual behaviour profiling [5] and nonconscious advertising techniques [30].

These developments fit Tausk’s conception of a “diabolical” “influencing machine” [63]. The machine “produces ... thoughts and feelings by means of ... mysterious forces which the patient’s knowledge ... is inadequate to explain” [63]. This paper shows how advertising is moving to affect people’s purchasing decisions beyond their conscious attention. It proposes an Anti-Influence Engine system to free people from the diabolical influencing machine of pervasive advertising.

Section 1: The Diabolical Machine
Establishing the extent to which people are manipulated by unwanted advertising is problematic. The first issue is in defining manipulation. Although Sunstein defines the term as influence that “does not sufficiently engage or appeal to [people’s] capacity for reflection and deliberation”, he notes that manipulation is often characterised by “a justified sense of ex post betrayal” [62]. Yet a sense of betrayal is unlikely given evidence that users are influenced by ads but unable to consciously recall them.

This evidence includes research showing that activating pre-existing associations outside of awareness can have an impact on consumer behaviour. Coates et al. showed that nonconscious brand priming increases selection [14]. Lee et al. demonstrated that even irritating animated ads, although initially disliked by consumers, ultimately generated a positive user attitude through the mere exposure effect [39]. The mere exposure effect is where exposure to even neutrally-valenced stimuli can increase the subsequent liking of those stimuli [67]. The effect extends to subliminal exposures, i.e. exposures where people cannot consciously recall being shown a stimulus [8].

Why might people be concerned about the impact of the content of advertising, regardless of recall? Even before the current boom in pervasive ads, researchers expressed concern about possible harmful associative learning such as positive views of unhealthy foods or alcohol [44]. Yet a further problem in determining the potential impact of pervasive advertising is that...
academic analysis of marketing on new technology platforms such as social networks lags behind their popularity [45]. Nevertheless, an analysis of social media drinking ads found regular exhortations to drink, and attempts by advertisers to link drinking with specific cues e.g. a given day [45].

In the broader environment, Dalton et al. note that advertising displays are becoming more pervasive and increasing in size [20]. However, few studies quantify exactly how many ads a given user is shown in naturalistic settings. In 2005, a UK newspaper used a glasses-mounted camera to record the number of ads one of its journalists saw in central London, UK, over 1½ hours [26]. The camera recorded 250 ads with 100 brands in 70 different formats, while the journalist could only recall 1 ad without prompting. Yet caution is needed for newspaper claims on pervasive advertising: some reported claims of 5,000 ads a day [61] appear to be without a research basis [13].

More rigorous research by Dalton et al. tracked shoppers’ eye movements in a large UK shopping centre [20]. On average, participants fixated on 16 ads for 0.318 seconds over a 15-minute task. They were not asked to recall any advertising instances. However, visual response times of 0.1-0.3s are considered to be rapid and automatic [42]. The results therefore suggest that people are unlikely to be able to consciously recall 16 ads seen over 15 minutes at a later date, and that low conscious attention means any “cognitive defences” from ad literacy [56] (outlined below) will not be in place.

There is a clear need for further research into the number, emotional valence, and mode of delivery of ads encountered in naturalistic settings. Without these facts, and in the absence of concrete user recall, it is difficult to complain about manipulation.

What, then, is the basis for believing that the current trends in advertising are any more pernicious than previously? Are advertisers using “dark patterns” [27] that exploit psychology to influence people beyond their intentions? Three broad developments in advertising are particularly concerning: (1) the increase in technology-driven behavioural targeting, i.e. the gathering of user interaction and other behavioural data and the use of it to personalise ads; (2) the rise in use of neuroscience-based physiological monitoring to fine-tune nonconscious responses to ads; and (3) a movement towards ‘native’ ads [38], ads that are concealed within content. These trends are outlined below, before addressing how behavioural science might provide some defences.

1. **Behavioural targeting**

   Behavioural targeting is the use of adaptive user profiles generated from both explicit user-shared information and implicit user information derived from their behaviour such as browsing activity [2]. The profiles are then used to tailor ads. The adaptation and tailoring may occur in real time: a recent patent seeks to “symbiotically” link ads shown in public spaces to personal mobile devices to allow interaction between them [15], while other research has demonstrated how to automatically augment user behaviour profiles with behavioural information [2].

2. **Nonconscious advertising**

   The need for technology to counter pervasive advertising is more urgent because of increased
interest in nonconscious advertising [17], in line with interest in nonconscious techniques in health and behaviour change [53,57]. Acar notes that advertisers are exploring “incidental advertising processing, states of unconscious learning, and preattentive exposure effects” [1]. Advertisers are also using “consumer neuroscience” [34,35] and “neuromarketing” [7,54] to try to maximise the nonconscious impact of ads on recipients. These techniques allow advertisers to establish the affective impact of their work over and above conscious self-report, and give advertisers the power to manipulate people’s emotions and attitudes beyond their conscious control.

3. ‘Native’ advertising
The advertising trend is towards ‘native ads’ —ads that are integrated within content so they are difficult to distinguish from content, such as Twitter and Facebook’s sponsored content, ads in search results and newspaper articles [19,38]. These ads are resistant to ad-blocking plugins [66], making opting out difficult. The difficulty in distinguishing ads from actual content is highlighted by the development of a plugin to specifically detect and flag up native ads [69]. Although some business researchers view this form of advertising positively as non-disruptive [11], there are serious concerns amongst less vested interests, including journalists, e.g. Robert Peston’s speech against “news that is a disguised advert” [49], psychologists e.g. Bargh [6] and legal scholars e.g. Sunstein [62].

These three trends combine to form a serious asymmetry between advertisers and their targets. Advertisers know who has been watching their ads and when, with what emotional affect and behavioural effect, with what interaction and in what context, while users are unable to consciously recall what ads they have seen. Couldry & Turrow [16] argue that this asymmetry of information and “deep personalisation” risk threatening democracy itself by eliminating collective experience: advertisers will be able to show different versions of reality to different audiences.

Users may agree to surrender their behavioural data in exchange for technological advance or social network access, but they are likely to remain unaware of exactly how their personal data is used to manipulate content to elicit strong affective associations towards products embedded into their daily lives. And, unfortunately, awareness affords no cognitive defence to nonconscious manipulation.

Social networks have been criticised for conducting large-scale field experiments in emotional manipulation without the knowledge of their users [37,41]. Yet advertisers are doing the same, without even ‘implicit’ user consent, without attracting controversy.

Possible solutions
Although users can avoid some ads on some platforms, e.g. by fast-forwarding through time-shifted television content, what defence do users have against manifestations of this diabolical influencing machine embedded within SNS, search engines, internet email, games, news sites, newspapers, magazines, public displays, etc.?

Policy-makers often advocate an advertising literacy strategy [3]. This approach implicitly accepts a conscious “cognitive defence” model: that critical thinking about advertising can mitigate its effects [56]. However, as Rozendaal et al. (ibid) point out, there are
both theoretical and empirical reasons to doubt the approach. On the theory side, in line with dual process theories [24] including the Elaboration Likelihood Model [50], the cognitive defence is useless in situations where conscious attention is not directed towards a given advert, and the evidence suggests that children, for example, are "avid multitaskers" [3]. On the empirical front, there is a lack of evidence to support the efficacy of such interventions. Likewise, strategies to limit media exposure [3] are unlikely to be successful in the context of advertising that increasingly encroaches into more technology and public spaces.

At the core of the Anti-Influence Engine is the assumption that people are affected by associative processes. The association of brands with positive affect is the key problem in advert-based manipulation: advertisers seek to associate memories of their brand with positive experiences in order to make their brand more likely to be recognized and retrieved in the future [54]. Plassman et al. argue that an important predictor of a person’s choice between brands is their memory of previous exposures to those brands, which may be formed on an unconscious level [54]. Dalton et al. note that advertising displays are focusing more on attempts to associate brands with positive affective experiences [20].

Stayman & Batra provide evidence this strategy works to boost retrieval: positive affect at the point of ad exposure can speed up retrieval, particularly in low involvement conditions [59]. Similarly, Pham & Vanheule, showed that even fragments of ads fragments can trigger activation in an associative network [51].

In an advert-free world, brand selection would be a function of expected utility, formed primarily from a person’s past experience together with some minimal inputs from brand packaging and perhaps from word-of-mouth. Manipulative advertising seeks to inflate the expected utility function by falsely associating the brand with positive affect. The key to solving the problem is therefore to (a) capture the valence of the false effect, and (b) to neutralise it by exposing the user to an association of the brand with a diametrically opposite negative affect. This is the basis of the Anti-Influence Engine.

Section 2: The Anti-Influence Engine
This section outlines our proposed solution to the problem of unwanted manipulative pervasive advertising. Our “Anti-Influence Engine” has two key subsystems: the first gathers information about ads a user experiences, and the second retrain them to counter the effects of those ads.

1. Gather subsystem – unobtrusively gathers information about all ads the user experiences and annotates them with relevant contextual and affective information.

2. Retrain subsystem – retrieves pre-seen ads and presents users with retraining in various forms depending on platform. Customisable to allow users to increase or decrease their preference for ads and/or other items.

The Gather subsystem operates continuously; the Retrain function runs at opportune moments—including during user sleep—as outlined below.
Gather subsystem
The Anti-Influence Engine first gathers candidate ads the user experiences within both the wider environment (bus stops, public displays, ads in magazines, etc.) and from personal screen technology (all computer displays, TVs, etc.). Candidate ads are processed to extract contextual and affective data, including:

- brand name or product itself
- length of exposure (ms)
- mean size of exposure in field of vision (mm)
- volume (for audio/video)
- visual field of exposure (peripheral vs foveal)
- platform (e.g. public display or magazine)
- emotional valence of adjacent stimuli, i.e. a measure of the affective images advertisers have used in their ad
- contextual information (location; time; physiological markers, etc.)

Ad objects are annotated with these extracted values. Each object is given an overall score as a function of these values to give a priority list of candidates for the retraining phase. The objects are stored to provide input to later retraining.

The gather subsystem comprises a set of Sense Monitors and a set of Technology Monitors. Sense Monitors comprise EyeWear and ContextWear modules. EyeWear is a lightweight, unobtrusive gaze tracker [10] with video and audio capture capability integrated into fashion glasses to gather visual and audio ad information, including the brand itself, affective images and/or words associated with the brand and, for visual ads, gaze activity around the ad. ContextWear is a smartphone module used to report a user’s current context, including collating data from any available physiological monitors (e.g. activity, heart rate, blood pressure monitors).

Technology Monitors include automated screen capture and processing plug-ins for all computer screens, specifically tailored to capture both native and non-native ads. When candidate ads are captured by both systems, for example when the EyeWear captures a native ad on a SNS, a disambiguation module runs to ensure the candidate ad is represented only once in the retraining database.

Retraining subsystem
The core of the retraining subsystem is the use of aversive evaluative conditioning and cognitive bias modification techniques to counter advertising exposures. The subsystem primarily uses an incidental approach to deliver training, i.e. the repurposing of existing behaviour [21,52]. The training is delivered via a distributed ecosystem across all a user’s devices: it may appear on any screen-based technology from a smartphone to a TV to a smart fridge. The Anti-Influence Engine uses an anticipatory behavioural model to predict which technology a user will use next and for how long. It then selects an appropriate retraining mechanism based on this platform, behavioural predictions and the ad attributes outlined above.

Retraining options are:

- Aversive evaluative conditioning: where a brand is juxtaposed with a negatively valenced image.
Cognitive bias modification training: *push or swipe away* gestures are hijacked to include a stimulus a user wants to avoid, e.g. rejecting an unhealthy food item using a swipe away to unlock a smartphone [52].

Aversive evaluative conditioning is the pairing of an unpleasant stimulus with a target item to alter its emotional valence. The Anti-Influence Engine pairs target ad images with unpleasant stimuli. The estimate of the valence of the positive affect for the brand generated from analysing the ad is used to select an unpleasant stimulus of the same, opposite valence. The overall aim is to ‘reset’ the affective association of the stimulus to neutral.

Aversive conditioning has been used elsewhere in HCI [18,36]. Our approach is based on evidence that negative emotional arousal is related to poorer associative memory recall [28]. Brands paired with unpleasant stimuli become less likely to be selected. These unpleasant stimuli may be an image [32], a sound [9], a smell [4] or even an electric shock [48]. For example, a user may have to pair an image of maggots with a brand image of crisps to unlock their phone or open their fridge, or an interim screen while switching TV channels might feature a brief unpleasant-brand pairing presentation.

Cognitive bias modification techniques aim to retrain problematic automated paths within the brain [31]. The Anti-Influence Engine’s implementation of these techniques is based on evidence that ‘push away’ gestures can retrain attention bias for unwanted stimuli and impact on user behaviour by ultimately reducing the real-life selection of those stimuli [65].

**User options**
The Anti-Influence Engine is also configurable by users. Users can configure:

1. User goals. These affect the types of ads shown in the retraining phase. For example, a user might choose to undo any associations of unhealthy foods with pleasure.
2. Training. Where, when and how the training should be delivered. Users might select a
specific time to perform their retraining, or prioritise a specific platform.

3. Specific stimuli. Personalization options enable users to alter the wantedness or otherwise of detected ads and to add their own stimuli. For example, say a PhD student identifies an unwanted fondness for a certain brand of crisps. She can upload an image of the crisps to the system and mark it as a problematic item on which to receive aversive training.

**Benevolence & Ethics**
This system gives people the power to influence what they believe and how they act. Although it has benevolent aims, i.e. to allow individuals to avoid manipulation by unwanted outside sources, there is a clear ethical tension in giving people this power. One usage scenario outlined above enables a user to devalue memories of crisps. However, other users might have more unpalatable aims, e.g. to enforce a gender or race bias, or seek to influence their nonconscious minds to avoid food altogether.

The Anti-Influence Engine might also offer a specific religion module, intending that users can use it to devalue their attitudes towards culturally-imposed religious beliefs. This might be used for the opposite purpose. Future creators of the system will need to consider whether they disable or restrict such reversals.

**Future Work**
Our solution considers time-shifted attempts to redress advertising manipulation: user exposure is tracked and users are retrained later. However, this does not address real-time in situ manipulation that affects consumer choice e.g. shelf placement [12] or music [47]. A remaining challenge is to counter real-time manipulation without disrupting users.

The simple pairing of unwanted ads with aversive images may be insufficient to reverse years of pre-Anti-Influence Engine exposure to ads. If so, the Anti-Influence Engine could be augmented with a Pain Module to administer aversive training with electric shocks, as with other consumer pain technology [48].

**Discussion**
The Anti-Influence Engine is a design fiction solution to a current real-life problem using near-future technology. The Engine seeks to return to individuals the power over their own associative memories, in response to advertisers altering these memories beyond people’s knowledge or control. The Engine gives people multiple escape routes from the diabolical influencing machine constructed by advertisers who are increasingly focused on pulling the levers of nonconscious control.

The technology-mediated future of pervasive manipulation looks bleak: it is easy to anticipate reactive pervasive displays that draw on data about user reactions from live physiological monitoring [40] and facial expressions [43], in-store movements [23] and data from user profiles on their own synced technology [15] to deliver personalised, maximally affective ads. Companies are likely to participate in real-time bids for the right to target particular users in particular locations via multiple channels. For example, a company may wish to present a particular food brand to a happy, hungry user on a large-scale display whilst sending a discount code to their smartwatch and directions to their smartphone for the nearest outlet.
Realistically, comprehensively countering such pervasive advertising is an onerous task. Research into the number and content of pervasive ads lags behind both technical developments and behind research into ever-more intrusive ways to target individual nonconscious processes. Likewise, there is a lack of research into effective means for countering advertising. Although research into the use of cognitive bias modification techniques on smartphones and tablets is starting to emerge [22,52], its efficacy is unknown, and unexplored on other platforms. HCI research into aversive evaluative conditioning is sparse despite evidence of its ability to alter implicit attitudes and subsequent behaviour [32,64].

This paper is intended to open a debate on how best to start countering the future manipulation of all corners of our lives—and all our technologies—by advertisers. We have much work to do.

Acknowledgements
I am indebted to the actual fiction of Jeff Noon [46], worryingly turning into a design reality. I would also like to thank all my reviewers for their helpful feedback.

References


41. Michelle Meyer. Everything You Need to Know About Facebook’s Controversial Emotion Experiment. WIRED. Retrieved January 9, 2017 from goo.gl/dpF2CO.


43. Jörg Müller, Florian Alt, Daniel Michels, and Albrecht Schmidt. 2010. Requirements and design space for interactive public displays. MM ’10, ACM, 1285.


65. Reinout W Wiers, Carolin Eberl, Mike Rinck, Eni S Becker, and Johannes Lindemeyer. 2011. Retraining automatic action tendencies changes alcoholic patients’ approach bias for alcohol and improves treatment outcome. Psychological Science 22, 4: 490–7.


