Towards Effective Ethical Behavior Design

Rodrigo de Oliveira*
Telefonica Research
Pl. de Ernest Lluch i Martín 5
08019 - Barcelona
oliveirar@google.com

Juan Pablo Carrascal
Universitat Pompeu Fabra
Carrer de Tànger, 122-140
08018 - Barcelona
jp.carrascal@upf.edu

Abstract
Many of today's persuasive systems are designed taking into account cognitive biases to foster positive changes in people's behavior (e.g. adopt greener attitudes). However, these biases are also exploited to shape the users' behavior in a way that not necessarily benefit them (e.g. user retention in a website). Scholars addressed this problem by developing design guidelines and methods for ethics in persuasive computing, but these approaches alone have proved to be inefficient since they require every designer to be aware, understand, and comply with the recommended ethical practices. We propose preventive approaches that shall support higher compliance, as well as a remediation-based approach that does not require compliance from every designer. These approaches aim to help users understand persuasive elements embedded in systems, as well as to take more rational decisions when interacting with them. We expect that using preventive and remediation-based approaches will more effectively implement ethics in behavior design.

Author Keywords
Persuasive computing; Ethics; Cognitive biases

ACM Classification Keywords
H.5.m [Information Interfaces and Presentation (e.g. HCI)]: Miscellaneous; K.7.4 [The Computing Profession]: Professional Ethics

* The author is currently affiliated with Google Inc., USA.

CHI 2014, April 26–May 1, 2014, Toronto, Ontario, Canada.
ACM 978-1-4503-2474-8/14/04.
http://dx.doi.org/10.1145/2559206.2581182

Motivation

One of the greatest faculties of the human brain is the ability to reduce the enormous amount of information surrounding us to a proper size for processing and yet keeping relevant information as proved by millions of years of adaptation. However, this extraordinary gift is subject to a number of cognitive biases that can lead people to take decisions not in their best interest. Some examples include people's frequent poor dietary choices and sedentary lifestyle [6], which can cause several illnesses, such as diabetes and heart disease.

Understanding and changing these behaviors have been the main research topic of a large body of work in persuasive computing [12]. Many scholars have leveraged persuasive techniques to shape people's behavior for preferable outcomes. For example, Arroyo et al. [7] developed Waterbot, a system that can be installed on household faucets to motivate people to turn off the tap when the water is not being used. Green attitudes have also been promoted for sustainable uses of energy resources [21]. In preventive health, Oliveira el al. [11] developed a mobile social game to help patients become more adherent to their medication prescription.
Lee et al. [16] evaluated several persuasion techniques to promote healthier eating habits in the workplace, including the default bias and planning strategies.
Consolvo et al. [9] used the addictiveness of game playing to fight obesity, And related to privacy, Wang et al. [20] proposed man-in-the-middle-like technologies to mitigate biases when sharing information in online social networks.

These are only a few examples highlighting how proficient our scientific community has become in fostering positive changes in people's behavior. However, in many cases it is not clear what a positive behavior change is. Should users be persuaded to spend more time checking their online social network feeds? Should they be influenced to buy more items in an e-commerce website? Designers of these technologies might think so, but it is not clear that these target behaviors are positive for users. Ethical[1] issues arise when technologies are designed to shape the users' behavior towards a target not intentionally defined by them. As pointed out by Smids [17], the voluntariness condition is key and must consider external influences and whether the user acts intentionally.

In persuasive computing, scholars have proposed guidelines and methodologies to support designers in ethical behavior shaping. Berdichevsky and Neuenschwander [8] were among the first researchers to propose guidelines, suggesting in their golden rule of persuasion that "creators of a persuasive technology should never seek to persuade a person or persons of something they themselves would not consent to be persuaded to do". In a similar attempt, Fogg [12] suggested 7 steps for designers to evaluate the ethical nature of a persuasive technology by examining its intentions, methods and outcomes. Discourse ethics was proposed to search for further guidelines in the field [18]. Several methodologies for ethical design have been proposed, including value sensitive design [10], persuasion profiling [14], and deep involvement with stakeholders and users [15]. And in personal informatics, systems have been designed to encourage and support self-reflection and self-behavior management [4].

Although very enlightening, the aforementioned principled approaches depend on designers' awareness, understanding, and commitment to ethical practices. However, there is a number of examples revealing that designers have not embraced such practices [1]. We believe there is an urgent need to combine today's

[1]In this paper we refer to *ethical* design as the process by which designers create persuasive technologies following guidelines and methodologies suggested for ethical behavior shaping [8, 12].

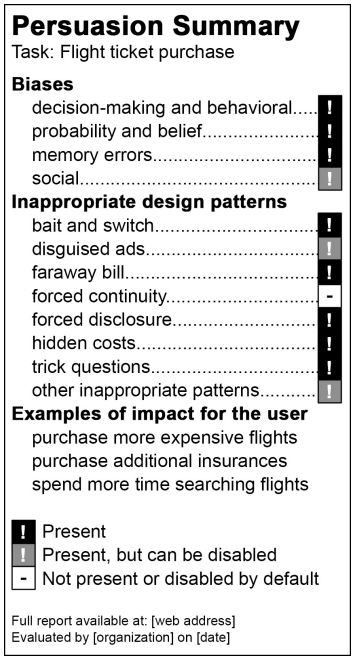

Figure 1: Mock up of a persuasion summary label describing persuasive elements used by a website that sells flight tickets. The label could be presented on the website, in online application stores, or attached to the physical software box sold in physical stores. More details on the list of inappropriate design patterns can be found in [1].

principled-based approaches with more effective solutions to implement ethics in persuasive computing.

Proposed Approaches

In self-beneficial behavior shaping, designers apply various persuasive techniques to shape the users' behavior towards targets defined *by the users* (e.g., eat healthier, quit smoking). In these cases, designers are not required to remove biases, but they rather introduce new ones, supposedly "stronger", to help users change some of their undesired behaviors.

However, this does not apply to cases where users do not know how they should best behave *and* would like to take a more rational decision about it (e.g. deciding whether to use a certain service more often or not). While some designers have addressed this user need supporting self-reflection for unbiased decision making [4], others have designed technologies that shape the user's behavior for their own benefit [1]. From an analysis of previous work in the field and today's motivations for building persuasive technologies, we highlight three main issues that should be addressed in order to effectively support ethical behavior design:

1. Lack of Awareness: Today's approaches are based on providing awareness of guidelines and methodologies for ethical behavior design, mostly in scientific journals and conferences (e.g. [8, 12, 10, 14, 18, 15, 4]). However, not every designer has access to these channels or is proactively engaged in the research community.

2. Lack of Understanding: As noted by Torning et al. [19], ethical considerations have been often mentioned, but not clearly addressed. We further highlight that not only it is unclear how designers should best apply related theoretical concepts in commercial products, but also that consumers are not knowledgeable about how persuasive technologies shape their behavior. Hence why some designers came up with initiatives to collectively create consumer-targeted educational content [1].

3. Lack of Commitment: By disregarding ethics in behavior design and leveraging persuasive techniques to increase user base and user retention, designers were able to quickly increase their profits [13]. Hence, designers have a conflict-of-interest that prevent them from committing to ethical practices.

Next we propose three approaches to address these issues.

Enforced Prevention

Preventive approaches that *enforce* the use of ethical guidelines [8, 12] could leverage the influence of regulatory entities, such as the government or organizations with decision power to authorize and deny commercialization of technologies in a certain territory. Specialized committees could be created to evaluate ethics of persuasive technologies. These evaluations should yield consumer-targeted summaries with information about the persuasive techniques used in each technology studied by the committee. These summaries could be designed based on consumer labeling efforts capable of presenting complex information in a concise and easy to understand format, such as the "Nutrition Facts" label [2]. Figure 1 shows an example of our proposed design.

The main advantages of this approach include: (1) making consumers *aware* of ethical issues in persuasive technologies, (2) providing them with the tools to *understand and judge* between competing technologies, and (3) ensuring the designers' *commitment* to ethical practices by means of regularization. However, it is limited by the complexity and time demand for proposing, discussing, voting, and implementing regulation laws that prevent bad cases of behavior design.

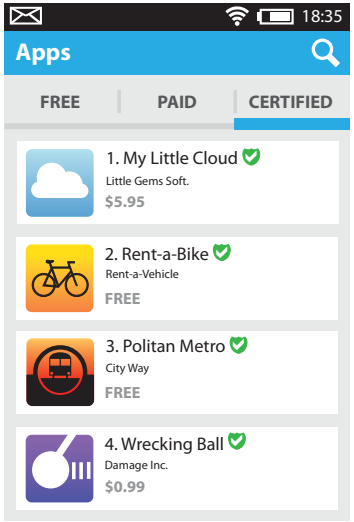

Figure 2: Example of active encouragement to ethical behavior design. The mobile search interface groups results by certified applications, i.e. apps evaluated by an expert committee and certified to meet minimum requirements for ethical behavior design. This special category gives higher visibility to designers of these apps, while also enabling consumers to enter a space where they feel safe when searching and downloading applications.

Actively Encouraged Prevention

The aforementioned approach can be envisioned in a scenario where designers are not obliged to adhere to ethical guidelines, but rather *actively encouraged* to do so. By active encouragement, we mean to *ease* and *motivate* the adoption of ethical design practices and the generation of consumer-targeted summaries.

In that sense, designers should have easy access to guidelines and methodologies for ethical behavior design, as well as to organizations that evaluate compliance to these ethical practices. In addition, consumers should be able to easily identify and check whether technologies meet requirements, e.g. with visual cues, like those from certified websites in the internet security domain [5].

In terms of motivation, designers' compliance to ethical practices could be rewarded by displaying certification stamps next to names of websites or applications—listed in search results—that were awarded the certification. Similarly, online application stores could include a specific category for searching "certified apps", thus supporting higher visibility of apps that comply with ethical design practices, as well as allowing consumers to enter a space where they feel comfortable searching and downloading applications (see Figure 2). Once these certifications become ubiquitous, we expect consumers to give preference for using certified persuasive technologies, and hence companies shall be more motivated to follow guidelines from persuasive computing. Further research is needed to identify appropriate incentives for the first adopters of these certifications, such as tax deductions, privileged governmental partnerships, among others.

Although actively encouraged prevention is not limited by the lengthy process of establishing regulation laws, it adds extra complexity for implementing different incentive schemes and does not guarantee commitment from every designer.

Remediation-based Approach

Preventive approaches are not perfect, and hence consumers might be always surrounded by technologies that do not comply with ethical design guidelines. We propose a remediation-based approach to empower users to reveal and remove—or at least mitigate—biases in persuasive technologies. We envision its implementation by groups of designers that implement adaptive solutions for technologies that misuse behavior design. Three activities shall be conducted by these groups:

1. Identify the most relevant biases in the given persuasive technology;
2. Provide interactive mechanisms that enable users to reveal these biases for self-awareness and educational purposes;
3. Provide intervention methods that empower users to remove or mitigate the effect of existing biases.

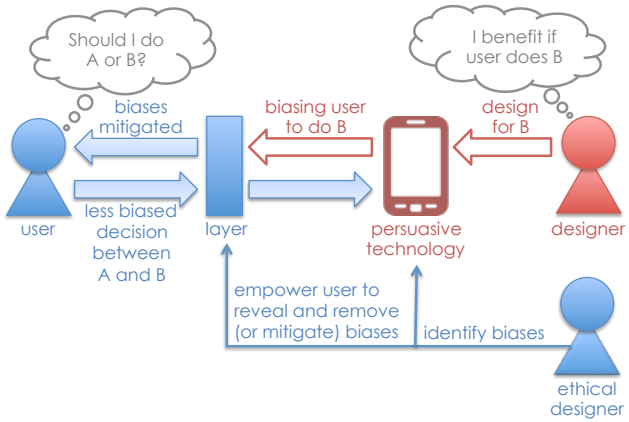

Figure 3: Schematic of our remediation-based approach.

An schematic representation of our approach is presented in Figure 3. It does not require commitment from every designer, like in [8, 12, 10, 14, 18, 15]. Smaller groups of ethically-conscious designers shall be able to implement it as an extra layer interfacing users and today's persuasive technologies. Next we present a few examples about how this remediation-based approach can be realized.

Example 1: Mobile. Consider a user that is thinking about temporarily disabling his/her phone's GPS sensor due to privacy concerns. Many of today's smartphones introduce biases that prevent this behavior. In some cases, the functionality is somewhat "hidden" in the settings' menu, requiring more effort on the user's side to find it (mental/time demand). In other cases, a discouraging message is displayed after disabling the sensor, further asking for the user's confirmation (loss aversion[2]). Our approach could be implemented by designers—not necessarily related to the phone maker—that *identify* the presence of the given biases, and develop a mobile application to empower users to *reveal* and *remove* these biases. In that sense, the mobile application could offer an easy and fast way for users to turn the GPS sensor on and off, such as through a shortcut button or a gesture based interaction.

Example 2: Web. Let us use the famous example of default bias in web forms [6] to exemplify our approach. Consider a user that is buying a flight ticket online using a desktop computer. After s/he makes the flight reservation, the website automatically adds a special customer service for €30, expecting that the default bias will discourage the user from opting-out the service. In this case, designers—not necessarily related to the site owners—could implement our approach by studying the website and identifying the presence of the given default bias; and implementing a browser plugin that can be installed in the users' computers, enabling them to *reveal* and *remove* the default bias. In order to remove it, the plugin could restructure the webpage including an extra modular step that inquiries users on whether they want or not to include the special service.

Example 3: Social Networks. Removing biases is not always possible. In these cases, our approach can be implemented by means of intervention methods that do not specify a target behavior. For example, should users browse social network websites for longer periods of time than they currently do? In this case, our approach could be implemented by revealing the main biases that increase user retention time, and enabling users to activate interventions that mitigate these biases. An example of intervention is to provide daily summaries about the user's time spent in these websites. Other interventions that go beyond self-reflection could prove to be more effective and should be systematically tested, thus opening different avenues of research in persuasive computing.

Conclusions and Future Work

In this paper we proposed preventive and remediation-based approaches to more effectively implement ethics in behavior design. We discussed the scope of each of these approaches, highlighting their advantages and disadvantages. We focused our proposal on a remediation-based approach that can be implemented by small groups of ethically-conscious designers, hence not requiring compliance from every designer, like in most of today's approaches [8, 12, 10, 14, 15].

Many questions remain open for future research and discussion. Besides evaluating our remediation-based

[2]Loss aversion bias affects how we think about our possessions, making the loss of giving up something more salient than the gains of acquiring something else [6]. The message for turning the GPS off focus on what the user loses when turning the sensor off, whereas no discouraging message is usually presented when turning the GPS on.

approach, we would like to investigate to what degree of automation it would be more successful. Although we give focus on a manual procedure for inspection of persuasive technologies conducted by a committee of designers, one could envision a hybrid approach that automatically identify certain biases for later manual inspection.

Another topic for discussion is the qualification required to be a designer of persuasive technologies. One could envision designers being certificated by authorities or organizations trusted for implementing appropriate guidelines and methods, in a similar way that it is done for other domains, like internet security [5] or quality management systems [3]. Alternatively, groups of designers could build their reputation by alerting users of biases present in the services they use everyday [1], and thus avoid going through a potentially long certification process. We believe both cases have potential when designers further provide interactive mechanisms that educate users in revealing biases in technologies, *and* empower users to remove or mitigate these biases.

We hope that our proposed approaches generate fruitful discussions in the HCI community to effectively implement practical solutions for ethical behavior design.

References

[1] Dark patterns. Available at: www.darkpatterns.org.

[2] FDA U.S. Food and Drug Administration. How to Understand and Use the Nutrition Facts Label. Available at: www.fda.gov/food/ingredientspackaginglabeling/labelingnutrition/ucm274593.htm.

[3] ISO - International Organization for Standardization. Available at: www.iso.org.

[4] Personal informatics. Available at: www.personalinformatics.org.

[5] VeriSign SSL certificates. Available at: www.verisign.com.

[6] Ariely, D. *Predictably irrational: the hidden forces that shape our decisions*, 1 ed. HarperCollins, Feb. 2008.

[7] Arroyo, E., Bonanni, L., and Selker, T. Waterbot: exploring feedback and persuasive techniques at the sink. In *Proc. of CHI*, ACM (2005), 631–639.

[8] Berdichevsky, D., and Neuenschwander, E. Toward an ethics of persuasive technology. *Communications of the ACM 42*, 5 (1999), 51–58.

[9] Consolvo, S., McDonald, D. W., Toscos, T., Chen, M. Y., Froehlich, J., Harrison, B., Klasnja, P., LaMarca, A., LeGrand, L., Libby, R., Smith, I., and Landay, J. A. Activity sensing in the wild: a field trial of ubifit garden. In *Proc. of CHI*, ACM (2008), 1797–1806.

[10] Davis, J. Design methods for ethical persuasive computing. In *Proc. of Persuasive*, ACM (2009), 6.

[11] de Oliveira, R., Cherubini, M., and Oliver, N. Movipill: improving medication compliance for elders using a mobile persuasive social game. In *Proc. of Ubicomp '10*, ACM (New York, NY, USA, 2010), 251–260.

[12] Fogg, B. J. *Persuasive technology: using computers to change what we think and do*. Morgan Kaufmann, 2002.

[13] Helft, M. The class that built apps, and fortunes. *The New York Times*, 7 May 2011. Retrieved from www.nytimes.com/2011/05/08/technology/08class.html?partner=rss&emc=rss&_r=0.

[14] Kaptein, M., and Eckles, D. Selecting effective means to any end: Futures and ethics of persuasion profiling. In *Proc. of Persuasive Technology*, Springer (2010), 82–93.

[15] Karppinen, P., and Oinas-Kukkonen, H. Three approaches to ethical considerations in the design of behavior change support systems. In *Proc. of Persuasive*, Springer-Verlag (Berlin, Heidelberg, 2013), 87–98.

[16] Lee, M. K., Kiesler, S., and Forlizzi, J. Mining behavioral economics to design persuasive technology for healthy choices. In *Proc. of CHI 2011*, 325–334.

[17] Smids, J. The voluntariness of persuasive technology. In *Proc. of Persuasive*, Springer-Verlag (Berlin, Heidelberg, 2012), 123–132.

[18] Spahn, A. And lead us (not) into persuasion?? persuasive technology and the ethics of communication. *Science and engineering ethics 18*, 4 (2012), 633–650.

[19] Torning, K., and Oinas-Kukkonen, H. Persuasive system design: state of the art and future directions. In *Proc. of Persuasive Technology*, ACM (2009), 30–37.

[20] Wang, Y., Leon, P. G., Scott, K., Chen, X., Acquisti, A., and Cranor, L. F. Privacy nudges for social media: an exploratory facebook study. In *Proc. of WWW* (2013), 763–770.

[21] Yun, R., Scupelli, P., Aziz, A., and Loftness, V. Sustainability in the workplace: nine intervention techniques for behavior change. In *Proc. of Persuasive*, Springer (2013), 253–265.