Benevolent Deception in Human Computer Interaction

Eytan Adar University of Michigan Ann Arbor, MI 48109 eadar@umich.edu

Desney S. Tan Microsoft Research Redmond, WA 98052 desney@microsoft.com Jaime Teevan Microsoft Research Redmond, WA 98052 teevan@microsoft.comm

ABSTRACT

Though it has been asserted that "good design is honest," [42] deception exists throughout humancomputer interac- tion research and practice. Because of the stigma associated with deception-in many cases rightfully so-the research community has focused its energy on eradicating malicious deception, and ignored instances in which deception is posi-tively employed. In this paper we present the notion of be- nevolent deception, deception aimed at benefitting the user as well as the developer. We frame our discussion using a criminologyinspired model and ground components in various examples. We assert that this provides us with a set of tools and principles that not only helps us with system and interface design, but that opens new research areas. After all, as Cockton claims in his 2004 paper "Value- Centered HCI" [13], "Traditional disciplines have delivered truth. The goal of HCI is to deliver value."

INTRODUCTION DEFINING DEVEIT DECEPTION IN HCL THE MOTIVE MENTAL MODEL VS. SYSTEM IMAGE INDIVIDUAL VS. GROUP INDIVIDUAL VS. GROUP DESIGN GOAL VS. DESIGN GOAL MIXED MOTIVE DECEPTIONS THE MEANS SYSTEM IMAGE DECEPTIONS BEHAVIORAL DECEPTIONS MENTAL MODEL DECEPTIONS

IMPLICATION FOR RESERACH THE OPPORTUNITY GETTING CAUGHT NOT GETTING CAUGHT IMPLICATIONS FOR RESEARCH CONSTRUCTION OF DECEPTION AND ETHICS CONDUSIONS AND FUTURE RESERCH ACKNOWLEGEMENTS

ARTICLES

The Elements of Typographic Style

Fuck Content

Crystal Goblet

- A Typographic Discourse for the Distaff Side of Printing
- A Too-Perfect Picture A Man of Letters In Defense of the Poor Image All Technology is Assistive
- Over the Bridge The Web's Grain A Vernacular Web

Users generally trust computer interfaces to accurately re-flect system state. Reflecting that state dishonestly– through deception–is viewed negatively by users, rejected by designers, and largely ignored in HCI research. Many believe outright deception should not exist in good design. For example, many design guidelines assert: "Do not lie to your users" (e.g., [40, 45]) Misleading interfaces are usually attributed to bugs or poor design. However, in reality, de- ceit often occurs both in practice and in research. We con- tend that deception often helps rather than harms the user, a form we term benevolent deception. However, the over- loading of "deception" as entirely negative coupled with the lack of research on the topic, makes the application of de- ception as a design pattern problematic and ad hoc.

Benevolent deception is ubiquitous in real-world system designs, although it is rarely described in such terms. One example of benevolent deception can be seen in a robotic physical therapy system to help people regain movement following a stroke [8]. Here, the robot therapist provides stroke patients with visual feedback on the amount of force they exert. Patients often have selfimposed limits, believ-ing, for example, that they can only exert a certain amount of force. The system helps patients overcome their percep- tive limits by underreporting the amount of force the patient actually exerts and encouraging additional force.

The line between malevolent and benevolent deception is fuzzy when the beneficiary of the deception is ambiguous. For example, take the case of deception in phone systems to mask disruptive failure modes: The connection of two indi- viduals over a phone line is managed by an enormous spe- cialized piece of hardware known as an Electronic Switch- ing System (ESS). The first such system, the 1ESS, was designed to provide reliable phone communication, but given the restrictions of early 1960s hardware, it sometimes had unavoidable, though rare, failures. Although the 1ESS knew when it failed, it was designed to connect the

As is the case with the 1ESS and placebo buttons, deception sometimes benefits the system designer, service provider, or business owner. However, this does not invalidate the fact that it might also help meet user needs. We believe that by not acknowledging that there is deception, and, more critically, that a line between beneficial and harmful deceptions might exist, research in the area is difficult to pursue-to the detriment of academics and practitioners alike.

A further example of benevolent deception are the "placebo buttons" that allow users to feel as though they have control over their environment when they actually do not. Cross- walk buttons, elevator buttons, and thermostats [33, 47] often provide no functionality beyond making their users feel as though they can affect their environment. Some of these buttons go far to provide the illusion of control; non- working thermostat buttons, for example, are sometimes designed to hiss when pressed [2]. In addition to providing the feeling of control, placebo buttons can signal the exist- ence of a feature to the user. Non-working crosswalk buttons, for example, clearly convey to a pedestrian that a crosswalk exists.

Whether intentional or not, implicit or explicit, acknowl- edged or not, benevolent deceit exists in HCI. Nonetheless, little is known about the motivation, mechanisms, detecta- bility, effectiveness, successes, failures, and ethics of this type of deception. Researchers have tiptoed around this taboo topic, concentrating instead on malevolent deception (e.g., malware or malicious software [14,17]) and unobjec- tionable forms of deception described using entertainment metaphors (e.g., magic or theater [32,54]). This limited view of deception does not capture its variety or ubiquity.

As we will see, one of the underlying reason for the ubiqui- ty of deception is that it can fill the many of the gaps and tensions that emerge with different design concerns (e.g., the good of the individual versus the good of the group), design goals (e.g., conflicting principles), or systems states (e.g., desired system performance versus actual system per- formance). In any situation where a poor fit exists between desire (e.g., the mental model or user expectations) and reality (e.g., the system itself) there is an opportunity to employ deception. This gap—which is extremely com- mon—both motivates and

either by hiding truthful information or showing false in- formation. Deception is an act of deceit with implied intent (e.g., telling the user the web page is 70% loaded when we know that this is not the case). On the other hand, decep- tive(ness) does not require intent (e.g., telling the user that the web page is absolutely 70% loaded based on some esti- mate with high error margins). Though this distinction is important as it speaks to motive, deceit exists with or with- out intent. In fact, when deciding whether an advertisement is illegal (false), the FCC only considers the deceptiveness of the message irrespective of intent. That said, proving motive/intent in a design is also a very convincing argu- ment for conviction. There is a notable difference between unintentional bugs, errors, or bad metaphors and ones that have been carefully designed for specific purposes.

either by hiding truthful information or showing false in- formation. Deception is an act of deceit with implied intent (e.g., telling the user the web page is 70% loaded when we know that this is not the case). On the other hand, decep-tive(ness) does not require intent (e.g., telling the user that the web page is absolutely 70% loaded based on some esti- mate with high error margins). Though this distinction is important as it speaks to motive, deceit exists with or with- out intent. In fact, when deciding whether an advertisement is illegal (false), the FCC only considers the deceptiveness of the message irrespective of intent. That said, proving motive/intent in a design is also a very convincing argu- ment for conviction. There is a notable difference between unintentional bugs, errors, or bad metaphors and ones that have been carefully designed for specific purposes.

Deceit occurs when

- I. an explicit or implicit claim, omission of information, or system action,
- 2. mediated by user perception, attention, comprehension,
- prior knowledge, beliefs, or other cognitive activity,
- 3. creates a belief about a system or one of its attributes,
- 4. which is demonstrably false or unsubstantiated as true,
- 5. where it is likely that the belief will affect behavior,
- 6. of a substantial percentage of users.

Figure 1. A definition of deceit, based on [44].

It is further useful to distinguish between deceit that affects behavior directly or indirectly (by modifying the user's mental model). A test of this impact is whether a user would behave differently if they knew the truth. Though theline is fuzzy, this distinction allows us to separate abstrac- tion from deception. Applying this thought experiment to deception interfaces, one might see that the 1ESS user who was connected to the wrong number might begin to blame and mistrust the system, the rehabilitation user may recali- brate what is on the screen and again be unwilling to push themselves, and so on. Though the line is fuzzy, this dis- tinction allows us to separate abstraction (or simplification), in which user behavior is largely unchanged, from decep- tion, in which it often is changed.

Building on behavioral and legal definitions introduced in earlier work [44] that deal with deceptive advertising, we put forth a working definition of deceit as it applies to HCl work in the Figure 1. Points 5 & 6, on substantial effect, are perhaps the most controversial, and are purposefully left ambiguous. How behavior is affected and what "substan- tial" means are left open, as there is likely no answer that works in every situation. A deceptive interface that causes physical harm in 1% of the user population may have a sub- stantial effect, whereas an idempotent interface with a but- ton that misleads significantly more users into clicking twice may not pass the substantial test.

In addition to intent, there are many other ontologies of deceit. Bell and Whaley [4] identify two main types of de- ception—hiding and showing—which roughly correspond to masking characteristics of the truth or generating false information (both in the service of occluding the truth). These forms of deception represent atomic, abstract notions of deceit that we refine in our discussion below. Related to the hiding/ showing dichotomy is the split between silent (a deceptive omission) versus expressed deception. Lying, as a special class, is generally considered to be a verbal form of deception [5]. Because HCI need not involve a verbal ele- ment, we expand the notion of the "lie" to include non- verbal communication between humans and computers.

Users generally trust computer interfaces to accurately re-T system state. Reflecting that state dishonestly– through deception–is viewed negatively by users, rejected by designers, and largely ignored in HCI research. Many believe outright deception should not exist in good design. For example, many design guidelines assert: "Do not lie to your users" (e.g., [40, 45]) Misleading interfaces are usually attributed to bugs or poor design. However, in reality, de- ceit often occurs both in practice and in research. We con- tend that deception often helps rather than harms the user, a form we term benevolent deception. However, the over- loading of "deception" as entirely negative coupled with the lack of research on the topic, makes the application of de- ception as a design pattern problematic and ad hoc.



plored features that enable this type of deception. While research into human-to-human deception has advanced our understanding of deception in general, it has largely focused on communication behavior that pre-exists, and per- sists through, computermediated-communication.

Most of the systematic research in the academic HCI com- munity focuses on malevolent deception, or deception in- tended to benefit the system owner at the expense of the user [14]. Such research frames deception negatively, and focuses on detection and eradication of malicious or evil interfaces (e.g., dark-patterns [17]). Such patterns include ethically dubious techniques of using purposefully confus- ing language to encourage the addition of items to a shop- ping cart or hiding unsubscribe functionality. Many forms of malevolent deception, such as phishing and other fraud, is in clear violation of criminal law.

Deceptive practices that are considered harmful by legal organizations are, for generally good reasons, considered harmful by designers as well. A possible exception to the negative frame for malevolent deception is in contexts where an obvious adversary exists or for military or securi- ty purposes. For example, to detect network attacks or illicit use, security researchers may deploy "honeypots" to imitate vulnerable

plored features that enable this type of deception. While research into human-to-human deception hcceptable even when it might affect legitimate clients who try to connect even though no services are available.

It can be difficult to draw the line between malevolent and benevolent deception. We frame the distinction from the end-user's perspective: if the end-user would prefer an ex- perience based on the deceptive interface over the experi- ence based on the "honest" one, we consider the deception benevolent. Note that this includes situations in which both the system designer and the end-user benefit from the de- ception, something economists sometimes term a Pareto white lie [19]. Arguably, all benevolent deceptions that im- prove user experience benefit the interface creator as well; otherwise, we might use the term altruistic deception [19].

The limited HCl research that has been conducted on be- nevolent deception has focused on the use of magic [54], cinema and theater [32] as instructional metaphors for HCl design. By using deception for the purpose of entertain- ment, playfulness, and delight, it becomes acceptable. This work has connected these art forms, in which illusion, im- mersion, and the drive to shape reality dominate, to situa- tions in which there is willing suspension of disbelief on the part of the user, or at least willing pursuit of acceptable mental models. Similar lessons have been drawn from ar- chitecture. For example, theme parks and casinos [20, 34, 37] are designed specifically to utilize illusions that manip- ulate users' perception of reality, entertainment, and partic- ipation in the experience. In this paper, we will describe a number of deceits in HCl systems that have parallel designs to magic, theater, and architecture. However, not all benevolent deceit can be framed in terms of creating delight for

as advanced our understanding of deception in general, it has largely fo- cused on communication behavior that pre-exists, and per- sists through, computer-mediatedcommunication.