Coercion and deception in persuasive technologies

Timotheus Kampik
tkampik@cs.umu.se
(Corresponding author)

Juan Carlos Nieves
jcnieves@cs.umu.se

Helena Lindgren
helena@cs.umu.se

Department of Computing Science
Umeå University
Sweden

Abstract

Technologies that shape human behavior are of high societal relevance, both when considering their current impact and their future potential. In information systems research and in behavioral psychology, such technologies are typically referred to as **persuasive technologies**. Traditional definitions like the ones created by Fogg, and Harjumaa and Oinas-Kukkonen, respectively, limit the scope of persuasive technology to non-coercive, non-deceptive technologies that are explicitly designed for persuasion. In this paper we analyze existing technologies that blur the line between persuasion, deception, and coercion. Based on the insights of the analysis, we lay down an updated definition of persuasive technologies that includes coercive and deceptive forms of persuasion. Our definition also accounts for persuasive functionality that was not *designed* by the technology developers. We argue that this definition will help highlight ethical and societal challenges related to technologies that shape human behavior and encourage research that solves problems with technology-driven persuasion. Finally, we suggest multidisciplinary research that can help address the challenges our definition implies. The suggestions we provide range from empirical studies to multi-agent system theory.

1 Introduction

In scientific literature, the term *persuasive technologies* appeared first in 1998 in an abstract (special interest group announcement) by the behavioral psychologist Brian J. Fogg in the proceedings of the CHI Conference on Human Factors in Computing Systems. While the document does not explicitly provide a definition of persuasive technologies, it refers to them as "interactive technologies that change attitudes, beliefs, and behaviors" [Fog98]. Later, in his book *Persuasive Technology: Using Computers to Change What We Think and Do*, Fogg defines persuasive technology as an "interactive computing system designed to change people's attitudes or behaviors" [Fog03b, p.1]. In addition, Fogg defines persuasion as "an attempt to change attitudes or behaviors or both (without using coercion or deception)" [Fog03a, p.15]. While he acknowledges that use cases for persuasive technologies can be unethical, Fogg puts the focus explicitly on "positive, ethical

20th International Workshop on Trust in Agent Societies, Stockholm, Sweden; July 14, 2018

applications of persuasive technology" [Fog03b, p.6]. In a later definition, the information systems researchers Oinas-Kukkonen and Harjumaa combine Fogg's definitions of *persuasion* in general and *persuasive technology* in particular by stating that a persuasive technology is a "computerized software or information system designed to reinforce, change or shape attitudes or behaviours or both without using coercion or deception" [OKH08b], thus setting a similar scope as Fogg. However, the persuasive influence of technology on human society is often perceived negatively, both by the scientific community and the general public, *because* it is associated with coercion and deception. Studies of past events suggest that spreading misinformation via social media can have lasting effects on the perception consumers have of the topic in focus [Ber17]. Political persuasion via social media is under increasing public scrutiny. For example, in April 2018 Facebook CEO Mark Zuckerberg was questioned by the Senate of the United States of America about the impact his company's products had on the 2016 presidential election [Sen18]. From the product design and software development community, the "time well spent" initiative emerged, calling for a more user-focused design of consumer applications that affect human behavior. The initiative argues that many of the most popular Internet applications persuade users to spend as much time as possible online to maximize the time users are exposed to (potentially deceptive) advertisements[1]. Scientific literature has established that while many persuasive technologies persuade users to engage in health-supporting [CSP+17] and educational [DKdT+17] activities, some online and mobile games make use of the same concepts for ethically questionable purposes, for example, to persuade the user to purchase virtual items [KT17]. Theoretical work, for example, the paper "Toward an Ethics of Persuasive Technology" by Daniel Berdichevsky and Erik Neuenschwander [BN99], reflects these concerns about persuasive technologies. By stating that "[p]ersuaders often tweak the truth or ignore it entirely", the authors imply that deception is a common feature of persuasive technology. Castelfranchi argues that the concept of deception is at the root of artificial intelligence research by pointing out that "remembering the so called *Turing's test* we could say that deception is the original, foundational challenge and proof for AI" [Cas00]. De Rosis et al. criticize the *severity assumption* that is often made in human-computer interaction and multi-agent system research and development and state that in the context of information technology, *deception occurs not only occasionally and unintentionally, but also on purpose* [FVGC].

In this paper, we analyze persuasive aspects of technologies in regards to conflicts of interests, deception, and coercion. From the findings of our analysis and from the background research we derive an updated definition of *persuasive technologies*. Finally, we discuss the implications the updated definition can have on future research.

1.1 Analysis approach

The overview of the status quo as established in the background section suggests that the line between persuasive, deceptive, and coercive technologies is blurred. To find confirming or refuting evidence for this hypothesis, we picked ten popular existing technologies that have prominent persuasive properties and analyzed them in regards to the following questions:

- What are the most important persuasive properties of the system? In this context, we consider the following sub-questions:
  - Who is persuader, who is persuadee[2], and how can the relationship between persuader and persuadee be described (*persuasion relationship*)?
  - What does the persuader want to achieve (*persuasion goals*)?
  - What does the persuadee want to achieve (*user goals*)?
  - How does the persuader achieve their goals (*persuasion strategy*)?
- Are there conflicts of interests between persuader and persuadee?
- To what extent do the persuasive properties of the system embrace coercion and deception?

[1] "Time well spent" is driven by the "Center for Humane Technology", an organization that wants to move "away from technology that extracts attention and erodes society, towards technology that protects our minds and replenishes society" [Cen18].

[2] In this paper we refer to persuaded individuals as *persuadees*.

For the analysis, we selected systems considering the following criteria:

- We considered a broad range of application types: learning and fitness applications, games, social networks, enterprise instant messaging systems, news applications, and software development tools to cover several domains with our analysis. This allows for a more confident generalization.
- The selection contains applications that are persuasive from a common sense perspective, as well as applications that do not seem to be persuasive at first glance. This helps to extend the analysis to subtler persuasion scenarios.
- The applications can be considered as relatively well-known by and available to a wide range of users. This improves the understandability of the analysis and helps ensure it is of societal relevance.

We rely on the following definitions of *coercion* and *deception* from the Stanford Encyclopedia of Philosophy:

- Coercion: "use of [...] power for the purpose of gaining advantages over others [...], punishing non-compliance with demands, and imposing one's will on the will of other agents." [And17][3]
- Deception: "to intentionally cause to have a false belief that is known or believed to be false." [Mah16][4]

Based on our analysis, as well as on the review of existing literature, we derived an updated definition of persuasive technologies that considers the existing definitions of Fogg, and Harjumaa and Oinas-Kukkonen, but also takes into account the properties of existing information systems.

2 Coercion and deception in existing persuasive technologies

We included the following applications in our analysis:

- **Mobile game** (Candy Crush[5]),
- **Language learning application** (Duolingo[6]),
- **Social network** (Facebook[7]),
- **Source code sharing site** (GitHub[8]),
- **Social network for researchers** (ResearchGate[9]),
- **Workout support and analysis application** (Runtastic GPS Running App[10]),
- **Enterprise instant messaging application** (Slack[11]),
- **Knowledge sharing website** (Stack Exchange[12]),
- **Online newspaper** (Washington Post, online edition[13]),
- **Video sharing platform** (YouTube[14]).

[3]For a more technical perspective, one can apply Nozick's definition of coercion as a claim that agent P makes to agent Q if "P's claim indicates that if Q performs [an activity] A, then P will bring about some consequence that would make Q's A-ing less desirable to Q than Q's not A-ing" [Noz69].

[4]As this definition requires *intent*, it is narrower than definitions used in other work on deceptive information systems. E.g., Castelfranchi's paper on deceptive multi-agent systems covers both *intentional deception* and *deception by ignorance* [Cas00]. Also, we do not distinguish between *deception* and *lie* (see: *simple deceptionism* in [Mah16]). For an investigation of the role of deception in multi-agent worlds, see [FCdR01].

[5]`https://candycrushsaga.com`

[6]`https://www.duolingo.com/`

[7]`https://www.facebook.com/`

[8]`https://github.com/`

[9]`https://www.researchgate.net/`

[10]`https://www.runtastic.com/en/apps/runtastic`

[11]`https://slack.com/`

[12]`https://stackexchange.com/`

[13]`https://www.washingtonpost.com/`

[14]`https://www.youtube.com/`

The appendix(6.1) contains a structured analysis of each application in regards to each question. We could identify persuasive properties in all of the selected applications. In six of the applications (Facebook, ResearchGate, StackOverflow, YouTube; cost-free versions of Duolingo, Runtastic), at least one of the persuasion goals was engaging with advertisements a third-party (a customer) provides, whereas three of the applications strive to persuade users to directly make a purchase decision (Candy Crush, GitHub, Slack). In five of the applications (Duolingo, GitHub, Runtastic GPS Running App, ResearchGate, StackOverflow) persuasive features help users to work more effectively towards their own goals. Three applications (Facebook, GitHub, Slack) could facilitate deceitful persuasion between end-users. GitHub and Slack allow users to deceitfully persuade others to view them as exceptionally productive or hard-working by intentionally optimizing indicators like online status and contribution counts, without actually producing impactful contributions. In the case of Facebook, existing literature confirms the application has effectively been employed for this purpose [Ber17]. As a persuasion strategy, seven applications (Duolingo, Facebook, GitHub, Runtastic GPS Running App, ResearchGate, Stack Exchange, YouTube) provide social incentives, while five (Duolingo, GitHub, Runtastic GPS Running App, ResearchGate, Stack Exchange) provide gamification features. Two applications (Slack, Candy Crush) coerce the user by requiring a payment to access user-generated content or a path to game victory, respectively. One application (Washington Post), occasionally uses deceptive (*forward-referring* [BH15]) article headlines[15] to persuade the reader to consume mundane articles[16]. Two applications (Facebook, Slack) enable end-users to deceitfully share content with other end-users to persuade them to adopt a particular opinion about the persuader or a more general topic. The six applications that provide targeted advertisements allow the advertisement customer to define an additional persuasion strategy. In these applications, there is a hierarchical relationship between the advertisement customer and the persuadee (end-user), in which the customer controls to some extent how the end-user is persuaded. The persuadee needs to accept (some of the) persuasive features of the application as a side effect of using other application features. The same is the case in the two coercive scenarios (persuasion to upgrade in Slack, persuasion to pay for virtual lives in Candy Crush); the persuader deliberately disempowers the persuadee to reach the persuasion goal. In the two user-to-user deception scenarios (deceiving other end-users in Slack and Facebook), the relationship between persuader and persuadee could be considered as coequal, because both persuader and persuadee have the same means and rights in the system. However, we assess the relationship as de facto hierarchical, as only the persuader is aware of the deceptive aspects of the interaction and limits–in the case of using Facebook to spread misinformation–their own usage largely to persuasive purposes. A similar deceptive relationship exists (albeit to a lesser extent) in the Washington Post's persuasion scenario, if the user is persuaded by forward-referring (deceptive) headlines. In the five applications that help users to reach their own goals by assisting with self-persuasion, the relationship is to some extent coequal (depending on possible additional persuasion goals), as the users apply the persuasive features as tools to reach personal goals and are aware of the persuasive intent of the application. In all applications, potential conflicts between persuader and persuader are inherent, albeit of different intensity. The conflicts are arguably most evident in scenarios where persuaders and persuadees have irreconcilable goals regarding how the persuadees should spend their time (targeted advertising) or what information should be provided to them (user-to-user deception, coercion/deception to trigger a subscription decision). Conflicts of interests are less intense in scenarios where the persuasive features are at least to some extent used knowingly by the users for self-persuasion to reach personal goals. For example, in the free version of the Duolingo language learning application, gamification features persuade users to absolve learning units, but also make them watch brief advertisements at the end of each unit. While in only some of the persuasion scenarios the key persuasion strategies are coercive or deceptive, all of the applications we analyzed have some persuasive properties that are at least *to some extent* deceptive or coercive. For example, when using the free version of the Runtastic GPS Running App, the user is persuaded to spend at least some of their time consuming advertisements, which could be considered as deceptive (giving the user the feeling the app provides an optimized workout experience, although the business goals of the application provider is increasing the time the user spends consuming advertisements). Also, users of an online social network might feel coerced into using the application because they are afraid of social exclusion. The intensity of persuasive properties and their deceptive or coercive aspects differ from application to application. The most common conflict of interests is between the end-users and the paying customers, who want to persuade the end-users to view advertisement to ultimately buy a product or service. Especially in the case of Facebook,

[15]For example: "In the age of tech giants, this company shows it's still possible to defy them", see: `https://www.washingtonpost.com/news/the-switch/wp/2018/04/06/in-the-age-of-tech-giants-this-company-shows-its-still-possible-to-defy-them/`

[16]Note that we do not consider the Washington Post's paywall as *coercive*.

the line between persuasion, deception, and coercion is blurred: the fear of missing out on important social communications and events might coerce users into spending a lot of time in the application, where they are targeted by advertisements, and possibly also by political agitators with deceptive intentions.

3 Persuasive technology: an updated definition

Based on the findings of the background research and technology analysis, we created the following new definition of persuasive technology:

We define persuasive technology as an information system that proactively affects human behavior, in or against the interests of its users.

In the context of the definition, we define the following core requirements technologies need to fulfill to be considered persuasive:

- **Intentionally persuasive**: The system needs to have a *persuasive intent*, that was designed by the system developers, steered by a user, or initiated autonomously by the system itself. For example, an application that informs users about the weather conditions and thus causes them to stay inside is not persuasive, because there is no persuading party that has an interest in influencing the users.
- **Behavior-affecting**: The system needs to *effectively* impact human behavior. For example, an application that is designed to encourage users to work out by sending motivational messages, but fails to affect user behavior–either due to software errors or poor design choices–does not qualify as persuasive. Note that we chose to define persuasive technology as *behavior-affecting* and not *behavior-changing* to cover technology that persuades humans to *sustain* behavior, for example to maintain healthy routines (see: [Bou14]).
- **Technology-enabled**: Technology needs to play a crucial role in the persuasion process that cannot be handled by non-technological means. For example, if a sales representative uses an information system to personally present a product or a service to a prospective customer via the Internet, the persuasion process is not significantly enhanced by the technology; hence the technology does not qualify as persuasive.
- **Proactive**: The system needs to *proactively* persuade the targeted users. This requires at least *some extent of autonomy*[17] or a *high degree of automation*. For example, a programmable alarm clock that can persuade its user to get out of bed in the early morning is not sufficiently autonomous or automated to qualify as persuasive technology.

The definition differs from the previous definitions by Fogg, and Harjumaa and Oinas Harjumaa in the following key aspects:

- It explicitly states that persuasive technology can act *against the interests of its users*. Thus, the definition covers deceptive and coercive systems. (Hypothetically a user could also *want* a system to persuade them to act against their own interest, but we do not consider this scenario practically important.)
- It does not require a computing system to be *designed for* persuasive purposes. We consider this distinction important because the practical purpose of computing systems can evolve in ways the system designers do not control.

4 Discussion

4.1 Limitations

The analyses of the persuasive properties of wide-spread technology provided in this paper are preliminary as they are based on somewhat subjective observations of the corresponding applications, which are only in some cases supported by empirical studies. While such studies are necessary to provide an accurate assessment of the detailed persuasive intentions and effects of each individual system, we are confident our

[17] While discussing definitions of *autonomy* is beyond the scope of this work, one can apply the definition of Wooldridge, who defines agents as autonomous if "they have control both over their own internal state and over their behavior" [Woo13]. It also makes sense to think of autonomy in terms of reasoning capabilities that inform the systems goals (see: "reasons-responsive conception of autonomous agency" in [BW18]).

conceptual analysis is sufficient to support the position that persuasive technologies practically often convey conflicts of interests and do not allow for a clear separation of persuasive from deceptive and coercive aspects. When expanding the scope from persuasive technology to persuasion in general, our argument is in addition supported by scientific history; already Aristotelian rhetoric discusses deceptive intents of persuasion (see: [RW85]).

4.2 Alternative approaches

An alternative approach to addressing coercion and deception in the definition of persuasive technologies can be coining an additional term–for example: *manipulative technology*–to encompass only coercive and deceptive persuasive technologies. This would allow for a clear distinction between persuasion *in* and *against* the interest of the user. However, such a term would encourage labeling specific systems with *either* of the terms, which would facilitate the misconception that a clear distinction is generally possible.

An insightful change of perspectives can be achieved when approaching the problem in focus from a somewhat technology-agnostic behavioral economics view. Thaler et al. introduce the concept of *choice architecture*, which can be summarized as designing objects in a way that optimizes the behavior of people interacting with those objects towards a specific goal [BPT07]. They then introduce the notion of *libertarian paternalism* as an economic maxim that allows for maximal choice (*libertarian*) but designs the choice architecture in a way that maximizes the welfare of all involved parties (*paternalistic*) [TS03]. One can assume that the concept of *libertarian paternalism* comes with the same pitfalls as the concept of *persuasive technology*; both rely on a (*choice architects* and *persuaders*, respectively), who are–at least in the moment of persuasion–to some extent controlling the fate of their subjects. In neither case, benevolence can be taken for granted.

4.3 Societal relevance

Fogg's definition of persuasive technology contains a *design* requirement; for him, persuasive technology requires that persuasive features have been devised intentionally by the technology's designers [Fog03a, p.17]. When he argues for this restriction, Fogg limits his examples of out-of-scope persuasion scenarios entirely on users who employ technology for *self-persuasion* purposes the designers have likely not intended. However, we can show by example–end-users employing a social network for deceptive persuasion–that for a technology to be persuasive, the intent to persuade does not need to originate from the designer, but could alternatively originate from a *third-party* user. One could even argue that highly autonomous systems can develop the intention to persuade without any human intent by evolving in a way the system designer did not plan.

While some of the identified scenarios seem to be trivial in their simplicity, the prevalence in many consumer and business applications suggests that generally, deceptive and coercive persuasion is used as a tool by technology designers and users, be it deliberately or not. Taking into account recent developments regarding the adoption of advanced data analytics methods in the software industry, one can expect that in the future persuasive features will be implemented with an increasing degree of personalization, which is likely to further disempower end-users. One can assume that in the future, persuasive technologies will increasingly blur the line between persuasion *in* and *against* the interest of the user. For example, organizations or states could require their employees or citizens to make use of persuasive technologies that increase productivity, decrease healthcare costs, or facilitate a specific ideology. Consequently, we consider a holistic definition of persuasive technologies that takes into account the possibility of deceptive and coercive persuasion as important to frame future research, for example when devising frameworks and algorithms for resolving conflicts of interests between humans and intelligent agents.

4.4 Implications for persuasive technology research

Given the research community agrees with our position that addressing coercion and deception in persuasive technologies is of societal relevance, our updated definition of *persuasive technology* could have the following implications for future research:

- **Empirical research on persuasive technologies**
  Empirical research could study the effects of deceptive and coercive use of persuasive technology to confirm or refute our position, both by analyzing data from everyday technology usage and by conducting trials in clinical settings.

- **Research-driven design and development of persuasive technologies**
  Design frameworks, as well as specific systems that mitigate the coercive and deceptive effects of persuasive technology could be developed. As a starting point, one could extend Harjumaa's and Oinas-Kukkonen's *framework for designing and evaluating persuasive systems* [OKH08a] to better account for deception and coercion. In practice, systems could be designed to comply with precisely defined and technically enforceable ethics rules and infer users' emotional states from sensor data and system logs to *avoid* exploiting vulnerable individuals.

- **Theory of multi-agent systems**
  New approaches in multi-agent systems theory could be devised to facilitate the development of the socially intelligent systems we described in the point above. As the basis for this, one could use argumentation-theoretic reasoning approaches [BDKT97]. In particular, one could apply Bench-Capon's extension of value-based argumentation frameworks for solving persuasive arguments to develop concepts of socially intelligent agents that consider social norms and ethics rules when persuading other agents (especially humans). An alternative approach for formalizing persuasive, deceptive, and coercive interactions between agents could be utility theory. For example, one could apply and extend the logic of utilitarian desires as developed by Lang et al. [LvdTW02]. Although multi-agent theory provides concepts that address a wide range of practical problems, it is often not applied in widely used systems. To facilitate the adoption of theory in applied research and software development, we recommend the development of higher-level abstractions of formal definitions that are more appealing to software engineers and ideally understandable for the general public. Such abstractions could possibly be provided by a graphical notation derived from the implicational logic for argument-centric persuasion as proposed by Hunter [Hun14].

5 Conclusion

While previous definitions of the concept of persuasive technology attempt to draw a clear line between persuasion and deception/coercion, such a clear separation does commonly not exist in actual information systems. Potential conflicts of interests between the initiator and the target of persuasion exist in popular consumer and enterprise IT applications. In this respect, persuasive technologies do not differ from other forms of persuasion, i.e. persuasive person-to-person communication. Moreover, we find that the well-established definitions' requirement of persuasion as a system feature *by design* does not always match the complexity and longevity of modern socio-technical systems that continuously evolve, both in respect to their technical properties and the role they play in society. Our updated definition of persuasive technologies as *information systems that proactively affect human behavior, in or against the interests of its users* reflects these findings. For future research, we suggest reviews and extensions of concepts for describing, evaluating and designing persuasive technologies, as well as more empirical research on how to manage conflicts of interests in human-computer interaction, in especially in the interaction of humans with increasingly autonomous agents.

Acknowledgements

The authors thank the anonymous peer reviewers for constructive criticism of the manuscript. This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

6 Appendix

6.1 Coercion, deception, and conflicts of interests in persuasive technologies: overview

Candy Crush
Persuasive properties: It is in the economic interest of the Candy Crush designers to persuade the game's users to spend as much money as possible on in-app goods. **Persuasion goals:** The game provider wants to persuade players to spend money to purchases virtual goods. **User goals:** The goal of the user is entertainment (having fun). **Persuasion relationship:** The relationship between persuader (application provider) and persuadee (user) is one-to-many and hierarchical, as the user cannot control the persuasive game mechanics and is not necessarily aware of the persuasion strategy that tries to exploit their short-term desire to win, which is not compliant with their long-term economic goals. **Persuasion strategy:** The application provider

persuades users to spend money by providing an additional path to victory that is only available if the users pay. **Conflicts of interests:** Users play for fun, probably often assuming they do not need to pay to play. The persuasion goal is against the economic interest of the users. **Coercive aspects:** The application approaches the users in situations of particular weakness and persuades them to pay. The short-sighted desire of users to win is intentionally build up and is in sharp contrast to the arguably more important long-term objective of fiscal responsibility. **Deceptive aspects:** -

Duolingo
Persuasive properties: Language learners use Duolingo to persuade themselves to spend more time learning languages. In contrast, it is in the economic interest of the application vendor to persuade the users to either spend as much time as possible with advertisements or to subscribe to the paid version. **Persuasion goals:** The persuasion goals are: 1. to spend time on advertisements (application provider, free-of-cost version only); 2. spend time learning languages (user). **User goals:** The goal of the user is to improve their skills in a particular language (persuasion goal *2.* is the user goal). **Persuasion relationship:** Persuasion goal *1.*: the relationship between persuader and persuadee is hierarchical and one-to-many. Persuasion goal *2.*: the persuasion relationship is one-to-self (self-persuasion) and the persuader is simply providing tools that help the persuadee to achieve their goals. **Persuasion strategy:** The persuasion strategy for both persuasion goals is to provide social incentives and gamification elements (competitions and sharing of achievements) to increase the time spent in the application (time spent on advertising and time spent learning, respectively). **Conflicts of interests:** Language learners use Duolingo to persuade themselves to spend more time learning languages. In contrast, it is in the economic interest of the application vendor to persuade the users to either spend as much time as possible with advertisements or to subscribe to the paid version. **Coercive aspects:** - **Deceptive aspects:** While Duolingo might not be deceptive per se, it is questionable towards whose objective its gamification features are optimized: towards the learning effect, towards time spent with advertisements, or towards to the users' desire to use the app continuously (and pay for it).

Facebook
Persuasive properties: Facebook and its customers want to persuade end-users to spent as much time as possible using the application and to watch (click on) ads as often as possible. In addition, some end-users try to persuade (or deceive) others to accept their version of reality, for example to change political opinions, or simply the perception others have of one's individual success in life. **Persuasion goals:** The persuasion goals are: 1. to spend time on advertisements and to spend time curating the user profile (that allows for better targeted advertisements), 2. to adopt a specific view on a person or topic. **User goals:** The goal of the end-user is to stay in touch with friends and family, and possibly to receive news updates from their favorite celebrities and organizations. **Persuasion relationship:** Persuasion goal *1.*: the persuasion relationship is hierarchical and one-to-many. Persuasion goal *2.*: the relationship appears as leveled and many-to-many. However, as the persuader uses the application to broadcast persuasive content, the relationship is practically hierarchical and one-to-many. **Persuasion strategy:** Persuasion goal *1.*: the persuasion strategy is to provide social incentives for consuming advertisements and for providing commercially exploitable data. Persuasion goal *2.*: the persuasion strategy is to spread inaccurate or false information that not all users treat with sufficient scrutiny (as it is just one of many news items they see when scrolling through their feed). **Conflicts of interests:** It is in the interest of the end-users to communicate efficiently and meaningfully with their social contacts. In contrast, Facebook's customers want to draw as much of the end-users' attention as possible towards advertisements and persuade them to purchase products and services. Moreover, other end-users might use the application for deceptive or coercive purposes. **Coercive aspects:** As Facebook is an important tool of staying in touch with social contacts, some users might feel coerced to use the application (and to be subjected to advertisements) to not miss something important or to appear as unsuccessful. **Deceptive aspects:** Facebook facilitates deception of end-users by its advertisement customers, as well as by other end users.

GitHub
Persuasive properties: Social features and gamification elements persuade users to stay engaged with the application. **Persuasion goals:** The goal of the application provider is to create a community of active users, who will eventually recommend GitHub's *private* repositories (paid subscription) to their employer. **User goals:** Individual users typically start using the application to create hobby and student projects, or to con-

tribute to open source software. Important additional goals are possibly improving one's personal brand or recognition within the software development community. **Persuasion relationship:** The relationship between persuader and persuadees is one-to-many, but cannot be considered hierarchical, when taking into account that users are likely to enjoy the persuasive features as tools that foster their motivation to contribute to software projects. **Persuasion strategy:** GitHub makes use of social features as well as gamification elements (contribution "scores") to persuade developers to keep using the application. **Conflicts of interests:** There is no clear conflict of interests between persuader and persuadee. **Coercive aspects:** - **Deceptive aspects:** While GitHub does not deceive users directly, the focus on very simple contribution metrics facilitates deception between users, as it encourages optimizing one's one behavior towards these metrics and does not consider the quality of the contributions.

ResearchGate
Persuasive properties: ResearchGate uses social features, as well as gamification elements like scores/performance metrics to persuade users to stay engaged. **Persuasion goals:** The goal of the persuader (application provider) is to persuade users to consume advertisements that are displayed in the application. **User goals:** The goal of the user is to increase their visibility in the academic community and to stay up-to-date on latest research findings. **Persuasion relationship:** The relationship between persuader and persuadee is one-to-many and hierarchical, as the persuadee is exposed to the persuasive features without being necessarily aware of it. **Persuasion strategy:** ResearchGate's social features, as well as its extensive use of email notifications aims to keep the users engaged and tries to persuade them to provide information that can be used for targeting advertisements. **Conflicts of interests:** Although one can assume that researchers have an interest using a platform that facilitates communication within the research community, there is a conflict of interests between persuader and persuadee, as the application provider needs to maximize exposure to advertisements, while it is in the interests of the users to spend limited time on ResearchGate to be able to dedicate more time to actual research work. **Coercive aspects:** - **Deceptive aspects:** Although the persuasive features are not generally deceptive, ResearchGate sends deceptive emails to facilitate engagement and data collection. For example, users are informed about new profile views by an automatic email that contains a "Find out more"-link to a page that asks users to complete their profile information. The email subject states "...your profile is getting attention", which is misleading, as it might be overstatement (it is sent even if there is only one additional profile view) and as it does not reflect the purpose of the email (persuasion to complete profile information).

Runtastic GPS Running App
Persuasive properties: Users of the Runtastic GPS Running App persuade themselves to work out more. Similar to Duolingo, it is in the application provider's economic interest to either spend as much time as possible with advertisements or to subscribe to the paid version. **Persuasion goals:** The persuasion goals are: 1. to spend time on advertisements (only free version), 2. spend time working out (user). **User goals:** The goal of the user is to work out more and to get a better overview of their workout metrics (congruent with persuasion goal *2.*). **Persuasion relationship:** Persuasion goal *1.*: persuader and persuadee have a hierarchical one-to-many relationship. Persuasion goal *2.*: the persuader is providing tools that help the persuadee to achieve their goal (self-persuasion). **Persuasion strategy:** The persuasion strategy is to provide social incentives and a gamification experience to encourage the user to spend more time in the application and to work out more, respectively. **Conflicts of interests:** Users of the cost-free version are exposed to in-app advertisements. While it is in the interest of the users to spend more time working out, the interest of the application vendor is to maximize the time the user spends exposed to advertisements, which is typically not during workout periods, but rather a distraction when users are preparing or analyzing workouts. **Coercive aspects:** - **Deceptive aspects:** The application might reward users to some extent for the interaction with the application (that helps creating advertisement revenue) and not only for the actual workout.

Slack
Persuasive properties: Slack's cost-free version has limited features (i.e. limited message history), which have little effect on usability right after an organization starts using the software. After the cost-free version has been successfully adopted, however, withholding the message history from the users effectively coerces the organization to switch to a paid subscription. Moreover, employers hope Slack persuades employees to facilitate transparent communications. In contrast, employees might use the application to persuade others to

view them as hard-working. **Persuasion goals:** The persuasion goals are: 1. to make organizations pay for a subscription (application provider-to-user persuasion), 2. to make other users adopt a specific view on the value an employee or colleague provides to the organization (user-to-user persuasion). **User goals:** The goal of the user is to communicate with their colleagues more effectively and efficiently. In addition (in the case of the the user-to-user persuasion scenario), the user goal is congruent with persuasion goal *2.*. **Persuasion relationship:** Persuasion goal *1.*: the relationship between persuader and persuadee is hierarchical and one-to-many. Persuasion goal *2.*: the relationship appears to be leveled and many-to-many, but is practically hierarchical and one-to-many, because only the deceiving user (or the few deceiving users) broadcasts their persuasive message and is the only one who is aware of the deceitful persuasion). **Persuasion strategy:** Persuasion goal *1.*: withhold user-generated information to encourage a paid subscription. Persuasion goal *2.*: adjust own usage of the application to resemble usage patterns that others associate with high-performing employees. **Conflicts of interests:** The application provider has the interest to establish Slack as the most important communication and collaboration system, as it helps secure long-term revenue. For this, the provider uses access limitations to user-created content (the message history) as a tool to persuade organizations to subscribe to the paid Slack version. In contrast, it is in the interest of organizations that use Slack to have access to their communication history at any time. Also, it is in the interest of Slack users to employ the tool to persuade colleagues to consider them as hard-working, while it is in the interest of the organization that employs Slack that users communicate efficiently and honestly. **Coercive aspects:** Users are coerced to pay for a subscription to access content (parts of the message history) they created themselves. **Deceptive aspects:** Users might feel encouraged to deceive others by making Slack present them as hard-working (answering messages late at night, spending more time on Slack than on value-creating work).

Stack Exchange
Persuasive properties: Users of Stack Exchange use the application's reputation and badge system to persuade themselves to learn and share knowledge with others. It is in the interest of the application provider that users see and click on job advertisements. **Persuasion goals:** The persuasion goals are: 1. to make users spend time exposed to advertisements/spend time curating their profile that allows for better targeted advertisements (application provider), 2. to persuade oneself to learn and help others, or to facilitate one's own professional reputation (users). **User goals:** The goal of the user is to receive helpful answers to their subject matter-specific questions, or to enhance and show off their expertise by answering the question of others (congruent with persuasion goal *2.*). **Persuasion relationship:** Persuasion goal *1.*: the relationship between persuader and persuadee is hierarchical and one-to-many. Persuasion goal *2.*: users persuade themselves (self-persuasion) using tools the application vendor provides. **Persuasion strategy:** The persuasion strategy is to provide social and gamification incentives for time spent exposed to advertisements, or for learning and strengthening one's professional profile, respectively. **Conflicts of interests:** The main economic interest of the application provider is to engage users as much as possible and to use the information users provide to display highly targeted advertisements; in contrast, most users want to find answers to their questions quickly or to acquire expert knowledge about a particular field by researching details that help answer questions. **Coercive aspects:** - **Deceptive aspects:** The gamification system of Stack Exchange encourages users to all kinds of participation, from asking and answering questions to reviewing, policing and copy-editing. The reward system provides users with a feeling of productivity, even when it distracts users from doing possibly more important value-creating work.

Washington Post
Persuasive properties: The online edition of the Washington Post employs persuasive features to facilitate paid subscriptions. **Persuasion goals:** The persuasion goal is to make the user pay for a Washington Post subscription. **User goals:** The core user goal is to stay informed and to only pay for a subscription if necessary. **Persuasion relationship:** The relationship between persuader and persuadees is one-to-many and hierarchical, as the persuader is not necessarily transparent about the tools of persuasion (see "Persuasion strategy" below). **Persuasion strategy:** Users can access three articles per month free of charge. Any additional article that is accessed by a user (or technically: by a client the application considers a unique user), is obscured by an overlay that asks the user to subscribe to get full access. Some of the article headlines[18]

[18]For example: "In the age of tech giants, this company shows it's still possible to defy them", see: `https://www.washingtonpost.com/news/the-switch/wp/2018/04/06/in-the-age-of-tech-giants-this-company-shows-its-still-possible-to-defy-them/`

use *forward-referring* headlines[BH15] to persuade the reader to consume mundane articles. **Conflicts of interests:** In scenarios where the user is not aware that a news article does not contain substantial additional information, there is a conflict of interests between the user, who wants to pay for only for high-quality journalism and the provider, who wants (and persuades) the user to pay in any case. **Coercive aspects:** - **Deceptive aspects:** The persuasion strategy is to some extent deceptive, at least if the user is persuaded by headlines that use forward-references to lure the reader into reading further (see "Persuasion strategy" above).

YouTube
Persuasive properties: YouTube's recommendation system is designed to persuade end-users to stay engaged using the application. The application's advertisement system then tries to persuade the user to consume third-party products. **Persuasion goals:** The main persuasion goal is to keep the user exposed to advertisements as long as possible and to consume advertisements. **User goals:** Likely user goals are information and entertainment. However, one can assume that most users' need for being informed and entertained by the application is limited. **Persuasion relationship:** The relationship between persuader and persuadee is one-to-many and hierarchical. Users of the application do not opt into the persuasion and are not necessarily aware of the persuasion and have only limited ability to influence the persuasion process (not using the application, installing a web browser plugin to block advertisements). **Persuasion strategy:** The application tries to persuade users to keep using the application by providing a constant stream of easy-to-consume videos that are selected based on the video consumption history of the individual end-user, and by continuously suggesting alternative videos the user is likely to watch. **Conflicts of interests:** Conflicts of interests are likely to arise because one can assume that it is often against the interest of the user, but always in the interest of the application provider that the user continues watching. **Coercive aspects:** One could argue the persuasion is to some degree coercive, as the persuader exploits the temporary inability of the persuadee to act in accordance with their long-term goals. **Deceptive aspects:** Advertisements can be to some degree deceptive.

References

[And17] Scott Anderson. Coercion. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2017 edition, 2017.

[BDKT97] Andrei Bondarenko, Phan Minh Dung, Robert A Kowalski, and Francesca Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial intelligence*, 93(1-2):63–101, 1997.

[Ber17] Adam J Berinsky. Rumors and health care reform: experiments in political misinformation. *British Journal of Political Science*, 47(2):241–262, 2017.

[BH15] Jonas Nygaard Blom and Kenneth Reinecke Hansen. Click bait: Forward-reference as lure in online news headlines. *Journal of Pragmatics*, 76:87–100, 2015.

[BN99] Daniel Berdichevsky and Erik Neuenschwander. Toward an ethics of persuasive technology. *Communications of the ACM*, 42(5):51–58, 1999.

[Bou14] Mark E Bouton. Why Behavior Change is Difficult to Sustain. *Preventive medicine*, 0:29–36, November 2014.

[BPT07] Shlomo Benartzi, Ehud Peleg, and Richard H Thaler. Choice architecture and retirement saving plans. *The behavioral foundations of public policy*, pages 245–263, 2007.

[BW18] Sarah Buss and Andrea Westlund. Personal autonomy. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2018 edition, 2018.

[Cas00] Cristiano Castelfranchi. Artificial liars: Why computers will (necessarily) deceive us and each other. *Ethics and Information Technology*, 2(2):113–119, Jun 2000.

[Cen18] Center for Humane Technology. The problem, 2018.

[CSP+17] Yang-Wai Chow, Willy Susilo, James G Phillips, Joonsang Baek, and Elena Vlahu-Gjorgievska. Video games and virtual reality as persuasive technologies for health care: An overview. 2017.

[DKdT+17] Sam Devincenzi, Viviani Kwecko, Fernando Pereira de Toledo, Fernanda Pinto Mota, Jonas Casarin, and Silvia Silva da Costa Botelho. Persuasive technology: Applications in education. In *2017 IEEE Frontiers in Education Conference (FIE)*, pages 1–7. IEEE, 2017.

[FCdR01] Rino Falcone, Cristiano Castelfranchi, and Fiorella de Rosis. Deceiving in golem: how to strategically pilfer help. In *Trust and deception in virtual societies*, pages 91–109. Springer, 2001.

[Fog98] Brian J Fogg. Persuasive computers: perspectives and research directions. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 225–232. ACM Press/Addison-Wesley Publishing Co., 1998.

[Fog03a] Brian J. Fogg. Chapter 1 - overview of captology. In Brian J. Fogg, editor, *Persuasive Technology*, Interactive Technologies, pages 15 – 22. Morgan Kaufmann, San Francisco, 2003.

[Fog03b] Brian J. Fogg. Introduction. In Brian J. Fogg, editor, *Persuasive Technology*, Interactive Technologies, pages 1 – 13. Morgan Kaufmann, San Francisco, 2003.

[FVGC] De Rosis Fiorella, Carofiglio Valeria, Grassano Giuseppe, and Castelfranchi Cristiano. Can computers deliberately deceive? a simulation tool and its application to turing's imitation game. *Computational Intelligence*, 19(3):235–263.

[Hun14] Anthony Hunter. Opportunities for argument-centric persuasion in behaviour change. In Eduardo Fermé and João Leite, editors, *Logics in Artificial Intelligence*, pages 48–61, Cham, 2014. Springer International Publishing.

[KT17] Marco Josef Koeder and Ema Tanaka. Game of chance elements in free-to-play mobile games. a freemium business model monetization tool in need of self-regulation? 2017.

[LvdTW02] Jérôme Lang, Leendert van der Torre, and Emil Weydert. Utilitarian desires. *Autonomous Agents and Multi-Agent Systems*, 5(3):329–363, Sep 2002.

[Mah16] James Edwin Mahon. The definition of lying and deception. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2016 edition, 2016.

[Noz69] Robert Nozick. Coercion. In M. P. S. Suppes White Morgenbesser, editor, *Philosophy, Science, and Method: Essays in Honor of Ernest Nagel*, pages 440–72. St Martin's Press, 1969.

[OKH08a] Harri Oinas-Kukkonen and Marja Harjumaa. A systematic framework for designing and evaluating persuasive systems. In *International conference on persuasive technology*, pages 164–176. Springer, 2008.

[OKH08b] Harri Oinas-Kukkonen and Marja Harjumaa. Towards deeper understanding of persuasion in software and information systems. In *Advances in computer-human interaction, 2008 first international conference on*, pages 200–205. Ieee, 2008.

[RW85] Robert C Rowland and Deanna F Womack. Aristotle's view of ethical rhetoric. *Rhetoric Society Quarterly*, 15(1-2):13–31, 1985.

[Sen18] Senate Committee on the Judiciary, Senate Committee on Commerce, Science, and Transportation. Facebook, Social Media Privacy, and the Use and Abuse of Data, April 2018.

[TS03] Richard H Thaler and Cass R Sunstein. Libertarian paternalism. *American economic review*, 93(2):175–179, 2003.

[Woo13] Michael Wooldridge. Intelligent agents. In Gerhard Weiss, editor, *Multiagent Systems*. The MIT Press, 2013.