Combating Misinformation Through Nudging

Evangelos Karapanos, Loukas Konstantinou, Ana Caraban

HAL Id: hal-02878644
https://inria.hal.science/hal-02878644

Submitted on 23 Jun 2020

Combating Misinformation Through Nudging

Loukas Konstantinou[1], Ana Caraban[1,2] and Evangelos Karapanos[1]

[1]Persuasive Tech Lab, Cyprus University of Technology, Limassol, Cyprus
[2]Instituto Superior Tecnico, Universidade de Lisboa, Lisboa, Portugal
lukekonsta@gmail.com, akc.caraban@hotmail.com,
evangelos.karapanos@cut.ac.cy

Abstract. Recent studies have shown that the spread and consumption of misinformation online can be attributed to errors in human decision making, facilitated by cognitive biases. The field of Behavioral Economics has contributed a repertoire of such cognitive biases that can be leveraged for the design of technological interventions. In particular, the concept of *nudging* refers to subtle changes in the 'choice architecture' that can alter people's behaviors in predictable ways. In this paper we present our ongoing work on the design of nudging interventions in the context of misinformation, including a systematic review of the use of nudging in HCI that has led to a design framework consisting of 23 mechanisms of nudging tapping to 15 different cognitive biases, the translation of this framework into a set of design cards, the *Nudge Deck*, and its use in a planned workshop that aims to explore the design space of misinformation in the context of nudging.

Keywords: Behavior change, misinformation, nudging.

1 Introduction

In the era of "post-truth", the rise of falsehood content dissemination in the information environment has sparked significant public concern, since it has been posed as a threat to the very existence of democratic societies. Meinert et al. (2018) define fake news and false information, in general, as purposeful publications of false, discrediting and deceitful content, motivated by various interests (e.g. political, financial, etc.) so as to manipulate and exploit its recipients.

A number of researchers have called for the examination of the factors that result in the dissemination and consumption of falsehood knowledge. As examined in the following sections, various cognitive and emotional biases facilitate the spread and consumption of misinformative content. For instance, Vosoughi et al. (2018) found that increased speed and depth at which fake news diffuse can be attributed to mere novel effect, while Badke (2018) discussed the role of the confirmation bias on the consumption and spread of misinformation.

Based on the concept of *nudging*, we suggest that knowledge about the different cognitive biases that we, as humans, are susceptible to, can be leveraged for the design of technological interventions that minimize the spread and consumption of

misinformation. This paper presents a preliminary account of our ongoing work to-wards this goal.

2   The role of human decision making on the spread and consumption of misinformation

Empirical studies have repeatedly highlighted that mis-informative content propagates faster, deeper, farther than truthful messages. Vosoughi et al. (2018), for instance, used a data set of rumor cascades on Twitter from 2006 to 2017, and found that the top 1% of false news cascades diffused to between 1000 and 100,000 people, whereas the truth rarely diffused to more than 1000 people. They found that fake news was approximately 70% more likely to be retweeted than true ones. A key question raised is: what role does human decision-making play, and how can technology enable humans to make better decisions? Recent studies have highlighted that cognitive biases in decision mak-ing can facilitate the spread, or the consumption of misinformative content. For in-stance, Vosoughi et al. (2018) found that, contrary to conventional wisdom, the spread of false news could not be attributed to the structure of social media outlets, website platforms and internet bots, but rather to a mere novelty effect. Novelty, as the authors claimed, "attracts human attention, contributes to productive decision-making, and en-courages information sharing because novelty updates our understanding of the world". False news was found to be more novel than true news, suggesting that people were more likely to share novel information.

Badke (2018) supports that humans see only what they expect or want to see, without inspecting news thoroughly. This, they argue, is a product of the confirmation bias, the internal tendency of people to seek out information that confirms and verifies what they already believe, instead of examining critically all the pieces of information. According to the theory of cognitive dissonance, whenever a presented piece of news includes information which conflicts with the currently held mental models of people, it imme-diately induces cognitive dissonance [4]. People are motivated to scale down this dis-sonance, thus they may avoid or even discount knowledge that contrasts their personal positions. Weeks (2015) argues that emotional experience moderates the influence of partisanship on individuals' responses to misinformation. Specifically, when individu-als experience anger, the influence of partisanship is boosted, making individuals more likely to believe claims that are associated with their political affiliation. On the con-trary, anxiety reduces the influence of partisanship and increases the chance of making other political affiliations believable. Schwarz et al. (2016) argue that whenever people come across a new piece of information, they tend to assess its truthfulness by focusing on five criteria. People usually ask themselves about the social consensus of the story, its supporting evidence, its consistency, coherence and credibility. However, instead of evaluating these questions analytically, individuals tend to use mental shortcuts in order to minimize the time and energy spent. This makes them susceptible to errors in deci-sion making.

3 *Nudging* away from misinformation

Thaler and Sunstein (2013) define a nudge as "any aspect of the choice architecture that has the capacity to change people's behavior in a predictable manner but without preventing any other alternatives or altering their economic incentives". Nudges leverage knowledge about systematic biases in decision making to support people in making optimal decisions. For instance, the status-quo bias reflects our tendency to resist change and to go along with the path of least resistance [8]. As such, we often chose the default option rather than taking the time to consider the alternatives, even when this is against our best interests. For example, several countries in Europe have changed their laws to make organ donation the default option. In such so called opt-out contexts, over 90% of the citizens donate their organs; while in opt-in contexts the rate falls down to 15%. Similarly, replacing cake with fruit in the impulse basket next to the cash register, has been found to lead people in buying more fruit and less cake, when both choices are still available [9]. As evident, nudges work on the assumption that people don't always make rational choices and they attempt to direct human behavior, yet without imposing a particular choice [9].

Our ongoing work seeks to explore the power of nudging in reducing the likelihood of spreading as well as consuming misinformation. Through a systematic review of the use of nudging in HCI research, we have identified 23 distinct mechanisms of nudging developed within HCI, clustered in 6 overall categories, and tapping to 15 different cognitive biases and heuristics [8]. One such mechanism is called *"reminding the consequences"*. Tapping on the *availability heuristic* which reflects our tendency to judge the probability of occurrence of an event based on the ease at which it can be recalled, thus making us susceptible to underestimate the probability of events when these are not readily available to our cognitive processing, such nudges prompt individuals to reflect on the consequences of their actions. One example is provided by Harbach et al. (2014) who redesigned the permissions dialogue of the Google Play Store to incorporate personalized scenarios that disclosed potential risks from app permissions. If the app required access to one's storage, the system would randomly select images stored on the phone along with the message "this app can see and delete your photos".

We have translated this framework of the 23 mechanisms of nudging into a set of design cards, the *Nudge Deck* [11]. Each mechanism is described in a card (see figure 1 for an example card), where the front side includes a definition of the nudging mechanism, an image of a possible implementation and an explanation. The back side illustrates suggestions, directions, instructions to provoke motivation and inspiration during the design conception. Moreover, to distinguish the situations in which each nudge should be used, the 23 mechanisms are mapped into the three trigger types, as suggested by Fogg's Behavior Model: *sparks* (i.e., ones to increase motivation), *facilitators* (i.e., ones that increase ability) or *signals* (i.e., ones to remind of the behavior).

Using the Nudge Deck as a design support tool, we currently plan to conduct a workshop with the goal of ideating on nudging interventions to combat the spread, or consumption of misinformative content online. To provide an example, suppose we want to minimize the spread of fake news on Twitter through preventing unaware users from retweeting those fabricated stories. One nudge mechanism that can be exploited is

called "*throttling mindless activity*" [8]. This mechanism taps into *regret aversion bias*, people's tendency to become more reflective and break their mindless activities whenever risk is identified. In this case, given that the user is about to post a tweet that contains bogus news, the tool could notify the user with a message like: "We estimate 90% chance of the article containing falsehood information. Are you sure you want to publish this tweet?". Through instilling doubt, this nudge encourages users to reconsider the tweet's content. A link to a fact-checking article may further provide an opportunity for learning. Our workshop will seek to explore how different nudging mechanisms can be applied in different platforms – from social media, to e-mail, and question and answer (Q&A) websites like WikiAnswers. We will engage students of interaction design in small design teams and seek to provide a first inquiry into the design space of technology-mediated nudging in the context of misinformation-resilience tools, through an analysis of the design ideas that come out of the workshop and the emerging design qualities of those ideas.
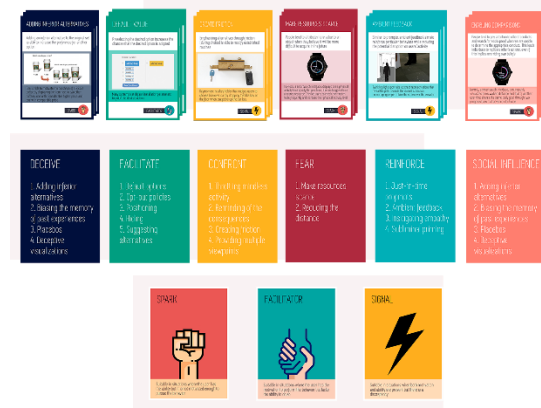

Figure 1: The Nudge Decks consists of 23 *mechanism cards* (top), six *category cards* (middle), and three *trigger cards* (bottom

4   Conclusion

This paper presented our ongoing work on the design of nudging mechanisms to combat the spread and consumption of misinformation online. Our future work will aim at exploring the design space of nudging interventions in the context of misinformation, as well as prototyping a set of intervention techniques with the goal of assessing their proximal effects on human behavior.

5   Acknowledgments

References

1. Meinert, J., Mirbabaie, M., Dungs, S., & Aker, A. (2018). Is It Really Fake? – Towards an Understanding of Fake News in Social Media Communication. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10913 LNCS*(July), 484–497. https://doi.org/10.1007/978-3-319-91521-0_35

2. Vosoughi, S., Roy, D., & Aral, S. (2018). News Online, *1151*(March), 1146–1151. https://doi.org/10.1126/science.aap9559

3. Badke, W. (2018). Fake News, Confirmation Bias, the Search for Truth, and the Theology Student. *Theological Librarianship*, *11*(2), 4–7. https://doi.org/10.31046/tl.v11i2.519

4. Festinger, L. (1957). *A theory of cognitive dissonance*. Palo Alto. California: Stanford University Press.

5. Weeks, B. E. (2015). Emotions, Partisanship, and Misperceptions: How Anger and Anxiety Moderate the Effect of Partisan Bias on Susceptibility to Political Misinformation. *Journal of Communication*, *65*(4), 699–719. https://doi.org/10.1111/jcom.12164

6. Schwarz, N., Newman, E., & Leach, W. (2016). Making the truth stick & the myths fade: Lessons from cognitive psychology. *Behavioral Science & Policy*, *2*(1), 85–95. https://doi.org/10.1353/bsp.2016.0009

7. Thaler, R. H., & Sunstein, C. (2013). Nudge: improving decisions about health, wealth, and happiness. *Choice Reviews Online*, *46*(02), 46-0977-46-0977. https://doi.org/10.5860/choice.46-0977

8. Caraban, A., Gonçalves, D., Karapanos, E., & Campos, P. (2019). 23 Ways to Nudge : A Review of Technology-Mediated Nudging in Human-Computer Interaction. *Computer-Human Interaction, 50,* 1-15.

9. Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

10. Harbach, M., Hettig, M., Weber, S., & Smith, M. (2014). Using personal examples to improve risk communication for security & privacy decisions, (April), 2647–2656. https://doi.org/10.1145/2556288.2556978

11. Caraban, A., Karapanos, E., Konstantinou, L. (2019) The Nudge Deck: A Design Tool for Technology-Mediated Nudging, M-ITI Technical Report.