On Social Acceptance of UI Intervention Mechanisms on Posting and Reading Comments on Online News

Joel Kiskola, Thomas Olsson, Heli Väätäjä, Veikko Surakka, Mirja Ilves

[1] Tampere University, Kalevantie 4, 33014 Tampereen yliopisto, Finland
{joel.kiskola, thomas.olsson, heli.vaataja, veikko.surakka, mirja.ilves}@tuni.fi

Abstract. Issues in the discussion culture in social media call for new approaches to improve, for example, the practices of commenting online news articles or similar public content. Our ongoing research aims to design and develop user interface mechanisms that could automatically intervene the reading or commenting experience in order to enhance emotional reflection and thus improve online behavior. While this aim might seem desirable, it is a conundrum where the solutions need to carefully balance various requirements and values. For example, automatic moderation of the messages might violate the fundamental right to freedom of opinion, and computationally tampering the intimate act of human communication might feel inappropriate. This paper discusses various issues from the perspectives of social acceptance and ethics by presenting three seemingly effective, yet problematic design explorations. Following the ideology of critical design, we contemplate how the design conventions in social media could be changed without introducing adverse behavioral consequences.

Keywords: Emotion Regulation, Critical Design, Social Media.

1 Introduction

In both the academic community and public discourse, we have recently seen heated discussions on how the various services have detrimentally affected the communication culture. Issues like social media rage, hate speech [5] cyberbullying, and increased polarization of the opinion sphere [6] could be considered as side effects of using digital media as the channel for public discourse and opinion exchange. However, the processes and reasons behind these symptoms are much deeper than people misbehaving in such digital communication services.

We suggest that the symptoms result from processes related to emotions and emotion regulation. The ability to regulate one's emotions and mood is a necessity practically for every area of life [4] but has been found to be challenging in technology-mediated textual communication. Emotions are widely expressed in textual format in digital media environments, such as social media services, online communities, and commenting threads of journalistic content, but it has been argued that the lack of nonverbal cues in textual communication deteriorates the ability to control emotions and empathize with other people [10]. Thus, we should better understand how emotions actually function in such communication and develop mechanisms that help individuals to regulate emotions.

Emotion processes operate largely unconsciously. An example of this is the case of emotional mimicry. People tend to react automatically to other people's emotion expression stimuli so that when we see or hear others' expressions of joy or anger, for example, we tend to mimic them without being conscious of what we saw or heard [3, 8]. Additionally, visually presented emotional words have been shown to evoke emotions. In digital media environments, it has been found that the conversation context, mood and other contextual factors can increase the probability of anyone writing uncivil comments [2].

Recent evidence shows that *affect labelling* (e.g., turning emotional cues into words) can attenuate emotional experiences and thus be one form of emotion regulation. Studies have shown that affect labelling does have significant effects on emotion related physiology, behavioral responding, and experiences. This is called as implicit emotion regulation because it does not require conscious intent to regulate emotional experience [9]. This type of process could be a potential option for unobtrusive emotion regulation in social media.

This challenging application area and research goal calls for critical thinking and systematic analysis of the existing UI mechanisms in computer-mediated communication. Consequently, we utilize *critical design* [1], which applies knowledge from social sciences and humanities for reflective design of artefacts, foregrounding the ethics of design practice, revealing potentially hidden agendas and values, and exploring alternative design values. Critical design has been argued to allow better understanding and shaping technologies that can lead to negative outcomes. Design artefacts are used to make consumers more critical about "how their lives are mediated by assumptions, values, ideologies, and behavioral norms inscribed in designs" [1].

Having said that, applying critical design to improve online discussion culture necessitates a careful analysis of the possible behavioral consequences of the developed UI mechanisms and how people could appropriate them in various ways, some of which might be detrimental. This position paper contributes a critical analysis from the viewpoint of social acceptance with regard to three preliminary and speculative concept designs. Rather than trying to theorize or define the notion of social acceptance, this paper identifies domain-specific risks and issues that could help doing so at the workshop.

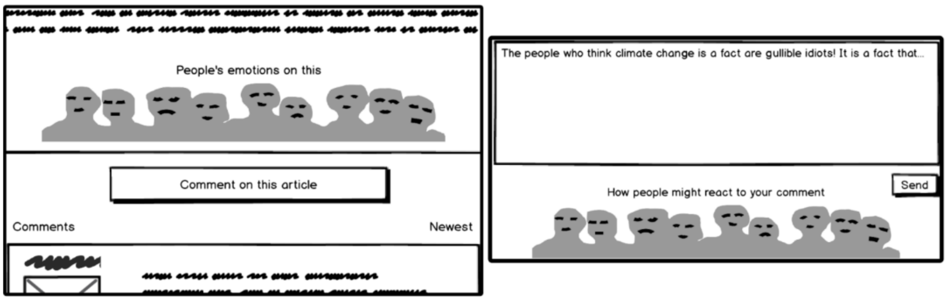
2 Designs and Critique

2.1 On the Design Space / Design Principles

We subscribe to the idea of implicit *affect labelling* by Torre & Lieberman [9], that is, making the emotionally loaded elements in a message more explicit. Our designs for this expect a future where we have advanced methods of natural language processing and human-labeled training data for supervised machine learning.

These designs are three handpicked examples out of 50+ ideas, included here because they elicit different kinds of social acceptance issues. However, the designs share the principle that affect labelling is meant to be purely personal and not visible to others (other, remote users of the platform).

2.2 Design 1: Virtual Audience


People's emotions on this
Comment on this article
Comments
Newest
How people might react to your comment
Send

The people who think climate change is a fact are gullible idiots! It is a fact that...

Fig. 1. Left: A virtual audience showing emotional reactions present in the discussion would appear on top of the comment section. Right: Anticipated reactions to user's writing.

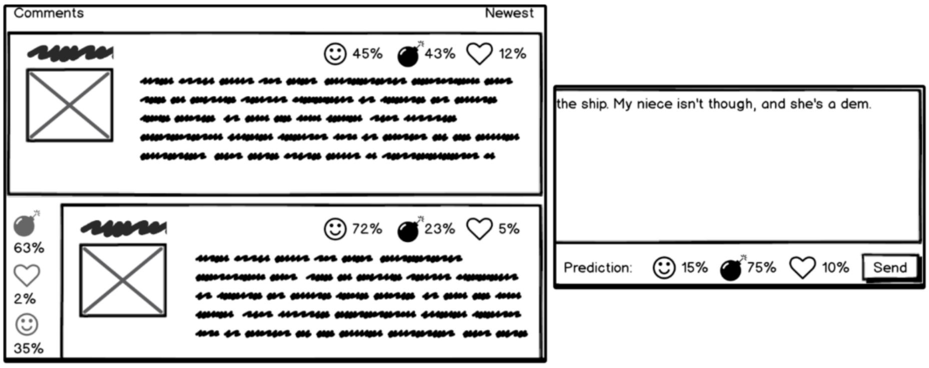
In the Virtual Audience design, the user intends to read the comments to an article when they see an array of abstract, yet animated anthropomorphic figures with various facial expressions (see **Fig. 1** left side). The facial expressions represent the emotional reactions present in the discussion. In addition, when one starts to write a comment, a similar visualization of the anticipated reactions (of different kinds of people) begins to form (see **Fig. 1** right side).

The design attempts to solve the practical problem that to understand how people feel about an article and the comments requires carefully reading the comments. The emotional reactions are summarized to give a sense of a live audience. The more specific critical design principles that the design utilizes include:

- Humanization of text that could otherwise seem impersonal.
- Social pressure: people generally want to produce positive emotions in others.
- People are wired to look at human faces.
- Ambiguity in how the facial expressions come about.
- Exaggeration of facial expressions and contrasts between the expressions.
- Gentle satire: imitating opposite emotional reactions to texts, to ridicule people.

The design introduces several potential issues of social acceptability and ethics. **(1)** The virtual audience may feel like an actual audience and this may evoke more real life like normative behavior in the digital environment. **(2)** The virtual audience may highlight or greatly increase the impact of the first comments; hence, the first commenters may feel that their comments are given a disproportionate amount of attention by the virtual audience. **(3)** Users might start to optimize their comments to reach positive audience reactions; alternatively, some users might be provoked to opposite behavior. **(4)** The virtual audience, being an easily observable UI element, may enable collocated people to judge the quality of a commenter's writing. **(5)** The virtual audience might become a key element of the public image of a certain digital platform or news broadcaster, which might contradict with how they want to be seen. Furthermore, some commenters might be considered as obedient or disobedient, affecting their public image.

2.3 Design 2: Emotion Symbols



the ship. My niece isn't though, and she's a dem.

Fig. 2. Left: Users rate comments for their emotional qualities and the system calculates percentages of ratings for comment threads and individual comments. Right: The system predicts what kind of emotional reactions the comment would elicit.

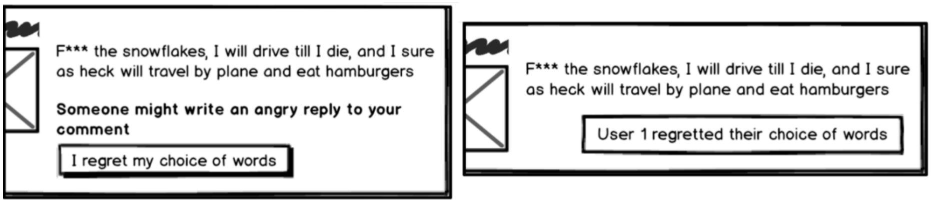
The Emotion Symbols design mimics the convention of giving certain reactions to posts, but approaches this by explicating one’s emotional reaction to a message. While Fig 2. displays only three types of labels (a general positive reaction, “this is explosive” and “loving this”), the vocabulary of labels could be very broad. The users can rate the comments by clicking the symbols. In addition, when a user is writing a comment, they will see the symbols and percentages change based on what they write, according to the system’s prediction on what kind of emotional reactions the comment would elicit.

The design attempts to solve the problem that there is no explicit information on the emotional content of the comments. It proposes to explicate the emotional quality of each comment and comment thread in a quantified way to help to select which comments or threads to read. Other principles that the design utilizes include:

- Playfulness: the symbols chosen to represent emotions (e.g., hearts and bombs) are visually playful.
- Gamification: e.g., users may try to get hearts or bombs.
- Ambiguity: leaving room for interpretation on what contributes to the percentages, which can encourage people to reflect on the messages they create.

The social acceptability and ethical issues include, for example, the following. **(1)** The commenter may feel that this design increases the risk that they will be bullied. Getting “bombed” or assigning other labels introduces new mechanisms of giving feedback, which might affect self-esteem. **(2)** Related to quantification, some may find it questionable that the nuanced and highly subjective semantics in their comments are reduced into numbers. As Lucy Suchman warns, any form of categorizing bears the risk of politicizing, with which minds can be formed and opinions made [7]. **(3)** While writing, it can feel awkward that an algorithm defines the *value* of the comment. **(4)** Related to the previous, some users may try to ”game the system” and try to maximize or minimize the metrics. This provides a new potential reason for writing comments, which undermines the primary communicative purposes of writing comments.

2.4 Design 3: Regretting one’s choice of words



F*** the snowflakes, I will drive till I die, and I sure as heck will travel by plane and eat hamburgers

Someone might write an angry reply to your comment

F*** the snowflakes, I will drive till I die, and I sure as heck will travel by plane and eat hamburgers

I regret my choice of words
User 1 regretted their choice of words

Fig. 3. Left: User is given a chance to regret one’s choice of words after publishing a seemingly uncivil comment. Right: User 2 sees a note that user 1 has regretted their words.

In the Regret design, user 1 has just published a comment and they are looking at it. Then they see a notification on their comment that allows regretting one’s words (see **Fig. 3**, left). Alternatively, the user 1 may regret after seeing what kind of a mess their comment caused. It is noteworthy that only the user sees the notification and only after clicking the regret button the other users see this as an extra label (**Fig 3**, right).

The design attempts to solve the problem that there are no quick and easy ways for a commenter to regret what they wrote or how they placed their words; editing a published comment requires more skill and effort, and deleting one’s comment entirely might not be desirable either. In other words, the design introduces a light-weight way for a user to notify others that they are not happy with their comment either, for example, to help resolving heated discussions. More specific principles include:

- Surprise: if the user does not realize their comment is controversial, notification by the system will surprise them. Moreover, regretting can be surprising to other users.
- Implying that messages should not be read too literally.
- Drama: it can be thought to be dramatic when someone regrets what they said.
- Social conventions: regretting is a universal behavioral pattern related to forgiveness
- Gamification: the design adds cost-benefit calculation to the discussion, making it more game-like; and the regret notification is “armor” against criticism.

The potential social acceptability and ethical issues include the following. **(1)** Users may consider regretting like this to be too easy to be counted as real regretting. **(2)** Some users might start writing more thoughtlessly than before, thinking, “you can regret it later, right?” The discussion might start resembling more synchronic communication, however, without the benefits of the multimodal face-to-face channel. **(3)** The system might feel patronizing and awkward in some cases (presuming it lacks “common sense” and does not recognize that strong language is sometimes ok).

3 Discussion and Conclusions

We presented work-in-progress on UI designs that aim to improve emotional reflection in social media discussions. While our intention is to create ethically sustainable designs and to avoid compromising social acceptance, this preliminary analysis implies that identifying a design that is at the same time effective and sustainable is challenging. Each design has their pros and cons. We would gladly continue the discussion on problematizing the existing UI mechanisms in social media and the presented designs. A more thorough analysis of the potential ramifications could be implemented by, for example, reflecting on certain items in the human rights declaration by the United Nations (e.g., freedom of opinion and expression, peaceful assembly, free participation in cultural life). Various moral philosophical doctrines (e.g., starting all the way from Nichomachean Ethics by Aristotle, and other virtue ethics) would also provide insightful viewpoints. That said, while Critical Design is all about questioning various conventions, we argue that especially in this kind of application area something that should *not* be deliberately twisted are the ethical principles—they also shapes people's perceptions of what kind of technology is acceptable.

References

1. Bardzell, J. & Bardzell, S. (2013). What is "critical" about critical design? Proc. of CHI '13. ACM, New York, NY, USA, 3297-3306.
2. Cheng, J., Bernstein, M., Danescu-Niculescu-Mizil, C., & Leskovec, J. (2017). Anyone can become a troll: Causes of trolling behavior in online discussions. In CSCW 2017, February 25–March 1, 2017, Portland, OR, USA.
3. Fischer, A., & Hess, U. (2017). Mimicking emotions. Current opinion in psychology, 17, 151-155.
4. Gross, J.J. (1998). The Emerging Field of Emotion Regulation: An Integrative Review. Review of General Psychology, 2, 271-299.
5. Guiora, A. & Park, E.A. (2017). Hate speech on social media. Philosophia, 45(3), 957-971.
6. Nelimarkka, M., Laaksonen, S. M., & Semaan, B. (2018). Social media is polarized, social media is polarized: towards a new design agenda for mitigating polarization. In Proceedings of the ACM Conference on Design Interactive Systems (DIS'18).
7. Suchman L. (1993) Do Categories Have Politics? The language/action perspective reconsidered. Proc. of ECSCW '93. Springer, Dordrecht
8. Surakka, V. & Hietanen, J.K. (1998). Facial and emotional reactions to Duchenne and non-Duchenne smiles. International Journal of Psychophysiology, 29 (1), 23-33.
9. Torre, J. B., Lieberman, M. D. (2018) Putting Feelings Into Words: Affect Labelling as Implicit Emotion Regulation. Emotion Review, 10, 116-124.
10. Walther, J.B. (1993). Impression development in computer-mediated interaction. Western Journal of Communication (includes Communication Reports), 57(4), 381-398.