Applying Critical Voice in Design of User Interfaces for Supporting Self-Reflection and Emotion Regulation in Online News Commenting

JOEL KISKOLA

Tampere University, Tampere, Finland, joel.kiskola@tuni.fi

THOMAS OLSSON

Tampere University, Tampere, Finland, thomas.olsson@tuni.fi

HELI VÄÄTÄJÄ

Lapland University of Applied Sciences, Rovaniemi, Finland, heli.vaataja@lapinamk.fi

ALEKSI H. SYRJÄMÄKI

Tampere University, Tampere, Finland, aleksi.syrjamaki@tuni.fi

ANNA RANTASILA

Tampere University, Tampere, Finland, anna.rantasila@tuni.fi

POIKA ISOKOSKI

Tampere University, Tampere, Finland, poika.isokoski@tuni.fi

MIRJA ILVES

Tampere University, Tampere, Finland, mirja.ilves@tuni.fi

VEIKKO SURAKKA

Tampere University, Tampere, Finland, veikko.surakka@tuni.fi

On digital media services, uncivil commenting is a persistent issue causing negative emotional reactions. One enabler for such problematic behavior is the user interface, conditioning, and structuring text-based communication online. However, the specific roles and influences of UIs are little understood, which calls for critical analysis of the current UI solutions as well as speculative exploration of alternative designs. This paper reports a research-through-design study on the problematic phenomenon regarding uncivil and inconsiderate commenting on online news, envisioning unconventional solutions with a critical voice. We unpack this problem area and outline critical perspectives to possible solutions by describing and analyzing four designs that propose to support emotion regulation by facilitating self-reflection. The design choices are further discussed in respect to interviews of ten news media experts. The findings are reflected against the question of how can critique meaningfully manifest in this challenging problem area.

CCS CONCEPTS •Human-centered computing~Interaction design~Interaction design theory, concepts and paradigms

Additional keywords and phrases: Design Research, Critique, Critical Design, Design Fiction, Social Media, Digital Media, Online News, Emotional reflection, Design conventions, Expert interviews

ACM Reference Format:

First Author's Name, Initials, and Last Name, Second Author's Name, Initials, and Last Name, and Third Author's Name, Initials, and Last Name. 2018. The Title of the Paper: ACM Conference Proceedings Manuscript Submission Template: This is the subtitle of the paper, this document both explains and embodies the submission format for authors using Word. In Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY. ACM, New York, NY, USA, 10 pages. NOTE: This block will be automatically generated when manuscripts are processed after acceptance.

1 INTRODUCTION

In both academic literature and public discourse, the communication culture on digital media services has been widely problematized, with scholars referring to behavioral and cultural issues like social media rage, use of uncivil language [22], and increase of hate speech [36]. People generally consider such issues as nuisances to be mitigated, which has motivated various solution approaches, ranging from human-based content moderation and enforcement of commenting guidelines [31, 61] to computational detection of hate speech [20] and toxic language [51]. However, the aforementioned behavioral and cultural issues remain hardly solved as the underlying reasons behind the behavior are probably numerous. Based on the long-standing discussion on computer-mediated communication [15, 63, 72], the present work poses that a central, yet relatively superficially understood factor behind the issues is how the user interface (UI) affords, conditions, and structures social interaction online. Computer-mediated communication [72] can be seen to force nuanced public discourse and opinion exchange through an inherently narrow channel, disregarding many emotional elements in interpersonal communication. From the perspective of emotional psychology, the current, largely text-based interfaces inherently limit the ability to control one's emotions or to empathize with other people [71]. This arises a need for creative design exploration to understand how UI design could provide new perspectives to this problem area and open new avenues towards UI solutions for emotion regulation.

In this work, we explore possible future UIs for self-reflection and emotion regulation in relation to the activity of commenting on news articles on online news sites. Online news commenting is a form of more or less anonymous public discourse between strangers [22] that takes place around journalistic content on the comment sections of online newspapers and broadcasters. We consider uncivil and inconsiderate commenting of online news as an intriguing problem area for design approaches where the critique of tradition and the status quo is emphasized. According to Bardzell & Bardzell [6], unconventional design artifacts may be introduced to make consumers more critical about "how their lives are mediated by assumptions, values, ideologies, and behavioral norms inscribed in designs." We particularly attempt to reflectively design artifacts that could support commenters' self-reflection and to unpack the role of UI design in this problem area [22, 61].

Incivility in online news commenting can be considered as a wicked problem: it is ill-defined, has no straightforward solutions, and manifests other "higher level" problems [14]. For example, the definitions of incivility (or related terms) are debatable and hard to apply in practice [61], and there is a long-standing academic discussion on what counts as legitimate expression of public opinion (e.g., [27, 37, 60]). Different forms of misbehavior in online news commenting may result from unknown combinations of behavioral and cultural issues (e.g., intentional trolling, commenting in an inconsiderate manner evolving into hateful discussion threads, unclear norms on online platforms). In addition to harmful effects to the involved commenters, uncivil comments can hurt news reporters and moderators who cannot easily avoid them [29], harm the publisher's brand [55], and evoke negative effects on the majority of readers who do not participate in commenting [17]. To this end, this problem area calls for audacious exploration of alternative solution proposals and research through design that could provide new perspectives to the related problems as well as open new avenues towards more sustainable UI solutions.

Our design exploration draws from two theoretical frames. The first is the evolving design philosophy of Critical Design (CD) [6, 7, 25, 54, 68]. Mindful of its various interpretations, we avoid subscribing to any specific school of thought. For example, CD may be expected to empower people or combat those in power [38], be associated with the term's originators, Dunne, Raby, and their students and disciples [54], or assumed to make a strong critical contribution in a broader sense [6]. To best reflect the design mindset in this study, we term our approach as designing with a critical voice. With this, we aim to find a balance between introducing thought-provoking perspectives (i.e., designs for raising questions) and creating design ideas that are potentially effective and socially acceptable as solutions (i.e., designs for solving problems). The second theoretical frame is the concept of *self-reflection* and *affect labeling* as an implicit form of emotion regulation

[70]. In affect labeling [70], emotion regulation can result from simply making the emotionally loaded elements in a message more perceivable. We attempt to propose sufficiently provocative forms of affect labeling to raise awareness of and discussion of the role of the UI.

However, exploring how critique can manifest acceptably in a problem-focused context is an ambitious aim. After all, there is little practical guidance or heuristics for using critique in the UI design practice, and there are relatively few examples of UI design projects where a critical voice would be emphasized. In a design project, it is difficult to judge, e.g., what went overboard and what remained inefficient in terms of provocation [8, 9]. For this reason, we first carefully analyze the criticality of our designs and then interview news media experts from media organizations to gain additional critical perspectives and feedback on their perceived risks and opportunities. Accordingly, we also remain critical of the concept of trying to nudge [16, 67] emotional reflection in online news commenting as well as what is and is not uncivil. The following related work section outlines relevant literature that the work builds on—related to political conversations [27, 37] and polarization [33, 45] in online discourse, discussion moderation [31, 61], and emotional regulation [34, 70]. That said, the contribution of this work targets the growing literature of critical design [6, 7, 11, 25, 38, 54, 68].

The contributions of this work include: (1) identification of critical perspectives to a particular problem area for UI design; (2) presentation of four selected design artifacts that embody different critical perspectives and could serve as solutions (or inspirations to other solutions) to mitigating incivility and inconsideration in online news commenting; (3) insights into the acceptability of the designs based on interviews of experts in administering online discussions in relation to news articles, and (4) reflection on applying design with a critical voice to problem-focused UI design case, contributing to the methodological development of critical design.

2 RELATED WORK AND POSITIONING

In the following, we first analyze how our design approach relates to the views and theories on criticality in design and discuss how prior critical design works inspired our design endeavors. Next, we cover moderation strategies for solving emotionally troubling online discussion. We further explain the concept of implicit emotion regulation and position the concept in relation to moderation, critical design, and the concept of behavioral nudging.

2.1 Criticality in Design Theory

While the notion of critique is often implicitly embedded in the design of interactive systems, there are several traditions that particularly encourage critique and consideration of alternative user-product relationships. Some notable examples include Critical Design [6], Reflective design [64], Design Fiction [10], Value-Sensitive Design [28], Ludic Design [30], and Critical Technical Practice [2]. In this paper, we apply critical design thinking somewhat like what has been done under the label Critical Design.

The design research described in this paper is inspired by Bardzell & Bardzell [6, 8] views on the appearance of criticality in design in particular because they provide a useful framework for the analysis of criticality. According to their view, the criticality of designs is tied to the display of some number of nonobvious or novel design features, which one can argue to perform a critical function, express criticality, etc. [8]. In other words, to create a design that performs a critical function, one should introduce "twists" (i.e., nonobvious changes) on the standard design [8, 41]. However, if the number or 'mass' of the features is too high, the object may be dismissed as art. "Presumably, critical mass is achieved when one believes that the judgment could credibly demand assent from others, or at least provoke constructive further discussion and analysis" [ibid.].

At the same time, the characteristics of provocative and unconventional designs that may facilitate critical thought depend on the user's ability to read designs insightfully [8], and this seems to be emphasized in many designs labeled as critical designs, for example, the works by Dunne and Raby [24]. However, our intention is to facilitate critical thinking about design *for everyone*, including individuals with little expertise to read designs. Furthermore, the context of digital media in terms of news websites and social media platforms provides opportunities and challenges that are not present in physical product design (e.g., publicity, a different type of interaction), which is where works labeled as critical designs usually seem to operate. Also, while we design in a more problem-focused manner than much of the prior literature on critical design describes (e.g., [9, 25]), we are still motivated to achieve audience reflection and seek to create designs that serve fairly obvious critical purposes. Hence, the presented design artifacts could thus potentially be read as critical designs [7, 8]. In other words, following Blythe et al. [11], we do not view construction and criticism as polar opposites.

We aim at designs that are as *easy to read* and *plausible as solutions* as those created by Khovanskaya et al. [42] and Raptis et al. [57]. Khovanskaya et al. [42] developed and studied a web-browser plugin that uses unconventional ways to display information about user's web-browsing activities to promote awareness of infrastructures behind personal informatics. Their design strategy was to display surprising perspectives to sensitive and highly personal aspects of gathered data. Raptis et al. [57] conducted research through design focusing on the element of provocativeness and designed a device that challenges families' energy-consuming practices. The device meddles with the availability of electricity for doing laundry and aims to change laundry practices by provoking reflection. While Khovanskaya and others [42] did not state behavioral change as their goal, the realization that "everybody knows what you're doing" online could also cause a change in users' web-browsing habits. Furthermore, the design by Raptis et al. [57] challenged the energy-consuming practices, went beyond persuasion, and made families reflect on their energy consumption and technology's role in it.

Further, we aim at designs that *ask*, rather than tell, what is good design and what is bad commenting. This aim arises from the knowledge of how difficult it is to accurately define the limits of incivility or “freedom of expression” [61]. We acknowledge the long-standing discussion on the (in)civility of public discourse (e.g., [27, 37, 60]), debating questions like whether dispassionate deliberation is synonymous with legitimate expression of public opinion [27] or not. While our design endeavor is motivated in part by this discussion, it is also why we avoid defining what is and is not uncivil: we believe doing so would make the design work too opinionated, unambiguous, norm-enforcing, expected, and to require a strong stance about the hard-to-demarcate concept of civility. After all, ambivalence can also be important for a design's criticality [41, 54].

2.2 Strategies for Moderating Uncivil Online Discussion

Ruckenstein & Turunen [61] identify two kinds of logic in content moderation [31] on commercial platforms: the *logic of choice* focuses on finding and deleting uncivil or ‘not neutral enough’ messages, while the *logic of care* may tackle all kinds of mess and disorder in the user-generated content and involves moderator-writer interaction. Most existing approaches to content moderation fall under the logic of choice. They involve little moderator-writer interaction, tend to break the natural flow of discussion, and even risk the freedom of speech (e.g., users flagging messages, paywalls, limited characters, algorithmic moderation to quarantine or delete messages). However, the authors [ibid.] argue that the logic of choice is not enough to improve online discussion as it fails to encourage behavioral change. In the logic of care, moderators attempt to persuade writers and readers to reflect, and/or to educate them, to improve the discussion. For example, a moderator could intervene in discussion, message a user privately, or hand out badges or prizes to civil writers. The drawback is that human moderator-driven approaches are costly, hard to scale, and potentially traumatizing for moderators as they need to deal with emotionally troubling writings. As a recent example of a relatively low-cost but hard to scale solution, Norwegian Broadcasting Corporation has incorporated custom-built quizzes to confirm the user read the article [35].

Machine learning-based solutions have been explored to address the issues of cost and scalability. One example is the Perspective API developed by Jigsaw [40]. It can detect toxic writing to some extent, and this can be shown to the writer as a score, an emoji, or made to trigger a notification that attempts to persuade the writer to reflect one's writing. The API has been integrated into Spanish language news site El País' comment writing system and it has been reported to have moderately improved the quality of discussion [21].

Algorithmic approaches may also be used to show the readers a sentiment analysis of the published comments, which may make some users stop to think. For example, Yahoo News features a row of three small emoji and percentages (smiling emoji, neutral emoji, sad emoji) to visualize the overall sentiment of the comments (see also Napoles et al. [50]). However, we could not find reports with evidence that the displayed sentiment analysis would affect the quality of news commenting.

While such algorithmic solutions are worth considering, we argue that they are not yet guiding enough (cf. [50]) and might introduce new ethical problems. As the problem of uncivil commenting persists, we argue for the exploration of alternative approaches, as explained in what follows.

2.3 Supporting Emotion Regulation by Facilitating Self-Reflection

To complement the dichotomy by Ruckenstein & Turunen [61], we suggest a third approach: supporting emotion regulation with the help of automatic identification of emotional elements. Building on research on emotion psychology, we suggest that many of the issues in the discussion culture on digital media 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 [34] but has been found to be especially challenging in computer-mediated textual communication. Furthermore, it has been argued that the lack of nonverbal cues in textual communication deteriorates the ability to control emotions and empathize with other people [71]. We explore ways to promote emotion regulation as well as ways to highlight the idea through UI design.

Recently, the concept of *implicit emotion regulation* has been discussed in literature. In contrast to explicit emotion regulation, which requires a conscious effort to for example suppress emotion responses, implicit regulation is effortless and potentially automatic [70]. Therefore, implicit emotion regulation appears promising as a design concept in the context of this study. Emotion regulation may be improved by affect labeling [ibid.]: for example, simply making the emotionally loaded elements in a message more perceivable. Still, the effect is counterintuitive [ibid.], and not well understood. We have found no research on using computational affect labeling in digital media to help understand the emotional nuances in ongoing discussion or to manage emotional reactions. In the present work, we take the idea of labeling as an inspiration rather than as a boundary and explore various tactics to make the users more aware of the emotional elements in the messages.

To further position our work, we recognize that the idea of supporting emotion regulation by facilitating self-reflection relates to nudging theory [67] and critical artifacts. In general, affect labeling can be an approach to nudging (towards emotion regulation) as it gently guides the user while preserving freedom of choice. However, proposing to do so in the context of online news commenting is likely to generate debate, which often is a goal for critical artifacts [8]. Critical artifacts can be seen to manifest nudging — of thought rather than action. That said, CD artifacts often contain more complex, provocative, and reflective arguments than nudging artifacts do (e.g., compare nudging artifacts discussed by Caraban et al. [16] to CD artifacts discussed by Pierce et al. [54] and Bardzell et al. [8]).

3 DESIGN EXPLORATION: PROCESS AND OUTCOMES

The following sections detail the main steps in our research-through-design exploration.

3.1 Identifying Cultural and UI Conventions

To create unconventional designs, current design conventions were first identified by analyzing social media platforms and news websites. Specifically, we examined the commenting systems in the 15 most popular—by traffic—news websites in the U.S. [26]. Further, as the research took place in a Finnish university, we examined them in four most popular Finnish news websites (tabloids Ilta-Sanomat and Iltalehti, national newspaper Helsingin Sanomat, and Finland's national broadcaster Yle) [3]. This resulted in lists of existing *UI conventions* (e.g., option to sort comments by recency) and *cultural conventions* (e.g., people are rarely specific about the intended audience). The lists were used in three ways: to find a convention to be twisted, to avoid reinventing existing solutions, and to reflect what kind of solutions might fit different news websites.

3.2 Idea Generation, Filtering and Selection

In sum, 60 concept ideas were sketched on paper based on several idea generation sessions. Based on an iterative selection process, four design artifacts were selected to be analyzed in this paper. The process included two major phases: idea generation and filtering & selection (Figure 1).

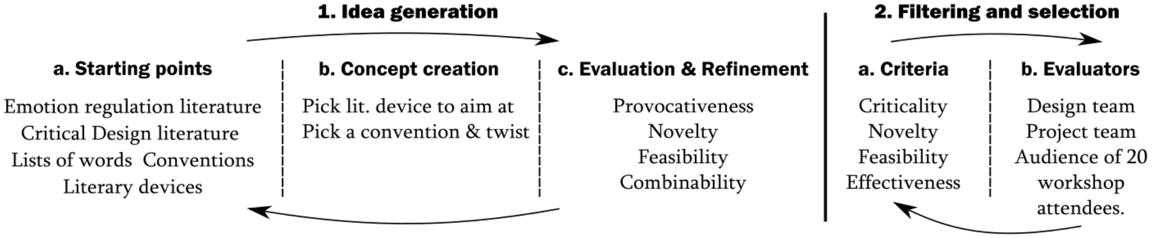

1. Idea generation
a. Starting points
Emotion regulation literature
Critical Design literature
Lists of words Conventions
Literary devices
b. Concept creation
Pick lit. device to aim at
Pick a convention & twist
c. Evaluation & Refinement
Provocativeness
Novelty
Feasibility
Combinability
2. Filtering and selection
a. Criteria
Criticality
Novelty
Feasibility
Effectiveness
b. Evaluators
Design team
Project team
Audience of 20 workshop attendees.

Figure 1. Approximation of the iterative idea generation and filtering and selection process.

3.2.1 Idea Generation

We began by explicating the different theoretical and conceptual sources of inspiration for the ideas (1a. in Figure 1). These helped both to envision new designs and analyze and refine the ideas and define a diverse selection of designs for further analysis. Literature on emotion regulation [34, 70] served as a central source of inspiration for the design work. One of the key ideas guiding the process was based on affect labeling [70]: it may help the user to regulate one's emotions if one recognizes that the text contains expression of emotion. Additionally, we drew inspiration from examples and design concepts in critical design literature [9, 39, 48], doctoral theses featuring designs labeled as critical designs [49, 53, 56], as well as from other types of design case studies [5, 43, 75]. The identified UI and cultural conventions, rhetorical strategies [65], and studies of why people comment on the news [4, 22, 66] served as important points of reflection. Having numerous sources was considered crucial to approach the wicked problem area from multiple angles.

In practice, the idea generation was conducted by a design team consisting of the first author, who has formal education in interaction design and industrial design, and of two colleagues, who both have formal education in user experience design and software engineering. To clarify our relation to the specific problem area, we did not have strong viewpoints on moderating news commenting and we tried to dissociate ourselves from specific political agendas and commitments. That said, we did subscribe to the idea that critical design is in part embodying the authorial and critical voice of its designers [54].

The first round of idea generation took 2 weeks, resulting in about 40 ideas and involved mostly the first author. While the concept creation was not guided by specific design creativity methods, such as fictional inquiry or brainstorming methods, two general strategies mentioned in critical design literature were used (1b. in Figure 1): (1) the designer picks a literary device (e.g., irony, sarcasm, parody, ambiguity) and tries to implement it in designs [41]; (2) the designer picks a convention (cultural or UI) and twists it, for example, by introducing a foreign concept, and then reflects on the result [8].

The first 2-week round of idea generation ended in an evaluation session by the design team. The concept sketches were evaluated for their provocativeness, novelty, feasibility, and combinability with other concept sketches (1c. in Figure 1). The evaluation session also resulted in ideas for more areas to explore. For this reason, we engaged in one more round of idea generation, which we ended when we had generated 60 ideas in total.

3.2.2 Filtering and Selection

Through the filtering process, the 60 ideas were narrowed down to the four presented in this paper. In the first round, the design team conducted two evaluation sessions, where the 60 ideas were evaluated for perceived criticality, novelty, feasibility, and effectiveness. This evaluation was based on the authors' subjective judgment on which designs might best yield diverse critical perspectives. In these sessions, 19 of the ideas were judged as more promising than the others. Following this, the first author created UI mockups of the 19 ideas.

Next, the 19 mockups were presented to and discussed by the whole project team, which was extended by two senior scholars who had formal education in psychology and one senior scholar with formal education in computer science. The psychologists speculated on the likely effects of the designs in terms of self-reflection, emotion regulation, and behavioral nudging. Based on this, the designs were narrowed down to 12.

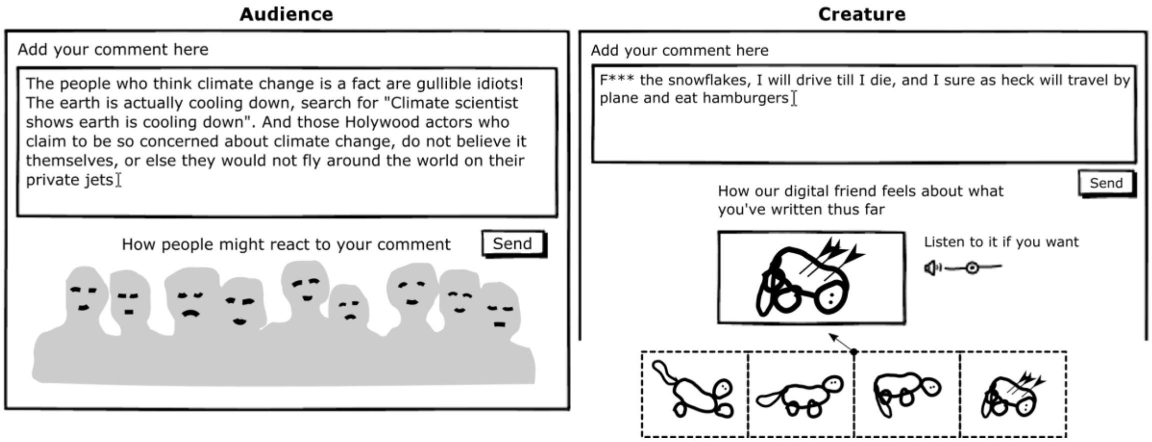
In the third round of filtering and selection, the first author chose 6 designs out of the 12 and presented them in an informal workshop with approx. 20 media scholars, journalists, social media managers, and researchers from other fields. As the designs were presented, the attendees were given a form to quickly rate the designs for acceptability and effectiveness and give short comments in writing. While the results of the evaluation are omitted from this paper, a key implication was that the same six designs presented in the workshop were chosen for the interviews of news media experts because the designs were considered to provoke thought and were not seen as completely unacceptable.

The final selection of the four designs took place based on the expert interviews and while writing the paper. We focus on what we consider the four most suitable designs for discussing the concept of criticality in this problem area in a diverse, nuanced, yet concise manner. The two left out designs are briefly described at the end of the following section.

3.3 The Selected Four Designs: Audience, Creature, Regret, and Promise

For ease of reading, this section introduces the four designs with respect to how they propose to facilitate reflection and emotion regulation, and only in the next section, we analyze why they may be considered manifestations of critical voice in design. Also, we do not go further into technical detail about the designs than noting that while the designs expect a future of advanced content analysis systems, it could be possible to have them work to some extent with existing systems.

The AUDIENCE is an animated graphical element that we propose to represent, with a single image, probable emotional reactions to a comment or discussion thread. As the user is writing a comment, an array of abstract animated anthropomorphic figures with various facial expressions would begin to form as the writing progresses and emotional elements are identified (see Fig. 2 left).


Audience
Add your comment here
How people might react to your comment
Send
Creature
Add your comment here
Send
How our digital friend feels about what you've written thus far
Listen to it if you want

The people who think climate change is a fact are gullible idiots! The earth is actually cooling down, search for "Climate scientist shows earth is cooling down". And those Holywood actors who claim to be so concerned about climate change, do not believe it themselves, or else they would not fly around the world on their private jets

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

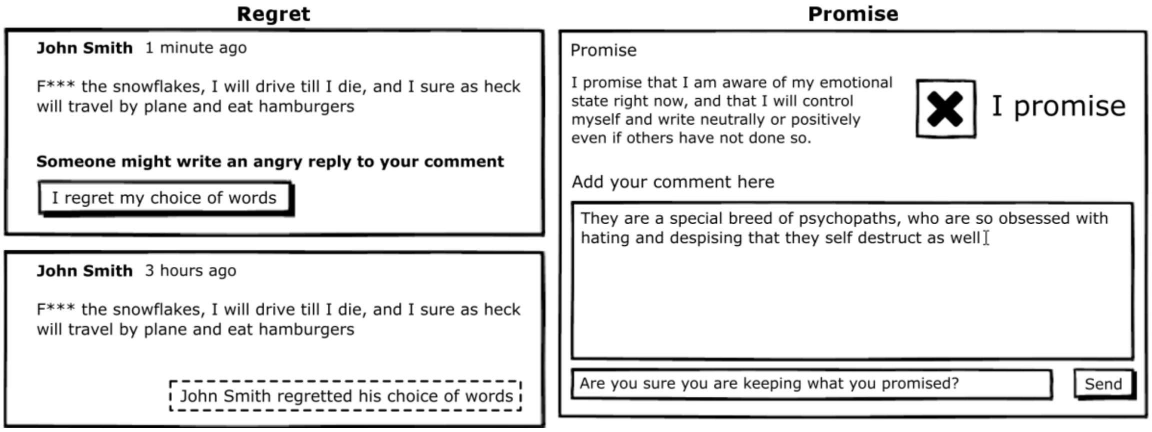
Figure 2. Left: The Audience as it appears for a comment writer. The audience shows the anticipated emotional reactions to the user's writing. Right: The Creature as it appears for a comment writer. The design features a virtual animal that thrives or suffers according to the user's writing.

With the Audience design, we propose to help a commenter to predict how the readers might feel about the comment. Hence, a variety of emotional reactions would be depicted to give a sense of a diverse live audience. Also, we propose the Audience would be placed above the comment section with the intent to show reactions to all published comments. This is because a person who intends to read the comments to a news article might appreciate a hint of what they are about to read. The design relates to affect labeling [70] by possibly helping the user to recognize that the text contains expression of emotion. Also, the design intends to evoke the sense of having a live audience, which may make one consider their self-presentation through writing (e.g., [32]). In this regard, the design also relates to prior work considering how social interaction norms could be applied in designing solutions for enhancing *collocated* social interaction [52].

With the Creature, we propose to highlight the positive and negative effects of emotionally pleasant or troubling commenting through an animated image of an animal right when the users write a comment (Fig. 2 right). How the animal looks like would depend on how emotionally troubling the writing is. If the writing is emotionally positive, the animal would look happy, while if it is very troubling, the animal would appear dead. The user could also listen to the animal by pressing a button if one wishes. Also, we propose to place the animal above the comment section, intending that it would act as a cue of what the user is about to read.

The Creature would work much like the Audience, but we intend it as a more direct take on emotional elements as it reduces the scale of emotions to one dimension (troubling–pleasant) and intends to represent it through the well-being of the creature. We believe this also makes the design relate more to the theory of affect labeling [70] than the Audience, because it may be easier to understand what emotional dimension it reacts to.

With the Regret, we propose to change the dynamics of discussion for the better by providing the writer with a chance to regret their choice of words explicitly and publicly. In Fig. 3 (top-left), John Smith has just published a nasty comment; a notification appears, allowing him to regret his words. Alternatively, the user may regret later, for example, after seeing what kind of a mess their comment caused. If the button is clicked, also the other users would see the writer having regretted their words (Fig 3, bottom-left).



John Smith 1 minute ago

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

I regret my choice of words

John Smith 3 hours ago

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

John Smith regretted his choice of words

Promise

I promise that I am aware of my emotional state right now, and that I will control myself and write neutrally or positively even if others have not done so.

☒ I promise

Add your comment here

They are a special breed of psychopaths, who are so obsessed with hating and despising that they self destruct as well

Are you sure you are keeping what you promised?

Send

Figure 3: Left: The Regret mockup. Top: The user is given a chance to regret their choice of words after publishing a seemingly overly uncivil comment. Bottom: User 2 sees a note that user 1 has regretted their choice of words. Right: The Promise mockup where the user is encouraged to promise to control one's emotions before writing a comment.

With the Regret we propose a way to solve the problem that a commenter typically cannot easily show remorse after posting a comment; editing an already published comment requires more effort and deleting one's comment entirely might not be desirable either. In other words, the design introduces what is intended to be a lightweight way for a user to notify others that they are not happy with their comment either, for example, to help to resolve heated discussions. For a user who reads comments to a news article, the label might act as a cue to skip the comment, or at least to take a deep breath before reading it. Compared to Audience or Creature, the design may relate less to affect labeling [70] as it might be harder to understand that the notification is being triggered by the system after identifying negative expressions of emotion or uncivil phrases. The design may rely more on the self-presentation theory [32] as we intend it to remind the user to manage impressions and follow social norms.

With the Promise, we propose to force the user to make an explicit promise to control their emotions. In Fig. 3 right, an attempt is made to force the user to promise good behavior before they can write a comment, based on predefined text and a large checkbox. If the user writes nastily after promising, a note would appear under the text area that would read, "Are you sure you are keeping what you promised?" With the design, we attempt to solve the problem that users might not stop to reflect on what they are about to write and how. Similar to Regret, this design may rely on the self-presentation theory [32] rather than affect labeling [70]. With the design, we attempt to inform the user that one must be in the right state of mind to comment.

That said, we now briefly describe the two designs we left out when writing the paper. The idea of the first left out design is that uncivil wording in comments is blacked out but can be revealed by clicking the text. We judged the design to be somewhat less credible and novel than the four above. In the other left out design, the idea is somewhat like common comment rating designs, such as up and down voting, except users would rate the comments for *explosiveness*, *love*, etc., using symbol-buttons (bomb, heart, etc.) and the ratings would appear as percentages next to the symbols. We judged this design as somewhat less provocative and more familiar than the four above. In other words, the two were left out as we subjectively judged the four to be more suitable for discussing the concept of criticality in this problem area in a diverse, nuanced, yet concise manner.

4 ANALYSIS OF CRITICALITY OF THE DESIGNS

The following outlines and discusses the various manifestations of criticality in the four designs, intending to show how the designs may be considered provocative and novel, or discursive artifacts. To this end, we assessed the mockups through the four dimensions of criticality suggested by Bardzell et al. [8]: *Changing perspectives, Enhancing appreciation, Proposals for change,* and *Reflectiveness*. While in the following we outline some aspects of criticality in this area, we acknowledge that these are not necessarily distinct categories and that there might be further relevant aspects of criticality. After all, Bardzell et al. [8] provide "support for, not a recipe for, judgment making."

The dimension of *Enhancing appreciation* (or judgment) is about making the user see the role of design in a socio-cultural issue of significance. We believe the dimension appears in all the designs, but most obviously in the Audience. With the design, we attempt to enhance the user's judgment on the present UI mechanisms for commenting and its possible role in uncivil and emotionally problematic commenting. The design is intended to underline how different text-based commenting is from public speaking and face-to-face discussion. Additionally, the fact that the Audience does not reveal details about who they are may remind the writer of the fact that one does not know the silent majority (i.e., the readers who do not reveal themselves in any manner through the discussion function). On the other hand, the reader may feel that the news publisher is judging the commentators because it has installed this system, and for a writer, the presence of the audience may feel like social pressure.

The dimension of *Proposals for change* [8], which is about proposing "an alternative way of being", also helps analyze the designs. With the Regret, we propose change by embodying a provocative proposal for a credible future, where the user is asked by a machine to publicly regret the overly negative wording that one used in a comment. The role of the design is to allow the user to quickly prevent fighting or calm down the readers of one's comment, which is unusual. Also, the label "username regretted their choice of words" may surprise the comment readers and cause them to question the writer's intentions or the truthfulness of the message. Next, the Audience and Creature may be read as assuming trust in an algorithm, or even obedience to it. The persuasiveness in avoiding the animal to suffer or the audience to frown if the user writes in a certain way may be understood to propose a future where people trust the interpretations of an algorithm and could act accordingly.

Additionally, there is a proposal for change in the sense of user-publisher power dynamics. In the Promise, the publisher would show the commenter that it is mightier than the commenter by forcing one to check an oversized checkbox and make a nearly impossible promise to control one's emotions. The other designs likewise may be understood to propose that the publisher has a strong voice in shaping the quality of discussion. The Regret is intended to present the publisher as something distant like a Catholic priest, as far as providing the context to express regret is akin to a confession booth. The Audience and Creature are also intended to show the publisher as having some form of an opinion about one's writing. The users, however, may not like the publisher taking this role, or at least not initially.

In our view, *Changing perspectives* helps to understand particularly Creature and Promise. In *Changing perspectives* "the design presents a framing or a point of view that is new, coherent, and interesting enough to help the user to perceive the particulars of a domain according to a new schema" [8]. The Creature is intended to present the concept of the wellbeing of an animal instrumentally as a tool for illustrating the emotional quality of text (i.e., change in designer's perspective on what one can use for this purpose). The design might also ask whether it is ethical to make users watch a virtual animal suffer because of emotionally troubling commenting. Especially when the creature is shown dead, pierced by arrows, is a concrete and provocative representation of the worst state. The morality may depend on whether the virtual creature is presented in an abstract or realistic form. In the mockup, the creature is cartoonish and abstract, which is likely less troubling. Furthermore, if the design is seen as cartoonish, it can be humorous. Finally, in the Promise, the size of a checkbox, a standard UI element, is intended to act as a signal of the publisher's power.

All in all, while this analysis provides several perspectives to how critique can manifest and inspire design in this problem area, this is not enough to judge which forms of critique are acceptable. For this reason, we conducted an interview study to bring additional viewpoints from domain experts.

5 INTERVIEW STUDY

5.1 Method and Participants

Ten Finnish news media experts (2 females, 8 males) were interviewed with a semi-structured interview procedure. Their domain expertise was expected to provide further insight into the risks and opportunities of the designs to users and media companies, hence contributing additional critical perspectives. The interviewees had experience in moderating online discussions in news media or were involved in developing solutions or policies for moderation and maintaining appropriate online discussion quality. Nine interviewees held executive positions in news media, such as digital development manager (participants P8 and P5), content manager (P4, P6, P9), or editor in chief (P1-P3, P7); in other words, the interviewees would likely be in key roles in the selection and deployment of future moderation systems in their organizations. One had recently moved to a company developing machine learning-based solutions for automated moderation. The range of experience in moderating or with discussion quality in online news media varied from 2 to 18 years, with the majority having experience of at least 10 years (P1, P3-P7). All the interviewees represented Finnish news media organizations. The gender imbalance of the interviewees is regrettable, and as more men were able to participate, it may reflect journalism being a gendered institution [62].

The interviews took place at the interviewees' workplaces and took from 50 minutes to 2 hours. The interviews were conducted by the third author. The first half of the interviews focused on, for example, moderation practices and ideals of online discussion quality. The selected designs were presented and discussed during the second half of the interviews. The interviews were audio-recorded and transcribed for analysis, and consent for participation and audio recording was asked at the beginning of the interview. This paper only covers the data related to the designs.

The interviewees were presented with the selected four mockups in a randomized order, along with a verbal explanation of the designs. The mockups were intentionally left unpolished because we wanted interviewees to feel free to share their ideas and opinions. They were then presented with an evaluation questionnaire with statements on the acceptability and effectiveness of the design (with seven Likert questions like "the solution improves the quality of commenting"), to provoke reflection and taking different perspectives. In other words, the questionnaire was used to support interviewing rather than as data collection per se. More importantly, the interviewees were asked to think aloud their reasoning and thoughts on the design and were asked follow-up questions to reach a deeper understanding of the reasoning behind the evaluation and on the thoughts on the design. The questions covered themes like first impressions on the design idea, possible effects on emotional reflectivity and online behavior, why the idea might or might not work.

5.2 Analysis

All qualitative coding and analysis were conducted by the first author, with iterative feedback on the coding and analysis from a colleague. The analysis followed a bottom-up approach. First, the transcriptions were read line-by-line and descriptive open coding was used to identify themes. Then, common themes across the data were identified and abstracted.

5.3 Findings: Additional Perspectives of Critique and Acceptability

In the findings, we analyzed how the expert interviewees' comments relate to the above-mentioned ways the designs might facilitate emotional reflection. We report how some of the designs were regarded as too distracting or shocking to facilitate behavior change. We also report the participants' considerations of expected effectiveness—whether the designs might *support* or *prevent* increased self-awareness and whether they might lead to improved discussion quality. The findings hence *complemented* the prior analysis of criticality of the design concepts. Therefore, we focused on the critical comments and omitted comments that overlapped with our own analysis or that relate to technical concerns, such as accuracy of text classification.

5.3.1 Shifting Users to a More Self-Aware Stance

The participants believed that the Audience design can facilitate a shift to a more reflective stance among people writing comments, but they were not sure about its acceptability. P8 expected that the user would start to reflect on their writing as they see the faces in the Audience design and found that a positive effect. P4 and P5 expected the design to be useful for certain kinds of users or in some hand-picked news articles. However, the design could also evoke anxiety. P3 foresaw that the feedback given by the audience could be made so visually impressive or invasive that it limits what the user dares to write. Moreover, P6 expected the audience would evoke anxiety for some users, making them think “this is how liked or hated I am.”

The participants likewise believed that Creature could facilitate a user’s shift to a more self-aware stance, but many of the participants also considered it too distressing or distracting. P8 thought that the idea is mostly the same as if a reporter intervened, except that “nobody could get angry at the dying virtual dog [laughter].” However, P4 thought the design would steal the user’s attention and make one forget what one was about to write. While P4, P6, and P8 did not consider the Creature design too distressing, many others did. P1, P5, P7, and P10 expressed that the concept of animal suffering is too cruel. P5 explained that the publisher could not in any circumstances use the concept of animal suffering to guide users. This is probably relevant especially on public sites with a broad spectrum of users. In addition, the concept caused P6 to laugh, after which s/he pointed out that it would not suit a news site but would work as a media education tool for children.

The notification in Regret was said to probably annoy users and be seen as unnatural but also to facilitate reflection. P1, P3, P6, and P8 pointed out that the notification “someone might write an angry reply to your comment” would annoy most users, and angry writers would not press the regret button. P1 stressed that the Regret would cause an angry user to think “What the ****?! I will not regret it!” In other words, P1 expected the behavioral effect to be the opposite of what we intended. Yet only a little later P1 contradicted themselves and said the design could slow down the hasty users.

Promise was seen by some participants to facilitate reflection but its more traditional UI features were expected to also cause many users simply to disregard it. P6 thought the design might be effective but also that adding a checkbox might annoy users. P9 said that the well-behaving commenters would feel annoyed and wonder why they see the intervention. P1 said the solution would drive users away, because “commenting should be as easy as possible” (P1). This comment underlines the value of free speech and frictionless participation. Furthermore, P3 commented on the checkbox: "I bet that most would just check it [without thinking]."

In sum, regarding the acceptability of shifting users to a more self-aware stance, Creature appears to have gone overboard with the concept of animal death and Promise appears to be too similar to existing designs to cause the shift. As for the remaining designs, it is hard to say whether Regret or Audience would be more acceptable in this regard. Furthermore, the findings on these intentionally provocative artifacts help to better understand what is proportional in the context. This might help to apply Acquisti et al.’s [1] guidance on nudging in follow-up research: “the direction, salience, and firmness of a nudge should be proportional to the user’s benefit from the suggested course of action”.

5.3.2 The Impact on Users’ Freedom and Agency

The comments in this subsection are about what the user would come to know or realize (if one reflects), what the users would do with the design, and whether the design might be misleading.

Worries that the design will limit the user’s freedom and agency (freedom of speech and freedom of opinion) were a common theme in the participants’ comments. In Audience, P1, P3, and P7–P9 feared that predicting the reactions that a comment will elicit in the audience would be considered a manipulation attempt. To exemplify, P1 said: “Someone might feel that this crowd is trying to create social pressure and that you cannot have this or that opinion. This is important [to understand]. I fear that it could be interpreted as a manipulation attempt.”

In Creature*,* the manipulation fears were mostly connected to the use of animal suffering as a tool, but also other aspects were brought out. For example, P5 said the design could give a false image that the publisher wants to flatten the conversation. We interpret that s/he said this because the design comprises one animal figure that has one emotional state at a time. However, referring to the creature becoming happy when the user writes well, P8 said it is very smart to use rewards instead of punishment to change the user's behavior. P8 further explained that using punishments will only cause a backlash and pointed out that the Creature and Promise work through positivity, while Regret represents a negative perspective.

In Promise, the fears of limiting the user's freedom or agency were centered on the text ("I promise…"). P1–P6 and P9 hinted that the text is patronizing or asking too much and must be changed. To exemplify, P2 said, "to promise that I control my emotional state is a patronizing starting point" and P6 said, "I shy away from the idea that we would only allow neutral and positive [writing]." P1 said the text should tell if breaking the promise prevents publishing; otherwise, the design would not work. However, after the interviewer explained that the wording could be changed, the design was seen to less limit freedom or agency. P7 and P8 considered that asking whether the user has done wrong is not directing or limiting their writing.

In Regret, P3 and P6 feared the user's freedom or agency would be limited because the user is only provided the option to regret, not to edit or remove their comment. P3 ironically pointed out that if there is just the regret button, it can make the discussion board look like a regret-board, and the welcoming message would be "welcome to regret on our forum." P5, however, said the design does not imply that the publisher is directing the users, like some other designs, but that it provides tools to improve the discussion.

In sum, regarding the acceptability of the critique in relation to user's freedom. Promise now appears to not only be too similar to existing designs, but also to distress users who would not skip it (and they are probably the better behaving users). Also, the experts' comments on Regret help to highlight that adding an edit option beside the regret option could increase the acceptability of critique in Regret. Next, Audience was judged more harshly in comparison, as the core idea of using a virtual audience was connected to the concept of manipulation.

5.3.3 Risks of Discouragement and Abuse

A recurrent theme in the interviews was that some of the designs could invite abuse or discourage some forms of positive commenting. The following complements well our analysis of criticality in terms of how users can appropriate the designs. What is told here significantly decreases how acceptable we view the critique in the designs to be.

Audience was generally seen to make the user more aware of the other people, but it was also brought out that realistic prediction of other users' reactions could lead to self-censorship. While, for example, P9 underlined the increased sense of audience with "This brings out that there are other people and not just the writer." The related risk of discouraging the act of commenting was also brought out. For example, P8 commented that showing how different users might think about one's comment could make the user worry about posting a critical comment or going against the opinions of the majority, hence increasing self-censorship. The human figures, even abstract ones, were considered central causes for such worries. In addition, P7, P8, and P9 feared that the audience could cause the users to regard commenting as a people-pleasing exercise, where the users try to follow a norm set by the system. This concern resonates with the risk of "infantilization" mentioned in literature on nudging [1, 12]: individuals may come to rely on nudges for guidance and become unable to make decisions on their own. Having said that, such behavior would require very detailed modeling of the text and certain unanimity in the audience's expressions.

Furthermore, the participants saw that three of the designs could produce the opposite behavioral effect in the case of problem users. Trolls and other users with questionable intent could abuse Audience, Creature, and Regret. For example, P4 thought some users could write comments with the purpose of making the Audience show expressions that they want to the other users. Moreover, P6 thought that some would use the audience as a guide to writing as offensively as possible. In addition, P4, P5, and P10 thought some users would use CREATURE as a guide to say as hurtful things as possible to see how the animation progresses. P10 thought some users would intentionally make the creature suffer. In REGRET, however, the unwanted use could be less like hate-speech and more like wreaking havoc. P1 and P3 feared that REGRET could be used excessively or ironically. Furthermore, P5 pointed out that the AUDIENCE, when shown to all readers, could, for example, reveal that there are people that "are joyful that a refugee boy has drowned in the ocean", and told us the publisher would not like that to be highlighted.

6 DISCUSSION

The following summarizes the main observations and insights into the designs, and what the designs revealed, and reflects on the research process and its implications for the development of Design. Additionally, we draw some preliminary design considerations that might help to mitigate some of the issues in this problem area.

6.1 Insights into the Designs and their Effectiveness

This study provided a multitude of perspectives to how critique could be manifested in this problem area and offers valuable insights into the social acceptance and possible effects of the designs. While the paper only presented four designs out of a broad array of alternatives and the opinions of a limited sample of experts, we argue to have managed to both unpack relevant problems and outline the solution space in this challenging area. The mockups seemed to both provide the participants with *substance* to reflect on and provoke the very *act* of reflection through nonobvious features. As a result, we argue having found insights that would not have resulted from a process with designs that follow present-day design conventions or try to optimize for effectiveness, social acceptance, or any single design quality. Furthermore, the participants seemed to have a relatively high degree of uncertainty about the general acceptability of the designs as possible solutions. This suggests that they did not expect the designs to be dismissed as art or seen as conventional UI design.

While the design exploration and the findings of the qualitative research can offer various insights for different readers, the following highlights and discusses insights that we, in retrospect, found particularly interesting and educational.

6.1.1 Highlighting the Positive Instead of Removing the Negative

A key insight is that while critique in design might sound negative, and negative connotations are easily attached to the problem of online discussion quality, designs exhibiting a critical voice could also focus on positive perspectives. For example, discomfort with the status quo could be displayed in positive or fun ways. P8 considered that CREATURE had a positive dimension in potentially rewarding the good writer. Furthermore, highlighting the positive may be interpreted to follow the logic of care [61] somewhat more than highlighting the negative.

6.1.2 Humanlike Qualities Might Imply Rich Judgment Capabilities

The AUDIENCE expects human-like interpretation capabilities from the system. Based on the participants' comments, a measurement-oriented design that features humanlike figures would imply that it reacts to nuanced opinions and subtext, not just to the use of certain words. If the audience did not react to such implicit semantics, the design would violate users' expectations of human-like behavior, which resonates with literature on social robotics and anthropomorphism [23, 44]. Also, the use of humanlike figures leads naturally to discussion on who decides what too offensive opinions are and are not. The participants were concerned that if the publisher defined the standards on offensiveness, it might lead to a violation of users' right to freedom of opinion.

6.1.3 The Risks in Accurately Predicting What the Majority Thinks

Based on the comments on AUDIENCE, a system that accurately predicts the majority's reaction may discourage diverse and civil discussion. It could effectively argue to the user that one should never say anything that the majority does not like. This reminds us of concerns that an algorithm could silently suppress the voices of minorities [19, 47]. Furthermore, this problem would likely be magnified if anything like faces were used to predict the audience's reactions, as the faces could be considered as the opponents' faces. Therefore, AUDIENCE should probably rather feature an audience of professional judges than an audience of other users of the same service and avoid giving the impression that a majority of opinions exists. That said, featuring an audience of professional judges might feel patronizing to users.

6.1.4 Showing What is Uncivil Can Support Trolling

One aspect that our design process failed to recognize is that trolls might abuse especially the AUDIENCE, the CREATURE, and the REGRET. Therefore, these designs could only be applied in limited contexts and under careful supervision or be supported by other mechanisms. As the participants had less such concerns about PROMISE, it gives us a hint on how to solve the problem of the potential for abuse. One approach could be to notify the user when they begin to write in an emotionally problematic manner but not keep modeling and visualizing how badly one does so, as this would help to optimize the text for malicious purposes. In future work, we plan to explore different solutions to this problem.

6.1.5 Intervening in Commenting Increases the Social Calculation that Users Have to Do

The interviews allowed us to better understand that none of the designs are exclusively for comment readers or writers. If an intervention of any type is targeted towards published comments and/or readers, it can also affect the considerations of comment writers. The writers may avoid or seek to be the target of the publisher's, the system's, and/or other users' attention. At the same time, the reader may dislike the comments being forced towards what the publisher considers as "good." This leads to the realization that by influencing multiple user groups at once, UI designs discussed in this paper may require *social calculation* from the active users.

6.2 Reflection on the News Media Experts' Values and Generalizability of the Findings

In the previous we focused on the designs; here we focus on the experts and the cultural context of the interview study. We highlight the following three issues from the findings: freedom of expression, moderation, and publisher's reputation.

The interviews reveal an ambivalent position toward the designs: the Finnish news media experts wish to prevent trolling and uncivil discussion, however not wanting to limit the commenters' freedom of expression. This may be partly explained by the experts we interviewed being first and foremost journalists, whose basic values include freedom of expression. According to literature, journalists tend to have an ambivalent position toward uncivil commenting in general [18, 46, 73, 74]. However, the experts also seemed, understandably, to value and guard the publisher's reputation. They seemed concerned the presented solutions could lead to the publisher being questioned in public discussion. This suggests that the publishers, in their view, are not known for experimenting with solutions. The publishers' wish to preserve journalistic brand and general conservativism may hence be an obstacle for identifying solutions that truly question the conventions.

Finally, while we suspect that the Finnish context of journalism is not significantly different compared to other western countries, the sample size naturally limits the generalizability of the findings. Freedom of the press in Finland is among the highest in the world [58], and relative to the size of its population, the number of visits to news websites owned by news media is exceptionally high [59]. Also, participation in politics and voting is rather high in Finland [69]. These factors may increase the diversity of opinions about news, the perceived importance of news sites' comment sections, and the importance of a low threshold for participation.

6.3 Reflection on the Research Process and its Implications

We argue that the design process and the interview study helped in defining relevant criteria and perspectives to consider when aiming to address this interdisciplinary and difficult problem area. The design approach was beneficial for further framing the problem area and for understanding the opportunities and pitfalls in designing solutions from different perspectives. Also, we argue the work contributes to the growing critical design literature space.

We identify two positive outcomes for using design artifacts exhibiting a critical voice—as opposed to more conventional artifacts: radical innovation and potential for change of behavior. Based on this exploration, we agree with Bowen [13] who argued that critical artifacts can be used to foster innovation by "provoking stakeholders and designers to consider alternative possibilities for products (etc.)." Bowen termed the instrumental use of critical artifacts to foster innovation as Critical Artefact Methodology. We argue that the artifacts we have created may similarly foster innovation. For example, the artifacts may help designers to consider possibilities for machine learning-based systems, to identify new ways and techniques to educate users and to take alternative perspectives to interaction design and user-publisher relationships. More importantly, we did not only use design artifacts to foster innovation by the stakeholders involved in the design process, but we also sought designs that could make the vast spectrum of users commenting on online news to think critically and reflect on their emotional processes. The interview study leaves us thinking that targeting critique directly to the end-users might persuade them to write with a better tone and consider limitations of the digital media as a channel for communication. Furthermore, we believe this might cause the designs to provoke more discussion among the public and to increase the public's ability to perceive the (dis)value of UI designs for online discussion and to better discriminate between them.

Further, taking a step back and looking at the paper instead of the artifacts, we argue that this research-through-design case can make critical design more accessible for HCI researchers and designers. Tharp & Tharp [68] write in recent work that it can be difficult for designers to find literature that explicitly helps them do discursive design work. Our description of the iterative idea generation, filtering, and selection process, as well as the analysis of the critical elements, may help in this regard. This study exemplifies that analysis of criticality according to Bardzell et al.'s framework [8] is beneficial for articulating the designs and the designer's thinking, and for consideration of different perspectives to the problem at hand.

6.4 Limitations and Future Work

This design exploration has limitations concerning, for example, the depth of analysis and certainty of the findings. The number and types of critical perspectives we could identify to this particular problem area was narrowed by our choice to focus on emotion regulation; taking other such fundamental perspectives remains an opportunity for further design exploration. The identified critical perspectives were also limited by the breadth and depth of the interview study. For example, we could have more often asked the interviewees to support their reflections on the concepts with more detailed, situated examples stemming from their personal experiences. This combined with the future-oriented and speculative nature of the discussion forced the comments on the designs to remain on a rather speculative level. Further, the breadth of the interview study was limited because of not including people who use online news commenting platforms. Finally, this exploration focused on provocative, yet tentative design concepts, rather than technical detail: each design contained only one example of how a user might interact with them. This limited the level of detail of our discussion on the acceptability of the design features.

Having said that, the limitations introduce relevant needs and opportunities for future work. Increasing the fidelity of the designs and creating detailed scenarios could open new perspectives to the design. Evaluating the designs with more participants, including people who use online news commenting platforms, and creating different versions of the designs to allow better comparison are opportunities for more extensive empirical studies. Such evaluation studies can help better answer if the designs or certain features in them would support emotion regulation or reflection to the extent that the quality of discussion would improve measurably.

Evaluation studies could also help understand if the tactics for content moderation inspired by the theory of implicit emotion regulation [70] are better than other tactics, and if this approach can realistically complement the dichotomy of the logic of choice and logic of care [61]. In other words, more extensive evaluation studies would help to understand whether any of the four designs offer appropriate solutions to the given problem in any news or user-group contexts. In addition, comparing the benefits and disadvantages of focusing to design with a critical voice compared to a more user-centered design process in this context remains a future opportunity.

7 CONCLUSION

Maintaining the quality of discussion is regarded as a persistent challenge in online digital media. This paper reported a research-through-design case study, aiming to challenge the status quo of UI design in the context of online news commenting. The study identified a broad array of critical perspectives to this problem area and designs of unconventional solutions that might nudge emotional reflection. The critical perspectives stemmed from a design process that brought together creative design work with a critical voice, theories on emotion regulation, and reflection on existing design conventions, as well as an interview study with domain experts. Four designs exhibiting a critical voice were devised in an attempt to reach a good balance between provocativeness, novelty, social acceptance, and expected effectiveness. The designs build on ideas of, e.g., improving reflection on the characteristics and feelings of possible readers of comments, and the use of social conventions like regretting and promising in UI design.

All in all, this work contributes possible solutions—or at least ways of thinking about better solutions—for improving online discussion. Our analysis and the interview findings offer qualitative insights into the design artifacts—such that probably would not have resulted from conventional designs. We argue that identifying various critical perspectives, unpacking the problem area, and outlining the solution space are necessary steps towards the creation of sustainable UI solutions in this problem area. From a methodological viewpoint, we contributed a case study on using a critical voice in design for not only provoking new questions but also addressing wicked problems through UI design.

ACKNOWLEDGMENTS

We thank all the participants of the interview study. The research was funded by the Academy of Finland (grants 320767, 320766). We thank Saara Ala-Luopa for her valuable comments on the manuscript.

REFERENCES

[1] Alessandro Acquisti, Idris Adjerid, Rebecca Balebako, Laura Brandimarte, Lorrie Faith Cranor, Saranga Komanduri, Pedro Giovanni Leon, Norman Sadeh, Florian Schaub, Manya Sleeper, Yang Wang, and Shomir Wilson. 2017. Nudges for privacy and security: Understanding and assisting users' choices online. *ACM Comput. Surv.* 50, 3 (August 2017), 1–41. DOI:https://doi.org/10.1145/3054926

[2] Philip E. Agre. 1997. *Computation and human experience*. Cambridge University Press.

[3] Alexa. 2019. Alexa - Top Sites in Finland. Retrieved April 28, 2020 from https://www.alexa.com/topsites/countries/FI

[4] Christie Aschwanden. 2016. We Asked 8,500 Internet Commenters Why They Do What They Do. Retrieved October 17, 2019 from https://fivethirtyeight.com/features/we-asked-8500-internet-commenters-why-they-do-what-they-do/

[5] Jae Eul Bae, Youn Kyung Lim, Jin Bae Bang, and Myung Suk Kim. 2015. Pause moment experience in SNS communication. In CHI '15: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, 2113–2116. DOI:https://doi.org/10.1145/2702123.2702435

[6] Jeffrey Bardzell and Shaowen Bardzell. 2013. What is critical about critical design? In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13, ACM Press, New York, New York, USA, 3297. DOI:https://doi.org/10.1145/2470654.2466451

[7] Jeffrey Bardzell, Shaowen Bardzell, and Mark A. Blythe. 2018. Critical Theory and Interaction Design. The MIT Press.

[8] Jeffrey Bardzell, Shaowen Bardzell, and Erik Stolterman. 2014. Reading critical designs. In Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14. DOI:https://doi.org/10.1145/2556288.2557137

[9] Shaowen Bardzell, Jeffrey Bardzell, Jodi Forlizzi, John Zimmerman, and John Antanitis. 2012. Critical design and critical theory: The challenge of designing for provocation. In Proceedings of the Designing Interactive Systems Conference, DIS '12. DOI:https://doi.org/10.1145/2317956.2318001

[10] Julian Bleecker. 2009. Design Fiction: A Short Essay on Design, Science, Fact and Fiction. Retrieved January 9, 2020 from

http://drbfw5wfjlxon.cloudfront.net/writing/DesignFiction_WebEdition.pdf

[11] Mark Blythe, Enrique Encinas, Jofish Kaye, Miriam Lueck Avery, Rob McCabe, and Kristina Andersen. 2018. Imaginary design workbooks: Constructive criticism and practical provocations. In CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, New York, USA, 1–12. DOI:https://doi.org/10.1145/3173574.3173807

[12] Luc Bovens. 2009. The Ethics of Nudge. In Preference Change. Springer Netherlands, 207–219. DOI:https://doi.org/10.1007/978-90-481-2593-7_10

[13] Simon John Bowen. 2009. A critical artefact methodology : using provocative conceptual designs to foster human-centred innovation. Doctoral thesis, Sheffield Hallam University. Retrieved March 6, 2019 from http://shura.shu.ac.uk/3216/

[14] Richard Buchanan. 1992. Wicked Problems in Design Thinking. Des. Issues 8, 2 (1992), 5–21.

[15] Simon Buckingham Shum. 2008. Cohere: Towards Web 2.0 Argumentation. In Proc. COMMA'08: 2nd International Conference on Computational Models of Argument, 97–108. DOI:https://doi.org/10.5860/choice.51-2973

[16] Ana Caraban, Evangelos Karapanos, Daniel Gonçalves, and Pedro Campos. 2019. 23 Ways to Nudge. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19, ACM Press, New York, New York, USA, 1–15. DOI:https://doi.org/10.1145/3290605.3300733

[17] Gina Masullo Chen and Yee Man Margaret Ng. 2017. Nasty online comments anger you more than me, but nice ones make me as happy as you. Comput. Human Behav. 71, (June 2017), 181–188. DOI:https://doi.org/10.1016/j.chb.2017.02.010

[18] Gina Masullo Chen and Paromita Pain. 2017. Normalizing Online Comments. Journal. Pract. 11, 7 (August 2017), 876–892. DOI:https://doi.org/10.1080/17512786.2016.1205954

[19] Thomas Davidson, Debasmita Bhattacharya, and Ingmar Weber. 2019. Racial Bias in Hate Speech and Abusive Language Detection Datasets. In In: Proceedings of the third workshop on abusive language online, Florence; 2019., Association for Computational Linguistics (ACL), 25–35. DOI:https://doi.org/10.18653/v1/w19-3504

[20] Thomas Davidson, Dana Warmsley, Michael Macy, and Ingmar Weber. 2017. Automated Hate Speech Detection and the Problem of Offensive Language. In Proceedings of the 11th International Conference on Web and Social Media, ICWSM 2017, AAAI Press, 512–515. Retrieved December 20, 2020 from http://arxiv.org/abs/1703.04009

[21] Pablo Delgado. 2019. How El País used Perspective API to make their comments section less toxic. Retrieved December 17, 2019 from https://www.blog.google/outreach-initiatives/google-news-initiative/how-el-pais-used-ai-make-their-comments-section-less-toxic/

[22] Nicholas Diakopoulos and Mor Naaman. 2011. Towards Quality Discourse in Online News Comments Human Factors. In Proceedings of the ACM 2011 conference on Computer supported cooperative work 2011 Mar 19, 133–142. Retrieved June 19, 2019 from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.188.3516&rep=rep1&type=pdf

[23] Brian R. Duffy. 2003. Anthropomorphism and the social robot. In Robotics and Autonomous Systems, 177–190. DOI:https://doi.org/10.1016/S0921-8890(02)00374-3

[24] Anthony Dunne and Fiona Raby. Dunne & Raby Projects. Retrieved January 31, 2020 from http://dunneandraby.co.uk/content/projects

[25] Anthony Dunne and Fiona Raby. 2001. Design Noir: The Secret Life of Electronic Objects. Birkhäuser.

[26] eBizMBA. 2019. Top 15 Most Popular News Websites | September 2019. Retrieved October 17, 2019 from http://www.ebizmba.com/articles/news-websites

[27] Nancy Fraser. 1990. Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy. Soc. Text 25/26 (1990), 56. DOI:https://doi.org/10.2307/466240

[28] Batya Friedman. 1996. Value-sensitive design. Interactions 3, 6 (December 1996), 16–23. DOI:https://doi.org/10.1145/242485.242493

[29] Becky Gardiner, Mahana Mansfield, Ian Anderson, Josh Holder, Daan Louter, and Monica Ulmanu. 2016. The dark side of Guardian comments | Technology | The Guardian. Retrieved October 16, 2019 from https://www.theguardian.com/technology/2016/apr/12/the-dark-side-of-guardian-comments?

[30] William W. Gaver, John Bowers, Andrew Boucher, Hans Gellerson, Sarah Pennington, Albrecht Schmidt, Anthony Steed, Nicholas Villars, and Brendan Walker. 2004. The drift table. In Extended abstracts of the 2004 conference on Human factors and computing systems - CHI '04, ACM Press, New York, New York, USA, 885. DOI:https://doi.org/10.1145/985921.985947

[31] Tarleton Gillespie. 2018. Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media. Yale University Press.

[32] Erving Goffmann. 1956. The Presentation of Self in Everyday Life. Doubleday.

[33] Kirsikka Grön and Matti Nelimarkka. 2020. Party Politics, Values and the Design of Social Media Services. Proc. ACM Human-Computer Interact. 4, CSCW2 (2020). DOI:https://doi.org/10.1145/3415175

[34] James J. Gross. 1998. The Emerging Field of Emotion Regulation: An Integrative Review. Rev. Gen. Psychol. 2, 3 (September 1998), 271–299. DOI:https://doi.org/10.1037/1089-2680.2.3.271

[35] Ståle Grut. 2017. With a quiz to comment, readers test their article comprehension. Retrieved January 20, 2020 from https://nrkbeta.no/2017/08/10/with-a-quiz-to-comment-readers-test-their-article-comprehension/

[36] Amos Guiora and Elizabeth A. Park. 2017. Hate Speech on Social Media. Philosophia (Mendoza). 45, 3 (September 2017), 957–971. DOI:https://doi.org/10.1007/s11406-017-9858-4

[37] Jürgen Habermas. 1991. The Structural Transformation of the Public Sphere: An inquiry into a category of bourgeois society. The MIT Press, Cambridge.

[38] Netta Iivari and Kari Kuutti. 2017. Critical Design Research and Information Technology. 983–993. DOI:https://doi.org/10.1145/3064663.3064747

[39] Netta Iivari and Kari Kuutti. 2017. Towards Critical Design Science Research. In International Conference On Information (ICIS) 2017. Association For Information Systems. Retrieved August 19, 2019 from http://aisel.aisnet.org/icis2017/ResearchMethods/Presentations/10

[40] Jigsaw. 2017. Perspective API. Retrieved October 16, 2019 from https://www.perspectiveapi.com/#/home

[41] Leon Karlsen Johannessen. 2017. The Young Designer's Guide to Speculative and Critical Design. Norwegian University of Science and Technology.

[42] Vera Khovanskaya, Eric P S Baumer, Dan Cosley, Stephen Voida, and Geri Gay. 2013. “Everybody Knows What You’re Doing”: A Critical Design Approach to Personal Informatics. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’13).

[43] Travis Kriplean, Jonathan Morgan, Deen Freelon, Alan Borning, and Lance Bennett. 2012. Supporting reflective public thought with considerit. In Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work - CSCW ’12. DOI:https://doi.org/10.1145/2145204.2145249

[44] Minae Kwon, Malte F. Jung, and Ross A. Knepper. 2016. Human expectations of social robots. In HRI ’16: The Eleventh ACM/IEEE International Conference on Human Robot Interaction, IEEE Computer Society, 463–464. DOI:https://doi.org/10.1109/HRI.2016.7451807

[45] Geoffrey C. Layman, Thomas M. Carsey, and Juliana Menasce Horowitz. 2006. PARTY POLARIZATION IN AMERICAN POLITICS: Characteristics, Causes, and Consequences. Annu. Rev. Polit. Sci. 9, 1 (June 2006), 83–110. DOI:https://doi.org/10.1146/annurev.polisci.9.070204.105138

[46] Anders Sundnes Løvlie, Karoline Andrea Ihlebæk, and Anders Olof Larsson. 2018. User Experiences with Editorial Control in Online Newspaper Comment Fields. Journal. Pract. 12, 3 (March 2018), 362–381. DOI:https://doi.org/10.1080/17512786.2017.1293490

[47] Donna Lu. 2019. Google’s hate speech AI may be racially biased. New Sci. 243, 3243 (August 2019), 7. DOI:https://doi.org/10.1016/s0262-4079(19)31505-2

[48] Matt Malpass. 2013. Between Wit and Reason: Defining Associative, Speculative, and Critical Design in Practice. Des. Cult. 5, 3 (2013), 333–356. DOI:https://doi.org/10.2752/175470813X13705953612200

[49] Matthew Malpass. 2012. Contextualising Critical Design: Towards a Taxonomy of Critical Practice in Product Design. Doctoral thesis, Nottingham Trent University. Retrieved March 6, 2019 from https://core.ac.uk/download/pdf/18419931.pdf

[50] Courtney Napoles, Joel Tetreault, Aasish Pappu, Enrica Rosato, and Brian Provenzale. 2017. Finding Good Conversations Online: The Yahoo News Annotated Comments Corpus. (2017), 13–23. DOI:https://doi.org/10.18653/v1/w17-0802

[51] Adewale Obadimu, Esther Mead, Muhammad Nihal Hussain, and Nitin Agarwal. 2019. Identifying toxicity within youtube video comment. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Verlag, 214–223. DOI:https://doi.org/10.1007/978-3-030-21741-9_22

[52] Thomas Olsson, Pradthana Jarusriboonchai, Paweł Woźniak, Susanna Paasovaara, Kaisa Väänänen, and Andrés Lucero. 2019. Technologies for Enhancing Collocated Social Interaction: Review of Design Solutions and Approaches. Comput. Support. Coop. Work CSCW An Int. J. (2019). DOI:https://doi.org/10.1007/s10606-019-09345-0

[53] James Pierce. 2015. Working by Not Quite Working : Resistance as a Technique for Alternative and Oppositional Designs. Doctoral thesis, Carnegie Mellon University. Retrieved from http://reports-archive.adm.cs.cmu.edu/anon/hcii/CMU-HCII-15-109.pdf

[54] James Pierce, Phoebe Sengers, Tad Hirsch, Tom Jenkins, William Gaver, and Carl DiSalvo. 2015. Expanding and Refining Design and Criticality in HCI. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI ’15, 2083–2092. DOI:https://doi.org/10.1145/2702123.2702438

[55] Fabian Prochazka, Patrick Weber, and Wolfgang Schweiger. 2018. Effects of Civility and Reasoning in User Comments on Perceived Journalistic Quality. Journal. Stud. 19, 1 (January 2018), 62–78. DOI:https://doi.org/10.1080/1461670X.2016.1161497

[56] Milena Radzikowska. 2015. Looking for Betsy: A Critical Theory Approach to Visibility and Pluralism in Design. Doctoral thesis, University of Alberta.

[57] Dimitrios Raptis, Rikke Hagensby Jensen, Jesper Kjeldskov, and Mikael B. Skov. 2017. Aesthetic, Functional and Conceptual Provocation in Research Through Design. In Proceedings of the 2017 Conference on Designing Interactive Systems - DIS ’17, 29–41. DOI:https://doi.org/10.1145/3064663.3064739

[58] Reporters Without Borders. 2020. 2020 World Press Freedom Index. Retrieved September 10, 2020 from https://rsf.org/en/ranking

[59] Esa Reunanen. 2020. Uutismedia verkossa 2020. Reuters-instituutin Digital News Report - Suomen maaraportti. Tampere. Retrieved August 31, 2020 from http://urn.fi/URN:ISBN:978-952-03-1610-5

[60] Ian Rowe. 2015. Civility 2.0: a comparative analysis of incivility in online political discussion. Inf. Commun. Soc. 18, 2 (February 2015), 121–138. DOI:https://doi.org/10.1080/1369118X.2014.940365

[61] Minna Ruckenstein and Linda Lisa Maria Turunen. 2019. Re-humanizing the platform: Content moderators and the logic of care. New Media Soc. (September 2019), 146144481987599. DOI:https://doi.org/10.1177/1461444819875990

[62] Iiris Ruoho and Sinikka Torkkola. 2018. Toward a Multidimensional Approach. 39, (2018), 67–79. DOI:https://doi.org/10.2478/nor-2018-0002.67

[63] Joseph Seering, Tianmi Fang, Luca Damasco, Mianhong Cherie Chen, Likang Sun, and Geoff Kaufman. 2019. Designing user interface elements to improve the quality and civility of discourse in online commenting behaviors. Conf. Hum. Factors Comput. Syst. - Proc. (2019), 1–14. DOI:https://doi.org/10.1145/3290605.3300836

[64] Phoebe Sengers, Kirsten Boehner, Shay David, and Joseph “Jofish” Kaye. 2005. Reflective design. In Proceedings of the 4th decennial conference on Critical computing between sense and sensibility - CC ’05, ACM Press, New York, New York, USA, 49. DOI:https://doi.org/10.1145/1094562.1094569

[65] Aaron Smale. 2016. Rhetorical Strategies: Logos, Ethos, Pathos, Kairos | University of Nevada, Reno – UWSC. Retrieved October 17, 2019 from https://writingcenter.blogs.unr.edu/2016/10/12/rhetorical-strategies-logos-ethos-pathos-kairos/

[66] Nina Springer, Ines Engelmann, and Christian Pfaffinger. 2015. User comments: motives and inhibitors to write and read. Information, Commun. Soc. 18, 7 (July 2015), 798–815. DOI:https://doi.org/10.1080/1369118X.2014.997268

[67] Cass Sunstein and Richard Thaler. 2009. Nudge: Improving decisions about health, wealth, and happiness. Penguin.

[68] Bruce M. Tharp and Stephanie Tharp. 2019. Discursive design : critical, speculative, and alternative things. MIT Press.

[69] The Economist Intelligence Unit. 2019. Democracy Index 2019 - A Year of Democratic SetbacksS AND POPULAR PROTEST. London.

[70] Jared B. Torre and Matthew D. Lieberman. 2018. Putting Feelings Into Words: Affect Labeling as Implicit Emotion Regulation. Emot. Rev. 10, 2 (April 2018), 116–124. DOI:https://doi.org/10.1177/1754073917742706

[71] Joseph B. Walther. 1993. Impression development in computer-mediated interaction. West. J. Commun. 57, 4 (December 1993), 381–398. DOI:https://doi.org/10.1080/10570319309374463

[72] Joseph B. Walther. 1996. Computer-Mediated Communication. Communic. Res. 23, 1 (February 1996), 3–43. DOI:https://doi.org/10.1177/009365096023001001

[73] J. David Wolfgang. 2018. Cleaning up the “Fetid Swamp.” Digit. Journal. 6, 1 (January 2018), 21–40. DOI:https://doi.org/10.1080/21670811.2017.1343090

[74] J David Wolfgang. 2018. Taming the ‘trolls’: How journalists negotiate the boundaries of journalism and online comments. Journalism (March 2018), 146488491876236. DOI:https://doi.org/10.1177/1464884918762362

[75] Gavin Wood, Kiel Long, Tom Feltwell, Scarlett Rowland, Phillip Brooker, Jamie Mahoney, John Vines, Julie Barnett, and Shaun Lawson. 2018. Rethinking engagement with online news through social and visual co-annotation. Conf. Hum. Factors Comput. Syst. - Proc. 2018-April, (2018), 1–12. DOI:https://doi.org/10.1145/3173574.3174150