

JOEL KISKOLA

Designing Critical and Practical User Interface Interventions in Uncivil Online Discussion

JOEL KISKOLA

Designing Critical and Practical
User Interface Interventions in
Uncivil Online Discussion

ACADEMIC DISSERTATION

To be presented, with the permission of
the Faculty of Information Technology and Communication Sciences
of Tampere University,
for public discussion in the auditorium B1096 (100)
of the Pinni building, Kanslerinrinne 1, Tampere,
on 8 December, at 12 o'clock.

ACADEMIC DISSERTATION

Tampere University, Faculty of Information Technology and Communication Sciences
Finland

*Responsible
supervisor
and Custos* Professor
Thomas Olsson
Tampere University
Finland

Supervisors University Lecturer
Dr. Anna Rantasila
Lappeenranta-Lahti University of
Technology
Finland

Dr. Aleksi Syrjämäki
Finland

Pre-examiners Professor
Netta Iivari
University of Oulu
Finland

Associate Professor
Evangelos Karapanos
Cyprus University of Technology
Cyprus

Opponent Professor
Jeffrey Bardzell
The Pennsylvania State University
The United States of America

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

Copyright ©2023 author

Cover design: Roihu Inc.

ISBN 978-952-03-3132-0 (print)

ISBN 978-952-03-3133-7 (pdf)

ISSN 2489-9860 (print)

ISSN 2490-0028 (pdf)

<http://urn.fi/URN:ISBN:978-952-03-3133-7>



Carbon dioxide emissions from printing Tampere University dissertations have been compensated.

PunaMusta Oy – Yliopistopaino
Joensuu 2023

ACKNOWLEDGEMENTS

One afternoon in early 2017 I was heading out of class, when a classmate, now PhD Candidate Wenyan Yang told me that a professor was looking for research assistants. Had Wenyan not told me this, I would likely never have become a researcher. I am similarly thankful for the people who told me I could become a PhD long before I became a researcher. They are my wife Hanna, and my friend Juhani Perälä. Their early support gave me courage to embark on this journey.

I am grateful to my primary Supervisor, Professor Thomas Olsson for having faith in me and hiring me to a project funded by the Academy of Finland, which enabled most of my PhD. You have guided me at every phase in completing my doctorate; you have patiently helped me to find solutions when I had too many ideas. I admire your way of instructing; you simultaneously give space and push for greatness. I am blessed to have had such a great supervisor and such a great opportunity to write my PhD.

I am grateful to my second Supervisors, Dr. Aleksi Syrjämäki and University Lecturer Dr. Anna Rantasila. Aleksi, I am particularly grateful for you for patiently helping me with the statistical analyzes. I have good memories of video-chatting with you about statistics-related questions. Anna, I am particularly grateful for you for helping me with analyzing participants opinions about my designs, I believe your guidance was crucial for the outcome. I am also grateful for Principal Lecturer Dr. Heli Väättäjä for having been my second supervisor for some time in the beginning of my journey.

Thank you, dear co-authors, for your support and our discussions: Professor Veikko Surakka, Dr. Mirja Ilves, Dr. Poika Isokoski, Dr. Heli Väättäjä.

Thank you, dear colleagues, in Olsson's TSI research group for your comments and all the time we have spent together: PhD Candidate Sami Koivunen, PhD Candidate Ekaterina Karjalainen, Saara Ala-Luopa, Dr. Zahra Hosseini, Amir Pakpour, and Rūta Šerpytytė. And extended TSI family: Dr. Jukka Huhtamäki and Dr. Jussi Okkonen.

Thanks for Finnish Foundation for Technology Promotion for the one-year grant for finishing my PhD. Thank you, Thomas, Aleksi, Anna, and Sami for helping me with the grant applications.

Lastly, I want to thank my families and relatives. Thank you, Hanna, my wife, for your love and support and enduring the time I have worked on this thesis. Thank you, Teija, mom, and Jarmo, dad, for your love and support. Thank you, Mauno, Sara, Iisa, Aamu, Pirjo, Antti, and Jaana for always being so supportive of me.

ABSTRACT

Millions of people use online discussion systems daily to share their opinions, knowledge, and feelings with others. At the same time, constructive online discussions can be hindered by both uncivil conduct and efforts to suppress it. This implies that interventions in uncivil discussion should be designed carefully.

This thesis supports designing for critical and practical functionality in user interface (UI) mechanisms intended to intervene in uncivil online discussion. Critical functionality refers to provoking reflection, discussion, imagination, appreciation of complexity, and consideration of new perspectives. This is useful for problem finding and nudging people to act in different ways. Practical functionality, on the other hand, refers to problem solving, here, mitigating incivility.

The thesis approaches the issue by conducting Critical Design of affect labeling UI interventions in uncivil online discussion, particularly focusing on news commenting. Critical Design refers to creating designs that may not be solutions but rather provoke discussion about the ethics of design, reveal potentially hidden agendas and values, and propose alternative design values. Affect labeling is seen as a promising strategy where the emotions present in the content are explicated, which, based on psychological emotion theories, is expected to have a calming effect.

The research resulted in four scientific publications. The research included ten interviews with Finnish journalists about four critical affect labeling UI intervention design proposals, as well as two international online surveys in which a total of 687 online news commenters reacted to a total of 14 intervention designs.

The thesis contributes relevant design aspects and dimensions for Critical Design. For example, the “Practical—Critical” dimension, which refers to whether the intervention should encourage a simple change in behavior or trigger a deeply reflective response to the intervention design. It also contributes knowledge on the user-perceived characteristics of high-quality. For example, online news commenters believe a high-quality UI intervention helps them to avoid reading uncivil comments and posting comments that they regret.

Overall, the thesis advances design of UI interventions in uncivil online discussion with a critical voice and contributes more generally to literature on design of online discussion systems.

TIIVISTELMÄ

Miljoonat ihmiset käyttävät päivittäin verkkokeskustelujärjestelmiä jakaakseen mielipiteitään, tietoaan ja tunteitaan. Samanaikaisesti, sekä epäasiallinen käytös että sen estämiseen tähtäävät toimenpiteet voivat ehkäistä rakentavia verkkokeskusteluja. Siispä epäasialliseen keskusteluun puuttuminen tulee suunnitella huolellisesti.

Tämä väitöskirja tukee epäasialliseen verkkokeskusteluun puuttuvien käyttöliittymämekanismien kriittisten ja käytännöllisten toimintojen suunnittelua. Kriittiset toiminnot viittaavat pohdinnan, keskustelun, mielikuvittelun, monimutkaisuuden havaitsemisen ja uusien näkökulmien huomioinnin herättämiseen. Tämä on hyödyllistä ongelmien esiin kaivamiseen ja ihmisten aktivoimiseen. Käytännölliset toiminnot puolestaan viittaavat ongelmien ratkomiseen, tässä tapauksessa epäasiallisen keskustelun hillitsemistoimiin.

Väitöskirja lähestyy aihetta tunteita merkitsevien käyttöliittymäinterventioiden *kriittisen suunnittelun* kautta. Kriittinen suunnittelu on suunnittelu ja tutkimusmenetelmä, jossa herätetään keskustelua suunnittelun eettisyydestä, paljastetaan mahdollisesti piilossa olevia arvoja ja ehdotetaan vaihtoehtoisia suunnitteluarvoja. Tunteiden merkitsemisellä viitataan sisällössä olevien tunteiden nimeämiseen, jolla odotetaan psykologian tunneteorioiden perusteella olevan rauhoittava vaikutus.

Tutkimus tuotti neljä tieteellistä julkaisua. Tutkimukseen sisältyi kymmenen suomalaisen journalistin haastattelu koskien neljää suunnitteluehdotusta, sekä kaksi kansainvälistä verkkokyselyä, joissa taltioitiin yhteensä 687:än verkkouutiskommenttoijan reaktiot yhteensä 14:sta ehdotukseen.

Väitöskirja antaa oleellisista tietoa kriittisen suunnittelun osa-alueista ja ulottuvuuksista. Esimerkiksi, “käytännöllinen—kriittinen” -ulottuvuus, joka ilmaisee, miten vahvasti intervention tulisi toimia kriittisesti. Lisäksi se antaa tietoa mitkä asiat tekevät interventiosta laadukkaan käyttäjien mielestä. Esimerkiksi, käyttäjät ajattelevat, että laadukas interventio auttaa heitä välttämään lukemasta epäasiallisia kommentteja ja olemaan kommentoimatta tavalla, joka voisi kaduttaa.

Kaiken kaikkiaan väitöskirja tukee epäasialliseen verkkokeskusteluun puuttuvien käyttöliittymäratkaisujen suunnittelua kriittisellä äänellä ja tuo merkittävän lisän verkkokeskustelujärjestelmien suunnittelua käsittelevään kirjallisuuteen.

CONTENTS

1	Introduction	1
1.1	Critical design and the research methodology	1
1.2	Research problem I: embracing critical design of socio-technical systems is difficult.....	3
1.3	Research problem II: mitigating uncivil commenting on online news sites.....	3
1.4	Research questions	5
1.5	Contributions.....	6
2	Critical design as a theoretical framework.....	8
2.1	Being critical about the role of design.....	8
2.2	The unclear differences between alternative design practices.....	12
2.3	Design space for criticality	16
2.4	Creating critical designs	18
2.5	Critical design's relation to other design approaches.....	22
2.6	Summary.....	30
3	The challenge of moderating online discussion.....	31
3.1	The difficult problem of uncivil online news commenting.....	31
3.2	Existing approaches to digitally mitigate uncivil online news commenting.....	34
3.3	Supporting users' emotion regulation with computational affect labeling.....	36
3.4	Summary and opportunities for critical design	37
4	Research process and methodological summary of articles.....	38
4.1	Research approach.....	38
4.2	Overview of the research process.....	39
4.3	Study for publication I: An analysis of critical designs based on designer reflection and interviews with Finnish news media experts	41
4.4	Study for publications II-III: An international online survey with online news commenters on critical UI intervention designs	49

4.5	Study for publication IV: An international online survey with online news commenters to find how discussion context affects alert design evaluation.....	58
4.6	Analysis for answering to the thesis research questions	62
5	Results: design space for critical design	65
5.1	Dimensions, aspects, and guiding ideas of the design space	65
5.2	Reactions to the designs.....	72
6	Results: characteristics of high-quality user-interface interventions	81
6.1	Discovered characteristics and requirements of high-quality UI interventions in uncivil commenting	81
6.2	Background factors that may affect the perception of quality	86
7	Discussion and conclusions	88
7.1	The uses of critical UI interventions.....	88
7.2	Significance and contributions	89
7.3	Limitations and future work.....	92
8	References.....	98

List of Figures

Figure 1. Funnel diagram summarizing the present work.

Figure 2. Explaining differences between traditional design and alternative design.

Figure 3. Excerpt of the A/B manifesto (Dunne & Raby, 2013) with added categorization.

Figure 4. ISO Human-Centered Design Framework and possible alternative advice.

Figure 5. Matrix of common argument types for design's criticality.

Figure 6. Summary of potential ways to create critical designs.

Figure 7. Illustration of the parts of a socio-technical system.

Figure 8. Some questions on STSs that CD of STS might highlight and ask.

Figure 9. The Audience V1 and Creature V1 designs.

Figure 10. The Regret V1 and Promise designs.

Figure 11. The Highlight, Creature V2, Symbols, and Evaluate designs.

Figure 12. The Philosophy, Regret V2, Warning, and Audience V2 designs.

Figure 13. The expected instrumental quality and inappropriateness ratings of eight designs.

Figure 14. Illustration of six alert designs explored in Publication IV.

Figure 15. Designs' irritating-pleasing and expected overall positive change in user behavior scores.

Figure 16. My estimation of all the designs' successful critical function vs. practical function.

Figure 17. The Regret V1 and V2 designs in short.

Figure 18. The Audience V1, V2 and V3 designs in short.

Figure 19. The Philosophy design in short.

Figure 20. A summary of the discovered characteristics and requirements of high-quality UI interventions in uncivil online news commenting.

List of Tables

Table 1. The relationship between the empirical research questions and publications.

Table 2. The ways discursive (alternative) designs (critical, speculative, or fiction) may be used (Tharp & Tharp, 2013).

Table 3. The design space schema for a media architecture project. Adapted from (Biskjaer et al., 2014).

Table 4. The relationship between the studies, publications, and research questions.

Table 5. Results of regression analyses investigating associations between background variables and instrumental quality ratings.

Table 6. Intention-related design dimensions and aspects.

Table 7. System-related design aspects related to achieving the intentions in Table 6.

Table 8. Interaction design aspects and dimensions

Table 9. The CD dimensions, aspects, and guiding ideas in the thesis studies.

Table 10. The CD dimensions and aspects and the Regret V1 & V2.

Table 11. The CD dimensions and aspects and the Audience V1, V2 & V3.

Table 12. The CD dimensions and aspects and the Philosophy.

ABBREVIATIONS

CD	Critical Design
CS	Computer Science
HCD	Human-Centered Design
HCI	Human-Computer Interaction
RtD	Research through Design
STS	Socio-technical System
UI	User interface
UX	User experience

ORIGINAL PUBLICATIONS

Publication I **Kiskola, J.**, Olsson, T., Vääätäjä, H., H. Syrjämäki, A., Rantasila, A., Isokoski, P., Ilves, M., & Surakka, V. (2021). Applying critical voice in design of user interfaces for supporting self-reflection and emotion regulation in online news commenting. *In Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1-13). DOI: 10.1145/3411764.3445783

Kiskola was in charge of the research planning and design, creating design artifacts, and analysis of the study. Vääätäjä was in charge of collecting user data. Kiskola, as the main author, planned and wrote most of the article.

Publication II **Kiskola, J.**, Olsson, T., Syrjämäki, A. H., Rantasila, A., Ilves, M., Isokoski, P., & Surakka, V. (2022). Online Survey on Novel Designs for Supporting Self-Reflection and Emotion Regulation in Online News Commenting. *In Proceedings of the 25th International Academic Mindtrek Conference* (pp. 278-312). DOI: 10.1145/3569219.3569411

Kiskola was in charge of the research planning and design, creating design artifacts, data collection, and analysis of the study. Kiskola, as the main author, planned and wrote most of the article.

Publication III **Kiskola, J.**, Olsson, T., Rantasila, A., Syrjämäki, A. H., Ilves, M., Isokoski, P., & Surakka, V. (2022). User-centred quality of UI interventions aiming to influence online news commenting behaviour. *Behaviour & Information Technology*, 1-33. DOI: 10.1080/0144929X.2022.2108723

Kiskola was in charge of the research planning and design, creating design artifacts, data collection, and analysis of the study. Kiskola, as the main author, planned and wrote most of the article.

Publication IV **Kiskola, J.**, Olsson, T., Syrjämäki, A. H., Rantasila, A., Ilves, M., Isokoski, P., & Surakka, V. Evaluating Alerts to Impolite Online News Commenters: The Impact of Previous Commenter's Politeness and the Form and Amount of Guidance. *Under review at Behaviour & Information Technology*.

Kiskola was in charge of the research planning and design, creating design artifacts, data collection, and analysis of the study. Kiskola, as the main author, planned and wrote most of the article.

1 INTRODUCTION

Recently, it has become apparent that online discussions are often unhealthy in several ways (Brey et al., 2019). In this thesis, I focus on focus on uncivil commenting on online news sites (Diakopoulos & Naaman, 2011; Rantasila et al., 2022; Stroud et al., 2016). By “uncivil” commenting, I refer to the unnecessary use of impolite, insulting, and toxic language or comments that, for example, deny opinion expression from others (adapting the definition of (Coe et al., 2014) of incivility). The incivility tends to harm users, moderators, and journalists, and undercuts the value commenting can have to societies (as a form of public participation) (G. M. Chen, 2017; Rantasila et al., 2022). Moderation of uncivil commenting is the primary sociotechnical problem addressed in this thesis, further elaborated in Chapter 3.

I focus on two problems in this area: (I) Critical Design (CD) of online news commenting appears necessary but it is methodologically difficult to implement due to a lack of literature providing guidance and inspiration, and (II) it is difficult to mitigate uncivil online news commenting as a cultural and behavioral phenomenon. The first is a problem related to the practice of designing sociotechnical systems (STS), which refer to systems comprising two jointly independent but correlative *interacting* systems—the social and the technical (Bostrom & Heinen, 1977; Whitworth, 2009). The second is a problem of mitigating bad behavior in a particular STS. I conduct CD and user-research on UI interventions in uncivil online news commenting to address both problems.

In the following sections, I elaborate the goals and research process, and why the work is important.

1.1 Critical design and the research methodology

This section introduces the design theory that is essential for understanding this thesis. This section also introduces the research methodology.

A significant catalyst for this thesis is the relatively recent ethical awakening in Human-Computer Interaction (HCI) and Computer Science (CS) regarding

Information and Communication Technologies (Dunne, 1999; Horton et al., 2022). Researchers have noted that production-related expectations of technology design can be problematic, as they can entail spending little time asking what problems the technologies propagate and what problems they may introduce (Dunne, 1999; Jakobsone, 2017; Pierce, 2021). Additionally, researchers and governments increasingly recognize that the internet, despite all its benefits to society, can also be correlated with harmful effects on individuals and society (Brey et al., 2019).

The ethical awakening may be seen to have begun in the early 2000s, when Dunne & Raby popularized Critical Design (CD) (Dunne, 1999; Dunne & Raby, 2001). CD is a Research through Design (RtD) methodology that foregrounds the ethics of design practice, reveals potentially hidden agendas and values, and explores alternative design values (J. Bardzell & Bardzell, 2013). RtD in HCI is about employing design practice methods, practices, and processes to generate new knowledge (Zimmerman & Forlizzi, 2014). In CD, designs are used to highlight ethical and social concerns or phenomena. The concerns or phenomena may range from highly speculative (e.g., child-pet relationship is one of exploitation only) to less so (e.g., violence toward women) (J. Bardzell et al., 2014; Dunne, 1999). These kinds of designs function like (societal) critique and may be called *critical designs* (see quote below and (J. Bardzell et al., 2014; J. Bardzell & Bardzell, 2013)).

“When is ‘critical design’ [critical]? When it functions as follows: Successful critiques are those that help others perceive and experience for themselves that work, phenomenon, or concern in more insightful, experientially worthwhile, and actionable ways than they could do themselves.” (Conference presentation by J. Bardzell, 2018)

However, creating critical designs contrasts with what people expect of designers in 2023. Few people expect designers to deliver designs that primarily work best for provoking conversations about designs and design problems (Pierce, 2021). Instead, designers tend to be expected to tolerate uncertainty, work with incomplete information, use imagination and constructive forethought in solving practical problems, and use drawings and other modeling media as means of problem-solving, and come up with novel, unexpected solutions (Lawson & Dorst, 2009; Pierce, 2021). While designers are expected to be critical (e.g., reflect what they do and challenge traditions of design) (Mazé, 2009), they are not expected to show it by creating critical designs.

In all of the thesis publications, I use CD strategies to create critical designs and use the designs to elicit reactions and opinions from participants. Regarding these methodological choices, this work is related to previous RtD work, for example

(Beheshtian et al., 2020; J. X. Chen et al., 2021). Overall, this thesis explores both how design features may function critically and how they may present practical functions. While most of the thesis publications focus on empirical user studies, in this introductory part of the thesis I analyze the criticality of the designs in more detail.

1.2 Research problem I: embracing critical design of socio-technical systems is difficult

As mentioned, CD is difficult to embrace because it involves challenging oneself and the conventions of the design discipline. Nevertheless, embracing CD can be more challenging when little previous CD work exists to support new work. Novice designers often rely upon previous work for guidance and inspiration (Ferri et al., 2014; Johannessen, 2017; Tharp & Tharp, 2019). It is generally true that the CD's quest for triggering a reflective response can involve much trial and error if one does not already possess knowledge about what is likely to trigger a reflective response (S. Bardzell et al., 2012).

One of the areas where embracing CD is difficult for the above reasons is STS design. Modern examples of STSs include, for example, transportation systems, social media, commenting platforms, workplaces, and email. STS use and management involves negotiation, for example, about the meaning of social order and freedom and how those are achieved (Whitworth, 2009). These facts seem to make STSs an apt area for CD. Critical designs may provoke reflection about sociotechnical issues and the idea that people need to be changed, not (merely) technology. Furthermore, as STSs are very complex (Whitworth, 2009), it may be valuable to show provocative critical designs to STS users and listen to what they have to say. However, despite these motives for doing CD of technological interventions in STSs, little previous work exists.

1.3 Research problem II: mitigating uncivil commenting on online news sites

Uncivil online news commenting seems challenging to mitigate. To begin with, it is not easy to define what good quality and civility even mean in the comments (Bossens et al., 2021; Diakopoulos & Naaman, 2011; Masullo Chen et al., 2019).

Academics have also spent decades debating what constitutes legitimate public expression (Fraser, 1990; Habermas, 1991; Rowe, 2015), focusing on questions such as whether dispassionate deliberation is synonymous with legitimate public debate or not (Fraser, 1990). Further, good comment moderation requires careful consideration of comments in their contexts and guiding commenters, making it difficult and expensive to conduct at scale (Rantasila et al., 2022).

These challenges have motivated a search for technological solutions that could work at a large scale and complement human moderators. For example, machine learning-based UI interventions that gently notify users when their comments may violate the guidelines have been developed (*Jigsaw(google) Perspective tool*; Simon, 2020). However, despite this quest for technological solutions, relatively little conventional design research exists in the area (e.g., (Bossens et al., 2021; Seering et al., 2019)). As a result, the perspectives and expectations of people affected by UI interventions in uncivil commenting are virtually unknown, such as the perspectives and expectations of online news commenters and news media. This complicates the design of both critical and conventional UI intervention proposals.

In the thesis, I address the aforementioned knowledge gaps by studying how people react to critical UI intervention designs, most of which propose supporting users' emotion regulation through various forms of affect labeling. Emotion regulation refers to the process and strategies that influence the quality, intensity, and timing of the experienced emotion (Gross, 2015). The need for emotion regulation arises when emotions are of strong intensity, duration, frequency, or the wrong type for a particular situation, or they maladaptively bias cognition and behavior (Gross, 2015; Mauss & Robinson, 2009). Emotion regulation can be enhanced through affect labeling (Torre & Lieberman, 2018): for example, simply making the emotionally loaded elements in a message more perceivable. As a basic example, a message containing the words “f*** you” could be labeled “this user may be angry,” and this could support users' emotion regulation implicitly and effortlessly (in contrast to explicitly trying to regulate emotions) (Syrjämäki et al., 2023).

The study of various affect labeling UI interventions arose from work on a research project with this focus. I was part of a team that included researchers with psychology, human-centered design, media studies, and computer science expertise. Furthermore, I co-wrote all of the thesis publications with colleagues from the research project, with me as the first author. Figure 1 summarizes the thesis focus areas.

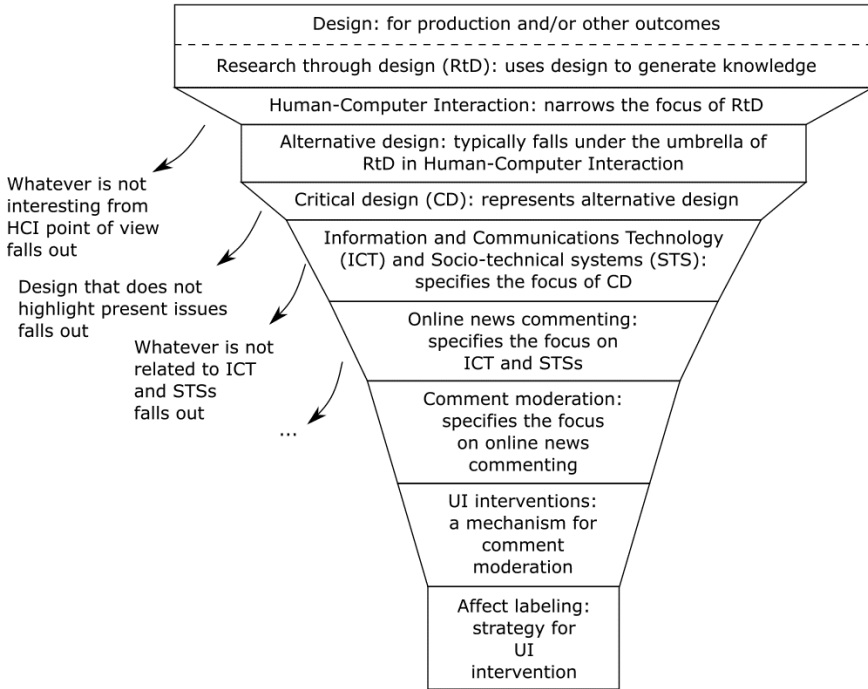


Figure 1. Funnel diagram summarizing the present work. The dotted line between design and RtD illustrates that their difference is unclear (Gaver et al., 2022).

1.4 Research questions

I address the above research problems by investigating two research questions that offer concrete viewpoints to the problems:

RQ1: What characterizes the design space for critical design of user-interface interventions aiming to influence online news commenting behavior?

Here, design space refers to “a conceptual space, which encompasses the creativity constraints that govern what the outcome of the design process might (and might not) be” (Biskjaer et al., 2014). Creativity constraints can take, for example, the following forms: design *aspects* and *dimensions* (e.g., functionality and size); design *options* (e.g., a specific function and size); and design *exemplars*. Constraints are selected based on situational requirements, knowledge, and beliefs (Biskjaer et al., 2014). Knowledge about what designs or design features engender discussion and reflection, imagination, appreciation of complexity, and consideration of new

perspectives may be used to constrain CD (J. Bardzell et al., 2014; S. Bardzell et al., 2012). Accordingly, the research question features the following two sub-questions:

- a) What design dimensions and aspects may reasonably be used to constrain CD in this context?
- b) What kind of designs and design features are likely to provoke reflection, discussion, imagination, appreciation of complexity, and consideration of new perspectives among people who are knowledgeable about online news commenting?

RQ2: What characterizes a high-quality user-interface intervention aiming to influence online news commenting behavior?

The question stems from a lack of knowledge about online news commenters' and comment moderation experts' expectations and requirements of UI intervention designs. This question is mostly about practical issues, where the previous is mostly about criticality. Aside from seeking to understand what people believe constitutes a high-quality UI intervention design, I seek to reveal what assumptions people might have regarding my designs. This is based on the CD strategy of "looking beyond the surface" to perceive and identify hidden phenomena that drive behavior and unequal social systems that are nevertheless widely accepted (J. Bardzell & Bardzell, 2013; Jakobsone, 2017). I also consider how the participants' standpoint might have impacted their quality-related comments.

1.5 Contributions

I address the first research problem concerning the difficulty of embracing CD of STSs by providing guidance for design in this context. I also address this problem by connecting CD with the design of UI interventions in STS. Finally, I address the second research problem of the difficulty of mitigating uncivil commenting by providing knowledge regarding what UI interventions might be acceptable to users and what needs to be considered to mitigate incivility while minimizing undesirable side-effects like driving many users away. Table 1 illustrates how the individual publications in the thesis contribute to answering the research questions.

Table 1. The relationship between the empirical research questions and publications.

Publication, data gathering method	Key contributions to RQ1: Characteristics of the design space for critical design	Key contributions to RQ2: Characteristics of high-quality user-interface interventions
P1, interviews	Four critical designs. Insight about conducting CD in this context. Knowledge on what designs Finnish journalists consider possible.	<i>Lesser role but gives insight into what Finnish journalists consider important in general.</i>
P2, online survey	Eight critical designs. Online news commenters' evaluations of the designs.	<i>Lesser role. Shows that those hoping for more comment moderation are more likely to believe that a design would work.</i>
P3, the same online survey as above	Insight into design dimensions- what to consider varying in CD in this context.	Insight into what online news commenters consider important and expect of UI interventions in general.
P4, another online survey	Three variations of a critical design and three variations of a more conventional design. Insight into constraints: what designs and design features online news commenters consider possible.	It was found that the apt use of textual and visual guidance in UI alerts positively impacts their user-perceived quality.

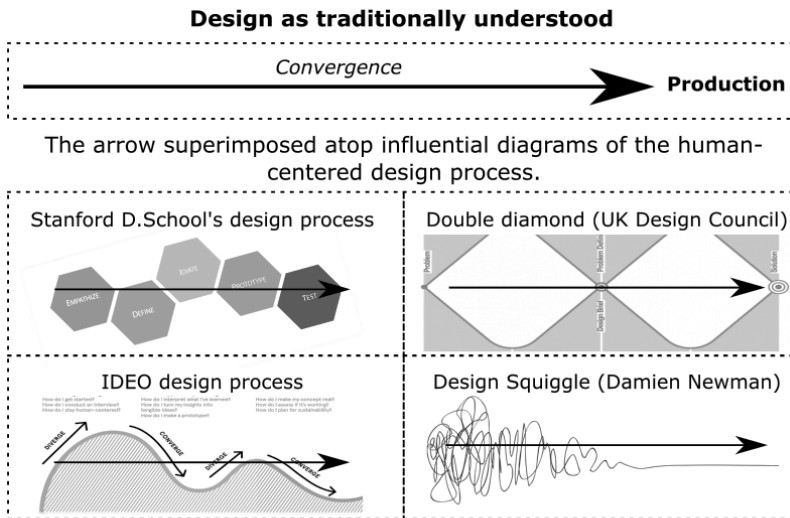
2 CRITICAL DESIGN AS A THEORETICAL FRAMEWORK

This chapter begins by describing the Critical Design (CD) philosophy and strategies in more detail. Then the chapter describes CD's relation to other design approaches relevant to this thesis, such as Human-Centered Design and intervention design.

2.1 Being critical about the role of design

In this section, I expand on what I discussed in the introduction and focus on the critique of personal and disciplinary aspects of design. I focus on what beliefs and expectations may threaten design discipline and how a designer may do harm by failing to reflect on their work. In other words, this section describes the philosophy of critical design.

Designers tend to be expected to come up with novel, unexpected *solutions* (Lawson & Dorst, 2009; Pierce, 2021). Pierce argued that this is evident from influential diagrams of the human-centered design process. The diagrams suggest that designers deliver solutions (see Figure 2. below).

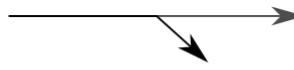


Alternative design

Introduces friction to productional expectations. For example:

*Exhibiting a critical stance
toward current tech, practices,
values, etc.*

*Departing from conventions
and expectations*



Alternative designs are compelling and potentially useful
because, not in spite, of resistance to production

Figure 2. Traditional design converges toward production, while alternative design introduces friction to productional expectations. I adapted the illustrations in the top-half from (Pierce, 2021). All the claims about alternative design were made first by Pierce.

Pierce (2021) argues that it is wrong to assume that designs must exhibit clear, direct, and straightforward progressional intentions and potentials. While traditional design processes (see Figure 2. above) aim toward production, proponents of alternative design practices argue that designs do not need to satisfy productional expectations nor attempt to address any known problem. Instead, designers may have purposes other than production in mind. In Discursive Design, Tharp & Tharp (2019) describe how designers can deliberately design objects to act as “intellectual prostheses” that provoke or support thinking about things or imagining alternatives to the present. Pierce (2021) explains that such designs “offer purposes, uses, functions, effects, and other types of value other than the production they literally, ostensibly prefigure.” For example, designs may help people think about complex societal or environmental issues involving technology, imagine future technologies,

rethink their use of technology, etc. This way, design could answer concerns regarding the risks of technologies and designers could demonstrate that they consider different ethical, societal, and environmental viewpoints.

In the previous section I briefly described that the expectation that design only converge toward production may not benefit design discipline. However, in the following I describe some current false and narrow approaches to and beliefs about design in more detail. For this purpose, I rely on Tharp & Tharp's (2019) list of issues, but they are not alone in their observations (see e.g., (S. Bardzell et al., 2012; Blythe et al., 2016; Dunne & Raby, 2013; Noortman et al., 2021; Pierce, 2021)).

Overly strong expectation for the practical, utilitarian use of designs. This expectation hinders the creation of designs primarily intended to make people ask questions, such as who is served, for how long, at what expense, and with what possible unintended consequences (Jakobsone, 2017; Tharp & Tharp, 2019). The expectation pairs with moving too fast to production from design, as asking questions takes time. Not asking questions may result in accidental but avoidable bad outcomes (Nelson & Stolterman, 2012). By making it harder to ask questions about a design, the expectation supports "solutionism": the creation of technological solutions that ignore the complexity of personal, political, or environmental issues (Blythe et al., 2016; Morozov, 2013). For example, people could too strongly expect designers to design functional, "real" UI interventions in uncivil online discussion. Designers, clients, and end-users can all fail to see the value in creating UI intervention designs that are intended to make people ask questions or discuss the topic.

Characterizing design as only a problem-addressing activity. This characterization hinders designers from asking "What if?" in the most open and projective sense, where the question "What if?" is not related to solving any known problem (Tharp & Tharp, 2019). Asking such open questions could lead to new opportunities. For example, designers could ask, "what if an AI would run a news website?" and then think about opportunities. It may also be helpful for designers to ask purposefully silly questions to free the mind (Blythe et al., 2016). For example, "What if dogs would run a news website?" In light of this, I note that while my CD work is related to the problem of uncivil online news commenting, I am not suggesting that CD work must address a known problem.

Overemphasizing the individual user's perspective and ignoring the society (Tharp & Tharp, 2019). This hinders designers from asking what the social implications of the design are and how others could use the design for good or evil. Instead, design should engage with the social implications of technologies, and designers should take responsibility as 'shapers' of society (Dunne, 1999; Tromp et al., 2011). For example,

the designers of UI interventions in uncivil online news commenting could ask if designs should reflect individualism (e.g., individual achievement and self-expression) over collectivism (e.g., group success and adherence to norms). Critical design can bring such ideological questions to the forefront (Jakobsone, 2017).

Placing excessive emphasis on how nice designs look and feel. Excessive emphasis on the form can hinder designers from making insightful intellectual contributions, for example, toward user experience (Tharp & Tharp, 2019). In order to make such contributions, not only should crude prototypes be accepted, but also purposefully distasteful, offensive, or silly ones. They can be used, for example, to elicit users’ values or to draw people into an open-minded creative process or discussion. In the first thesis publication, I presented a design I intended to feel distasteful and silly. In the design, I propose to show a virtual dog-like creature dead, pierced by arrows, if the user engages in uncivil commenting. Among other things, the design helped the interviewed senior-level journalists to consider how far news media organizations would be ready to go to intervene in uncivil commenting.

In addition, Dunne & Raby critique the abovementioned false approaches to and beliefs about design in their well-known A/B manifesto (Dunne & Raby, 2013; Johannessen, 2017; Pierce, Sengers, Hirsch, Jenkins, Gaver, & DiSalvo, 2015).

(A)	(B)
BASIC ORIENTATION	
design for production	design for debate
affirmative	critical
in the service of industry	in the service of society
consumer	citizen
HIGHLY VALUED PROCESSES AND OUTCOMES	
problem solving	problem finding
provides answers	asks questions
ergonomics	rhetoric
innovation	provocation
fun	satire
makes us buy	makes us think
RELATION TO THE PRESENT	
for how the world is	for how the world could be
change the world to suit us	change us to suit the world
futures	parallel worlds
science fiction	social fiction

Figure 3. Excerpt of the A/B manifesto (Dunne & Raby, 2013) with added categorization.

Dunne & Raby listed how design is usually understood (A) and what else design could do (B). Figure 3 presents an excerpt of the manifesto, with my categorization of the items. Overall, the A/B manifesto may be considered tactical advice for widening the horizon of design (Dunne & Raby, 2013; Johannessen, 2017; Pierce, Sengers, Hirsch, Jenkins, Gaver, & DiSalvo, 2015).

Nevertheless, all the previous should not be interpreted to claim that conventional design does not involve any critique or critical thinking. In general, being a critical designer means being critical about all three aspects of design work: *personal* (e.g., reflecting what one does and why), *disciplinary* (e.g., trying to challenge or change traditions and paradigms of design), and *public* (e.g., trying to address pressing issues in society) (Mazé, 2009). People also expect all of this from industrial designers to some extent. As Dunne & Raby (2013) put it, “all good design is critical.” That this is true can also be seen at the level of designs, for example, in social media, where a bad design may be said to make people scroll cat videos for hours on end but a good design to provoke people to rethink their social media use (Mujica et al., 2022). Thus, rather than viewing criticality and practicality as enemies, it is wise to note they are not mutually exclusive (Ghajargar & Bardzell, 2021).

In summary, some designers perceive the expansion and evolution of design to be limited by several false approaches to and beliefs about design. These may also negatively affect the design of UI interventions in uncivil online news commenting. The CD work in this thesis is partially motivated by a suspicion that the current online news commenting systems result from premature narrowing down on solutions. Accordingly, I seek to avoid narrowing down on solutions in the research while contributing knowledge relevant to designing UI interventions and addressing the problem of uncivil commenting.

2.2 The unclear differences between alternative design practices

“There is much overlap between [critical design, speculative design, discursive design, design probes, and design fictions], the differences are subtle and based primarily on geographical or contextual usage: all remove the constraints from the commercial sector that define normative design processes; use models and prototypes at the heart of the enquiry; and use fiction to present alternative products, systems or worlds.” (Auger, 2013)

As said, alternative design practices like CD involve introducing friction to productional expectations and seeking outcomes other than production. They

appear to share the same core idea of using friction to productional expectations. However, it is unclear which alternative practices (critical, speculative, reflective, fictional, etc.) feature which, more specific frictional tendencies (Pierce, 2021). The alternative practices are useful toward many, sometimes competing, and often intersecting aims and ends (ibid.). Researchers engaging in alternative design claim they have difficulty in differentiating between the various alternative design practices (Auger, 2013; Tharp & Tharp, 2019).

Furthermore, I note that the names of the alternative practices are interpreted differently by different people. For example, calling an alternative RtD project “critical” may confuse people because there are many interpretations of what designers and researchers can rightly label as “critical.” For example, the label may be expected to suggest that the design empowers people or combats those in power (Iivari & Kuutti, 2017) or that the design is associated with Dunne, Raby, and their students and disciples (Pierce, Sengers, Hirsch, Jenkins, Gaver, & DiSalvo, 2015), or that there is a strong critical contribution in a broader sense (J. Bardzell & Bardzell, 2013). Further, Tharp & Tharp (2019) note it is even possible to think that the label “critical” means primarily that the work is simply critical to the operation of a society or a system.

For the abovementioned reasons, the following explains my interpretation of the three most well-known alternative designs: Critical Design, Speculative Design, and Design Fiction (Johannessen et al., 2019; Pierce, 2021; Tharp & Tharp, 2019). In my view, *Critical Design* (CD) is an alternative design practice that particularly emphasizes critique and consideration of alternative (to the conventional) user-product relationships. It is not enough to relax progression assumptions a bit; an attempt must also be made to make at least one group of people think critically about the present situation. When critical designs work, they work like critique (J. Bardzell et al., 2014; J. Bardzell & Bardzell, 2013). A critique points out the assumptions and familiar, unchallenged, unconsidered modes of thought on which the practices we accept rest (Foucault, 1981). Critical designers intend critical designs to perform a “critical function,” to make people think critically, enhance appreciation, change people’s perspectives, etc. (J. Bardzell et al., 2014). However, CD can also be user driven, rather than designer driven, as identified by Iivari & Kuutti (2017). Nevertheless, critical designs feature irony, an internal conflict that reaches out into the world, intending to involve the audience emotionally and arouse the audience’s interest. They may also be argued to feature satire, diminishing themselves, individuals, the design discipline, or society by making them appear ridiculous (Malpass, 2017). If a critical design is viewed for a very short time, it may, however,

appear like a conventional product design. In contrast, Speculative Design and especially Design Fiction seemingly allow other types of outputs (such as narrative, role-playing, or image collages).

However, while I believe critical designs must emphasize criticality, I subscribe to the notion that criticality does not intrinsically exclude practical use (Ghajargar & Bardzell, 2021). For example, designers could incorporate a diorama of a coal power plant and dirty workers into an office lamp design. Further, while a critical design's "critical function" can address a clear and pressing issue in society, like the use of coal power plants or uncivil commenting online, I believe this is not a requirement. Critical designs may also ask if something people do not generally consider a problem is a problem. For example, a critical design could ask whether generic coffeemakers are instruments of oppression.

Additionally, while CD often seems to use shocking and questionable design metaphors and imagery, I believe this is not required either. For example, while critical, Alice Wang's Peer Pressure Project (A. Wang, 2008) does not employ disturbing imagery. Wang's design artifacts are a printer, a mobile phone, a keyboard, and headphones; the criticality depends on how Wang described them. For example, they described the mobile phone as "a phone that randomly receives text messages to make you look popular in public."

Next, I believe *Speculative Design* is a more future-oriented, somewhat less didactic offspring of CD. Speculative designs attempt to trigger debate about *future* challenges (Dunne & Raby, 2013). Speculative design may propose technologies that are so advanced that they might as well be magic. I interpret speculative designs to both accelerate to the future and noticeably diverge from the unexpected-expected designs (i.e., designs that answer to problems or needs that we were not [fully] aware we had) (the terminology comes from (Nelson & Stolterman, 2012; Pierce, 2021)). For example, consider a proposal that we all wear extended-reality glasses in the future, and send three-dimensional animated virtual characters (dogs, butlers, etc.) to deliver important emails to other people. I believe this proposal is far removed from today and seemingly belongs in an alternative universe inhabited by people with different needs and problems. Hence, I believe it is speculative. However, suppose the proposal were to be changed so that what is sent are birthday cards, and the receiver would not have to wear extended reality glasses all the time. In that case, I believe the design would no longer diverge from the unexpected-expected. Hence, it would cease to be speculative and begin to be a concept design or design vision.

Next, in my understanding, *Design Fiction* is like Speculative Design (i.e., also accelerational and diverging), except a designer creates an entire narrative world

around the designed artifact (Blythe & Encinas, 2016; Noortman et al., 2021). In other words, I interpret the word “fiction” to suggest that the design features imaginary events, relatable people, and interaction between the main characters in the form of a written story or play. For example, three people could perform a play on how the media uses artificial intelligence to keep people interested in the news commenting in the year 2100. A designer could write the play so that the first actor plays a technician in charge of the artificial intelligence, the second actor plays a business executive, and the third actor plays a user manipulated to continue commenting. Further, while these examples do not feature magic or anything supernatural, design fiction may use magic, wonder tales, or be rooted in the supernatural (Blythe & Encinas, 2016). However, I know that *Design Fiction* may also be interpreted as somewhat less diverging from the unexpected-expected and as keeping in the near future and the real world. For example, Bleecker et al. (2022) define it as follows: “Design Fiction is the practice of creating tangible and evocative prototypes from possible near futures to help represent the implications, outcomes, and consequences of decision making.”

In addition to all the above, I believe each alternative design practice can be distanced or brought closer to the HCI research tradition or the Design and Art tradition. Unfortunately, this seemingly results in projects with the same name (e.g., “critical”) looking different in practice. To explain, I adapt Tharp & Tharp’s (2013) proposal that discursive designs may be used in four ways (see Table 2. below). I will also use this table to help position my work.

Table 2. The ways discursive (alternative) designs (critical, speculative, or fiction) may be used (Tharp & Tharp, 2013).

	Instrumentally	Terminally
Internally	To study what the design team thinks, assumes, etc. to support the design project or future design projects. (Addition by the thesis author: it is unlikely that the design needs to look like fine art for these purposes.)	To make other designers think critically about or as a result of the design, hoping to change the world that way. (Addition by the thesis author: it is somewhat likely that the design needs to look like fine art for these purposes.)
Externally	To study what non-designers think, assume, etc. to support the design project or future design projects. (Addition by the thesis author: it is somewhat unlikely that the design needs to look like fine art for these purposes.)	To make non-designers think critically about or as a result of the design, hoping to change the world that way. (Addition by the thesis author: it is likely that the design needs to look like fine art for these purposes.)

The instrumental and terminal options are separated by whether the designer is present to communicate what the design does not communicate and to draw attention where the design does not draw attention (Tharp & Tharp, 2013). In terminal design, the designer puts more effort into the design because it will be like a “message in a bottle” that must both draw attention and communicate by itself. The design is the terminus of the designer’s efforts; hence the approach is terminal (ibid.). The terminal approach also seems to have a fundamental goal of reaching as many people as possible with the design, to change the world. The goal of reaching as many people as possible is sometimes assumed to be a primary goal of CD (e.g., (Rynning, 2017)).

Conversely, in instrumental design, the design does not need to draw as much attention and communicate as much by itself. For example, if designers conduct an interview study, the designer or a colleague is present to direct attention to the designs and answer questions. (I did this in thesis Publication I). Alternatively, suppose a survey is conducted on the designs. In that case, the participants have committed themselves to paying attention to the survey and hence to the designs in the survey (see Publications II, III, IV).

This distinction between instrumental and terminal alternative designs further helps me position my CD work in this thesis. My work in the thesis is mostly instrumental in nature; thus, it differs from the kind of CD work presented in museums. Pierce et al. (2015) argued that in Design, unlike in HCI, critical design is “quite polished and borrows the sleek and seductive visual and form language from fine art traditions and high-end design and advertising.”

In summary, in Section 2.2, I informed that designers are uncertain about how alternative design practices differ. People interpret even their labels (e.g., “Critical” in “Critical Design”) to mean different things. Therefore, I presented my interpretation of Critical, Speculative, and Fictional Design practices and the variety within such projects. Thus, I described Critical Design (CD) more concretely than in the previous section.

2.3 Design space for criticality

In the previous sections, I described CD philosophy and goals, what CD is, in my opinion, and what to expect from CD projects. In this section, I get closer to design practice. I explain how designers may construct and use design spaces and how designers may understand a design space in CD.

Designers often construct design spaces based on client input (Biskjaer et al., 2014). Biskjaer et al. (2014) proposed that an overview of design space could be presented with a “design space schema” (see Table 3. below for an example). This explains and demystifies the concept of “design spaces.” A “design space schema” depicts the key *aspects* and *dimensions* of a design to consider (e.g., “Do not forget to design for location and time and make the thing heavy”). It also presents the design *options* (e.g., “The design should feature a wall, an entrance...”).

Table 3. The design space schema for a media architecture project. Adapted from (Biskjaer et al., 2014).

Location	Situation	Interaction	Purpose	Experience	...
Entrance, wall ...	Arrival, exploration ...	Gesture, bodily ...	Information, branding ...	Playful, subtle

Seemingly relating to the above theory, J. Bardzell et al. (2014) proposed that a matrix of interaction design dimensions and key aspects of design and criticality dimensions could be used to help to reflect on how a design is critical (see Figure 4. below). Right now, it is unnecessary to understand what “Changing perspectives,” and other terms mean; I will explain (or interpret) them in the following section. It should only be understood that the researchers’ matrix could be used to describe design space for criticality. For example, “I plan the material for the entrance to help to change people's perspectives on something,” or less precisely, “I plan the material for the entrance to have a critical function.” Alternatively, “I plan some aspect of the design to change perspectives but do not yet know which one.”

	Changing perspectives	Proposals for change	Enhancing appreciation	Reflec-tiveness
Topic				
Purpose				
Functionality				
Interactivity				
Form				
Materiality				

Figure 4. Matrix of common argument types for design’s criticality. Adapted from (J. Bardzell et al., 2014).

Furthermore, based on the previous, it seems that what characterizes CD is that a number of design aspects, dimensions, and options typically associated with creating solutions do not bind the designer. In other words, a “critical designer” is free in some ways that a “designer” is not. It also seems that a critical designer is free to

question conventional relationships between the design aspects, dimensions, and options (i.e., they may use unorthodox or nonsensical combinations). Continuing this thought, it seems reasonable to call a designer a “critical designer” or a “speculative designer,” whenever they are expected to be free in these ways.

2.4 Creating critical designs

In this subsection, I present some ways a designer can create a critical design and argue it is critical. I expand on what I described above and in Publication I. Much of what I describe here may also apply to Speculative Design and Design Fiction.

To create a critical design, one should introduce nonobvious changes or “twists” (aka tropes) on the standard or unexpected-expected design (J. Bardzell et al., 2014; Johannessen, 2017). However, before explaining how the twisting is done, I will briefly explain how somebody may create a standard or unexpected-expected design. Designers first learn the “precedents,” which are whole or partial pieces of existing design solutions combined with knowledge about their usefulness (Lawson, 2004). For example, it could be a precedent that to post a comment, one writes in a text-area HTML element that has a white, gray, or black background color. After the designers feel that they have learned enough about the “precedents”, they begin sketching new designs and let the “precedents” influence the process (Lawson, 2004). Hence, with the help of the “precedents,” the resulting designs are in keeping with what is considered standard, conventional, or not too unexpected.

In CD, that which would be considered a standard design is twisted. The twists may be introduced, for example, by subverting (i.e., destroying or damaging something essential), exaggerating (e.g., representing something as better, worse, more prominent than it is), or juxtaposing (i.e., placing different things together for a contrasting effect). The twisting is done because the criticality of designs is tied to the display of nonobvious or novel design features, which one can argue to perform a critical function, express criticality, etc. (J. Bardzell et al., 2014; J. Bardzell & Bardzell, 2013). For example, the presence of a paradox, a contradiction, irony, and satire in critical designs results from their features’ strangeness and ambiguity (Malpass, 2017). Dunne (1999) argued that a critical design should have a “slight strangeness”; it should not be dismissed outright by the intended audience nor instantly adopted.

Additionally, based on CD literature and what I report about my design process in Publication I, the twisting may be conducted in at least the following ways (or some combination thereof):

- Simply doing twists without knowing what one is trying to do. That is, “twist” first and analyze the results later.
- Twisting the design features or aspects desired to have critical functions and leaving others untouched.
- Choosing a literary device (e.g., irony, ambiguity) and doing twists to that end. That is, holding the literary device in mind and “twisting” to create it.
- Using knowledge of what design feature or characteristic would likely be perceived as “slightly strange” by the intended audience and designing to that end.
- Using one of the four aspects of criticality identified by Bardzell et al. (2014) as a starting point and doing twists to that end. (I will show what the aspects of criticality are a couple of paragraphs later).

This twisting process may also be informed by scientific theories (e.g., on emotion regulation). In this case, the resulting critical design might appear scientific or like science-fiction (see Blythe & Encinas’s (2016) discussion on scientific and other kinds of design fiction). Alternatively, the twisting may be informed by critical theories, often on the hidden significance of everyday life, structures, or objects (J. Bardzell et al., 2018). For example, a designer could propose that the design of the online discussion UI propagates a dualistic belief that mathematical-logical intelligence is superior to emotional intelligence. Then, this interpretation could inform the twisting, resulting in designs that highlight emotional intelligence's role. In connection with this idea, Dunne & Raby (2013) propose that “one of critical design’s roles is to question the limited range of emotional and psychological experiences offered through designed products.”

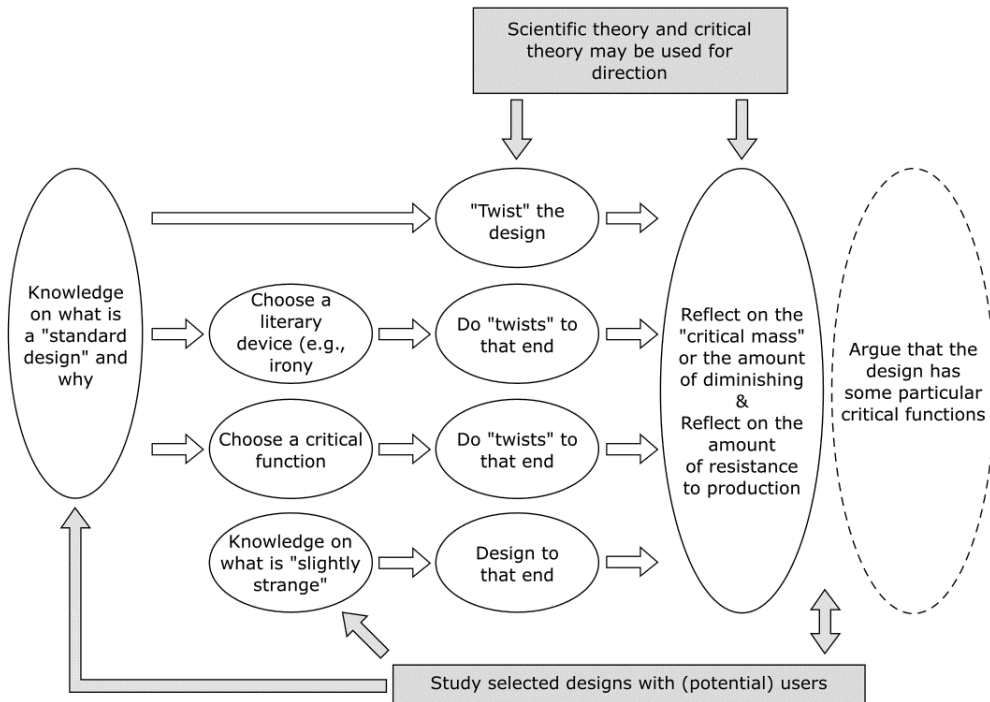


Figure 5. Summary of potential ways to create critical designs.

After some twisting, it should be considered whether or not the number or “mass” of the twists is appropriate (see Figure 5.). Regarding considering the “mass” or the number, J. Bardzell et al. (2014) note, “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.” However, this is not an easy task. It may take multiple rounds of design and studies with the designs to find a design that is neither too strange nor too mundane (S. Bardzell et al., 2012).

Further, it may be helpful to analyze the criticality of the designs in detail, that is, after one feels confident the designs are provocative enough. To do the detailed analysis, one could use the four aspects of criticality of designs identified by Bardzell et al. (2014). According to the researchers, a critical design may simultaneously manifest one or more of the four aspects and possibly others. A critical design can:

- *Change perspectives* by presenting “a framing or point of view that is new, coherent, and interesting enough to help the user perceive the particulars of a domain according to a new schema.” This may be achieved by featuring new, interesting, and coherent combinations of design features (purpose, functionality, interactivity, form, etc.). For example, an online discussion

moderation system where a comment can be deleted only after it has been read aloud. This is arguably a novel combination of removing comments (functionality) and voice-user interface (interactivity).

- *Propose change* by embodying “a provocative proposal for an alternative way of being; the proposal is grounded in possibility, cannot be easily dismissed as ‘science fiction,’ and the user can imagine her or himself in its universe.” I interpret this may be achieved by combining things and the idea of living with the design in a provocative way but still grounded in possibility. For example, a jacket featuring a screen that displays uncivil online news comments.
- *Enhance appreciation*. “The design contributes to the user’s appreciation of or judgment on design’s role(s) in a sociocultural issue of significance, by making the user more perceptive, imaginative, or aware of the complexity (political, symbolic, etc.) of a domain.” I interpret this may be achieved by combining things to create contrast. For example, a proposal to gamify online news commenting could reveal the complex relationship of news media and comments.
- *Encourage reflectiveness*. “This can be understood in two senses. One is the sense of encouraging user reflectiveness, that is, facilitating the user’s shift from direct perception and action to a more reflective or self-aware stance. The other is the design itself embodying reflectiveness, for example, by revealing or foregrounding the tropes by which it distinguishes itself from design conventions as the rhetorical devices that they are.” I interpret the former may be achieved by contrasting things and the knowledge of the present way of living or acting. For example, a proposal that online news commenting could be policed by actual police. I interpret the latter may be achieved by presenting an unusual combination of things in a design.

In summary, in Section 2.4, I described general rules and strategies for CD, emphasizing the importance of introducing twists to challenge conventional design norms. Designers can achieve criticality and provoke reflection by twisting, subverting, exaggerating, or juxtaposing elements. I suggested various approaches to the twisting process, including using literary devices, leveraging audience perceptions, and focusing on critical aspects.

2.5 Critical design's relation to other design approaches

In this section I explain CD's relation to other design practices relevant to this thesis. Note that while the first research question of the thesis explores criticality, the second relates to user/human-centered design and explores practicality. Yet, both questions are explored in studies by using critical UI intervention designs.

2.5.1 Differences between critical and human-centered design

In the following, I further explain what CD is by describing how CD goals may be related to the goals of well-established Human-Centered Design (HCD) practice, which are linked to the second research question. To remind, the second research question is about the characteristics of high-quality intervention designs.

Notably, the HCD framework (ISO, 2019) and other influential diagrams of the HCD process manifest progressional assumptions, expectations, and arrow-like movement (Pierce, 2021). The diagrams outline roughly comparable stages for design: "empathize," "brainstorm," "make prototypes," and "deliver solutions that work" (ibid.). Critical designers create friction toward the progressional expectations (ibid.). To illustrate this, I took what is contained in the HCD framework and added some alternative advice (see Figure 6 below). The alternative advice follows Dunne & Raby's (2013) tactical advice for critical designers, Bardzells' (2013) arguments on what makes a critical design project "critical," and Pierce's (2021) arguments on what makes alternative designs "alternative." However, note that the alternative advice in the figure lays out neither stages for CD nor steps designers must take in CD. The alternative advice rather illustrates how CD departs from HCD.

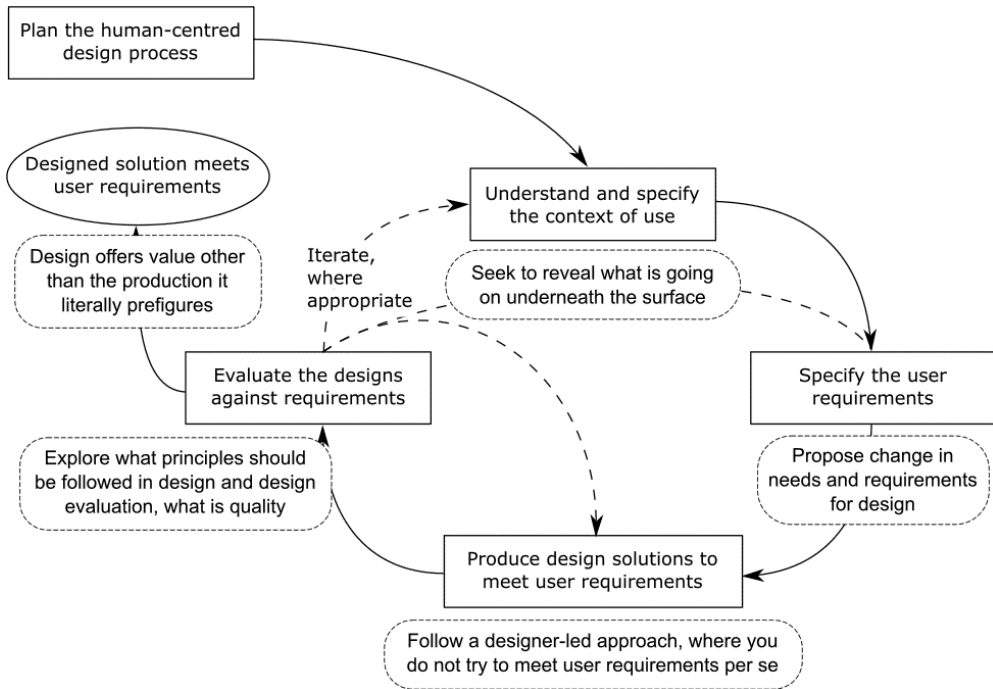


Figure 6. ISO Human-Centered Design Framework (see (ISO, 2019)) and possible alternative advice (rounded rectangles) (based on (J. Bardzell & Bardzell, 2013; Dunne & Raby, 2013; Pierce, 2021)).

Next, I discuss different parts of the Figure 6.

Seek to reveal what is going on underneath the surface (see Figure 6. top). The spirit of CD calls for challenging, highlighting, and trying to reveal and make people consider why they think and act in the ways they do (Dunne & Raby, 2013; Jakobsone, 2017). HCD, however, appears more willing to accept the way things are. For example, in a critical design in Publication I, I propose that online news commenters should benefit the collective more than they do currently. I propose placing a virtual audience of other users in front of the user to judge their commenting.

Further, peoples' reactions and comments to critical designs may be analyzed to reveal hidden assumptions and power dynamics. Critical designs may provoke people to say things they would not usually say out loud. This makes critical designs potentially useful for research purposes. Additionally, the type of information gained may be helpful to an HCD project that occurs alongside or after the CD project. It is possible to use CD to generate knowledge that may support conducting a production-oriented HCD project (Bowen, 2009; Johannessen et al., 2019; Tharp & Tharp, 2013).

Propose change... Follow a designer-led approach... CD often has an attitude that the designer knows the best (Iivari & Kuutti, 2017). However, according to Norman (2005), one of the main goals of HCD is to ensure that designers do not ignore the users, thinking they know the best. CD can hence appear very wrong from the HCD point of view. However, HCD and CD may not conflict with each other in this regard because critical designers do not, in all seriousness, argue for implementing the kind of critical designs they know might be dangerous or unethical to deploy. However, I note critical designers may pretend to advocate implementing or deploying potentially unsafe or untested critical designs as this may help ensure that audiences do not simply ignore the designs as “art” (Pierce et al., 2015).

Explore...what is quality. The spirit of CD calls for treating the designs and what they should be required to be like as part of the unknown and as in transformation (S. Bardzell et al., 2012). Following HCD, one could study what the online news commenters need, design a UI intervention to uncivil online news commenting, and then evaluate it against what the different users require. However, following CD, one could instead seek to *transform* how people conceive designs and relevant concepts in the first place (ibid.). For example, given that online news commenters probably have folk theories on what freedom means in commenting, one could encourage them to re-evaluate these theories reflectively. Accordingly, in Publication III, I studied what online news commenters speculated characterizes a good UI intervention, where their speculations were informed by the critical designs, their subject-matter knowledge, and their folk theories. Further, the thesis RQ2 “What characterizes high-quality?” treats characteristics of quality as unknown and *is thus related to CD aims*.

In summary, it appears that CD is in friction with the traditional HCD practices, not against them. These approaches may have a mutually beneficial relationship. For example, there is always a chance that a CD project provides valuable knowledge for a production-oriented design project. The thesis features friction against running HCD cycles to solve the problem of uncivil online discussion. I present and use critical designs intended to resist moving them into production. Yet, study participants reactions to them also reveal opportunities for HCD.

2.5.2 Critical design and software design

The above was about creating critical designs in general; this subsection describes the differences between designing software and physical artifacts. Describing them

is relevant for two reasons. First, I have done CD of UI interventions, which are software. Second, designing software that has critical functionality is seemingly not discussed much in alternative design literature (J. Bardzell et al., 2014; Dunne & Raby, 2013; Iivari & Kuutti, 2017; Pierce, 2021; Tharp & Tharp, 2019).

The design of software differs from that of material products. Only the design of physical products requires consideration of materials, constructing the product in a workshop or factory, and packaging and shipping it. As software design has no physical limitations like that, software can be changed and updated much quicker and cheaper than a physical product. Additionally, software designs may be displayed or made available to a broader audience faster than physical designs. For example, with the help of social media marketing tools and expiring links, a digital design might be shown to ten thousand adults interested in Kantian philosophy, for only ten minutes.

This suggests that meeting CD goals like exploring alternative designs and provoking reflection is more straightforward, cheaper, and faster in the digital realm than in the physical realm. However, despite these opportunities, I believe many designers may still find it challenging to do and present critical software designs. This is because the CD goal of *problem finding* (Dunne & Raby, 2013) appears to be in friction with the current culture of extremely rapid advancement of digital *solutions* (Dufva, 2020; Perrigo, 2023). However, I believe designing software critically might be considered desirable by organizations that can use discoveries and insights gained in CD projects in future projects.

To go into more detail on designing software critically, I discuss how CD may relate to User Experience Design, Interaction Design, and UI Design, all of which are relevant to software design. User Experience refers to all aspects of the end-user's interaction with the company, its services, and its products (IxDF, 2023; D. Norman & Nielsen, 2023). For example, regarding movie databases, User Experience Design might begin by asking whether the user's favorite movies are included (D. Norman & Nielsen, 2023). This is because if they are not, the experience may be ruined (ibid.). Considering CD and problem finding, trying to reveal what is hidden, proposing change, and other CD tactics are possible in User Experience design.

Interaction Design (IxD) is “the creation of a dialogue between a person and a product, system, or service” (Kolko, 2009). To use the previous example, interaction Design does not focus on if the right movies are featured in a movie database. Instead, it could focus on how the user interacts with an intelligent movie selection

assistant. Regarding interaction, CD may be used to provoke discussion about user behavior and complex systems behavior.

UI design is sometimes regarded as a sub-practice of IxD, where UI design focuses on the arrangement and form of static interface elements (Cooper et al., 2014). Here, a CD project could aim to provoke discussion through provocative or unconventional arrangements or forms of UI elements.

2.5.3 Critical design and intervention design

Having said the above about critically functioning software designs, I now discuss creating and reading a critical *intervention* design (aka persuasive design, or nudge). This is relevant as I design critical UI interventions in uncivil online news commenting. In persuasive design, attempts to influence user behavior should match the user's level of motivation and ability to act (Caraban et al., 2019; Fogg, 2009). Presumably, when persuasive designs are critical, they attempt to provoke reflection on assumptions or beliefs that (somehow) relate to the user's motivation or ability to act. For example, they may attempt to challenge assumptions about using persuasive designs or desirable user behaviors.

Notably, critical and everyday design artifacts may be considered arguments in material or digital form, arguing for some user action (e.g., a chair argues for using it as a chair) (Redström, 2006). However, critical designs may be perceived to contain more complex, provocative, and reflective arguments than conventional designs. This can be proven, for example, by comparing the nudge designs discussed by Caraban et al. (2019) to the critical designs discussed by J. Bardzell et al. (2014).

2.5.4 Designing socio-technical systems critically

In this subsection, I discuss CD within the context of STSs. Essentially, I apply the theory in the previous sections to STSs. First, I present questions related to STS that designers may focus on in CD. Then I present related work where designers have used critical and speculative designs to explore or highlight questions related to STSs.

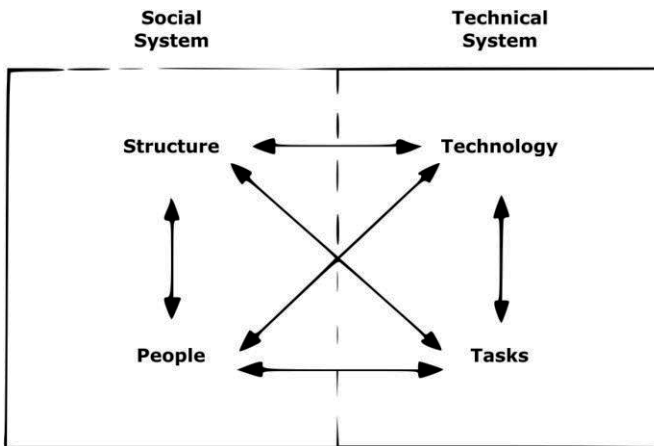


Figure 7. Socio-technical system. Adapted from Boström & Heinen (1977).

STS comprises two jointly independent but correlative *interacting* systems—the social and the technical (Bostrom & Heinen, 1977). The technical side concerns tasks, processes, and technology needed to transform inputs into outputs (Figure 7.). The social side is concerned with the attributes of people (e.g., attitudes, skills, values), the relationships among people, reward systems, and authority structures. Any system with interacting social and technical subsystems can be considered an STS (Bostrom & Heinen, 1977). For example, workplaces, transportation systems, and social media platforms may be considered STSs.

STS design considers hardware, software, personal, and community levels and requirements on each level (Whitworth, 2009). For example, STS design considers how communal requirements such as order, freedom, and openness are supported by and balanced with individual users' requirements, software requirements, and hardware requirements (ibid.). Examples of STSs, relevant to my thesis, where designers should consider all the levels to avoid disasters include Facebook, Twitter, chat rooms, and online commenting. For example, designers of social-media sites need to consider if webservers will be able to serve the design to a large number of people in a given period. Overall, a successful STS can only be achieved by jointly optimizing both the technical and social subsystems (Badham et al., 2000; Whitworth, 2009).

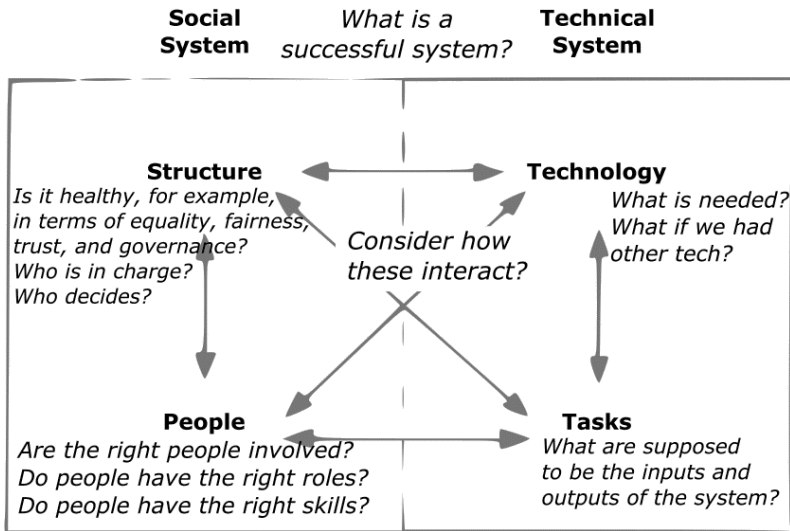


Figure 8. Some questions on STSs that CD of STS might try to provoke people to ask. I adapted the non-italicized parts from Boström & Heinen (1977).

Adapting the goals of CD (see Sections 2.1. and 2.3.) into the STS design context, CD into STSs would presumably seek to challenge whether the present STSs have healthy social structures, roles, and rights (see Figure 8. above). It would also presumably seek to transform how the system parts, their interplay, and the social behavior are thought about and draw attention to possible (hidden) issues in them.

In creating a critical STS design, one would presumably twist one or more of the parts of some present STS. For example, one could propose a slightly weird social structure for online news commenting and how it would be supported to draw attention to the existing structure and make people think critically about it. Further, as I subscribe to the view that CD is not about creating science fiction (see Section 2.2. and (J. Bardzell et al., 2014)), I presume that CD of STS seeks to ensure that the designs are grounded in the real world. For example, a critical design proposal for social media STS that could not run on today's hardware may be too much like science fiction to take seriously. Accordingly, to help ensure that people would not disregard my critical designs as science fiction, I kept technical feasibility considerations in mind when creating them (as opposed to not considering them at all).

Next, I present related work that I interpreted as CD into STS. As the first example of related work, Rynning (2017) explored combining graphic design, social app visual design, and visual identity branding with speculative design. Rynning presented three examples of students' speculative graphic design projects. For

example, one student project is Feel, The True Feeling Social Network. The project examined the seemingly false perfection people are exposed to through Facebook. The designers proposed that when people post to Feel, their true feelings are expressed as a background color when writing comments and incorporated in posts through graphical elements such as colored borders. In addition, their true feelings would be measured by a smartwatch they need to wear while posting to Feel.

The following example of related work also presents a critical STS design. Van Kleek et al. (2016) applied a speculative design approach to explore tools that assist in pro-social forms of online deception. The researchers interviewed people concerning five designs. One of the designs was lieCal, a tool that automatically generates fake calendar appointments to act as excuses on behalf of the user, optionally including friends in the deception and strengthening alibis by posting on social media. The researchers reported that their study resulted in a better understanding of the design space, how online deception might occur, and what factors lead to it.

The next example of related work focused on technology's role in social behavior more broadly but, like the first publication in the thesis, involved interviews of people who might be considered domain experts. Wong et al. (2017) created a design workbook featuring 15 critical, speculative, and fictional designs that would heighten privacy experts' awareness of potential privacy issues in technologies. Their designs included, for example, NeighborWatch Pro, an identification system that automatically detects and flags "suspicious people" who enter a neighborhood, and TruWork, an implanted chip that allows employers to keep track of their employees. The researchers report that many of their study participants (students who had taken courses on social aspects of technology) had generative visceral and affective reactions to the designs. The researchers found this a good thing for eliciting values reflections. The researchers concluded that critical designs, speculative designs, and design fictions presented in a design workbook could help institutions look around corners, an essential component of privacy work. The study illustrates how researchers can use designs that violate design and cultural conventions to have illuminating conversations with experts that result in understanding their values and knowledge about possible futures that may be of value to institutions and organizations.

In summary, in subsection 2.5.4, I explained what STSs are and what questions could be focused on in CD of STSs. I also discussed that few studies showcase CD into STSs, but the work is possible and has several potential benefits. The related work illustrated critical and speculative design's potential to help researchers to dive

deep into a given domain. The related work also illustrated different ways to use critical and speculative designs in user studies. I use designs in my research similar to the related work.

2.6 Summary

In my understanding, CD, rather than being a clear theoretical framework for analysis or creation of critical designs (i.e., a recipe for criticality), is more about positionality, attitudes, and values in design. CD emphasizes offering alternative perspectives and challenging the status quo, and forgetting the constraints of the commercial sector, (potentially) at the cost of discovering practical solutions. CD of STSs critiques the present ways of supporting social structures, roles, and peoples' rights in STSs. They are critiqued by introducing a suitable number of twists (i.e., nonobvious changes) on what could be considered a standard STS design and showing it to people to provoke them to discuss, reflect, consider new perspectives, imagine alternative behaviors, and appreciate STS complexity. To this end, this thesis offers knowledge on what the standard design space is and how it can be tampered with in CD. Furthermore, previous research and this thesis illustrate that showing critical designs to people and reflecting on them can result in new insights (e.g., regarding the design domain) and opportunities for other designs and studies. Nevertheless, the value and potentials of CD are not widely understood, and CD on software and STSs is still rare, which represents a research gap.

3 THE CHALLENGE OF MODERATING ONLINE DISCUSSION

In the following sections, I expand on the introduction and explain why CD is suitable in the context of UI interventions in uncivil online news commenting. I also describe the problem of uncivil online news commenting and current approaches to mitigating it in more detail.

3.1 The difficult problem of uncivil online news commenting

Uncivil online news commenting (i.e., unnecessary use of impolite, insulting, and toxic language or comments that, for example, deny opinion expression from others (Coe et al., 2014)) is a problem. The incivility harms users, moderators, and news sites and undercuts the value commenting can have to societies (as a form of public participation) (G. M. Chen, 2017; Rantasila et al., 2022). Further, to explain the phenomenon in more detail, I note that it is predicted by several factors. These include, for example, hard news topics (e.g., politics) (Coe et al., 2014), where people tend to have fights about controversial issues; existing uncivil comments (Ziegele et al., 2018), which can lead to a spiral of negativity; and the anonymous nature of comments, and hence it being easier to attack others without consequences (Nitschinsk et al., 2022). Additionally, when commenters possess contrasting political identities, commenting may be more likely to be uncivil (Rains et al., 2017).

However, it is not only the average users' behavior that explains why commenting can sometimes be uncivil. It is known that some commenters have psychopathic (i.e., high impulsivity, thrill-seeking, and low empathy and anxiety) and Machiavellian (i.e., manipulative) tendencies and enjoy leaving uncivil comments (Kluck & Krämer, 2020; Saresma et al., 2022). In addition, there may be "trolls," that is, users who are there to trick the other users into wrestling with them or with each other (Hardaker, 2010; Paakki et al., 2021). Because of all these factors or the lack thereof, the commenting may be horrible or pleasant regardless of the discussion technology and moderation efforts. For example, if the "trolls" find a new playground, the overall quality of the comments may suddenly improve.

At the same time, the abovementioned factors are hard to control without also resulting in major drawbacks. For example, suppose the solution is to enable only the paying subscribers to comment, where the thinking goes, ‘The trolls are not going to pay to get to troll.’ In that case, this solution also prevents other people unwilling or unable to pay from commenting. In general, every action taken to mitigate uncivil commenting has potential drawbacks. To illustrate, I adapt some of the social requirements and their tensions from Whitworth's STS requirements model (2009):

- *Order*. Enforce strict rules (e.g., on what words or topics commenters are not allowed to discuss)—This may decrease the freedom to discuss different topics, such as those relating to complex issues.
- *Freedom*. Increase the freedom to discuss different topics—May decrease the ability to enforce strict rules.
- *Transparency*. Increase the transparency of commenting (e.g., by asking the users to prove who they are)—May make the users less shielded and possibly more reluctant to comment.
- *Privacy*. Shield the users (e.g., by allowing them to remain anonymous)—Decrease the transparency of commenting.
- *Synergy*. Make the commenters build something together (e.g., arguments for and against something)—May disallow relaxed, open discussion and the ability of the community to show self-governance.
- *Morale*. Increase the need for commenters to police and take care of the commenting section—The commenters may have less time left to comment.

According to Whitworth (2009), the requirements of *order* and *freedom* on the social level of an STS correspond to reliability and flexibility on the HCI level. However, the two requirements are difficult to achieve simultaneously in the comments because it is difficult to define the right balance between them and what that looks like in the comments (Diakopoulos & Naaman, 2011; Masullo Chen et al., 2019).

To further discuss the trouble in balancing *order* and *freedom*, I discuss comment moderation practices and their relationship to the two requirements in the following. Rantasila et al. (2022) combined existing classifications of moderation into a single framework to provide a concise way to think about current approaches to comment moderation. The researchers classify moderation approaches from four perspectives: intention, form, scale, and specificity of moderation.

The intention of moderation ranges from governing to guiding commenters. (Here, the governing intention rhymes with imposing order). Concrete moderation practices range from hard to soft in form (e.g., removing comments vs. displaying

commenting guidelines) (c.f., imposing *order*—protecting user *freedom*). The scale refers to how much material moderators moderate and ranges from small-scale moderation done by community members to large-scale work outsourced to moderators who process hundreds of comments. Finally, specificity refers to the degree to which socio-technical contexts are accommodated (e.g., a moderation solution that works in most contexts vs. one that works in a particular context).

Regarding specificity, a moderation practice that, for example, removes all posts containing the word “f***” appears highly reliable (c.f., *order*, a strict rule) but also highly inflexible, affecting the *freedom* to comment about complex or sensitive topics. In contrast, highly flexible moderation may appear highly unreliable; it can appear to users that sometimes they are allowed to comment about some controversial topic and other times not, for example. Further, the intention to guide commenters is associated with soft moderation actions conducted in a context-specific manner. The guiding intention could also involve discussing with commenters to establish trust that moderation balances *order* and *freedom*. However, the guiding and trust-building moderation approach is difficult to scale, expensive, and emotionally challenging since moderators need to deal with troublemakers, negativity, and conflicts.

Next, regarding the *transparency* and *privacy* requirements (Whitworth, 2009), one of the main explanations identified for uncivil online news commenting in the literature is that online environments tend to reduce behavioral constraints (Lapidot-Lefler & Barak, 2012; Suler, 2004). People can engage in behaviors online that they would not engage in during face-to-face interactions. Suler (2004) refers to this phenomenon as the “online disinhibition effect.” Suler proposed that online disinhibition is influenced by several factors that differentiate online interactions from face-to-face interactions—for example, asynchronicity, relative anonymity, and invisibility of the interaction partners. Empirical studies have confirmed that many of these characteristics indeed increase both uncivil behaviors (e.g., (Lapidot-Lefler & Barak, 2012; Lowry et al., 2016; Rösner & Krämer, 2016)) and experienced online disinhibition (Wu et al., 2017). However, the level of disinhibition varies significantly across individuals (Stuart & Scott, 2021; Suler, 2004). All of this seems to suggest that increasing the *transparency* of commenting would be wise. However, research also suggests that user *privacy* has several potential benefits. For example, users may want to be shielded from threats, have more control over personal information disclosure, and lower the barrier to new relationships (see, e.g., (Kang et al., 2013)). These are all valid needs, especially considering that the threats are not imaginary: some users enjoy hurting the other users (Kluck & Krämer, 2020; Saresma et al., 2022).

Next, regarding *synergy* and *morale* (Whitworth, 2009), online news commenters are known to be motivated to comment to create things together that they could not create alone (i.e., synergy). These additional benefits include, for example, enjoyment, understanding between people, and understanding of the world (Springer et al., 2015). Simultaneously, the commenters have many seemingly more self-focused motives, such as expressing their feelings or establishing their identity (ibid.). As a whole, the motivations to comment seem to match open discussion systems where everyone can say what they want and interact with whomever they choose. However, at the same time, the user behaviors might be easier to control and monitor, and the behavior might be more civil in limited commenting systems (Bossens et al., 2021).

While the abovementioned issues and challenges may discourage “conventional” intervention design efforts, they only make the area more opportune for CD. CD emphasizes asking questions rather than finding practical solutions, and there seem to be many questions to explore. Additionally, while CD can result in designs with critical and practical functionality, CD does not have to result in working solutions.

3.2 Existing approaches to digitally mitigate uncivil online news commenting

The cost, scalability, and emotional challenges of moderating uncivil online news comments might be addressed by algorithmic solutions or modifying the commenting UI. Bossens et al. (2021) examined the impact of interface design on the civility of online news comments. The system they designed asked users to comment on a statement by a political figure featured in the news article. As a result, the comments were more civil than in a control where the researchers asked people to comment on the news article. However, the researchers noted that the system may limit public political participation and discussion. This follows what I stated above. If commenters are required to work together on something, relaxed and open discussion is likely disallowed in the process. In an extended abstract, Bossens et al. (2022) reported they developed the design further by adding, for example, a machine learning tool that notified the commenter if they used offensive or rude language. The researchers reported that some of the participants who used the new system were frustrated that the machine learning tool indicated something as not well-argued, while they thought it was. The users were also worried the system would limit the discussion and lead to the formation of filter bubbles.

Another example of machine learning-based solutions is the Perspective API developed by Jigsaw (owned by Google) (Jigsaw, 2017). It can detect some toxic writing, which can be configured, for example, to trigger an alert that tries to influence the writer to change their writing. El País' commenting system was incorporated with the API and it is reported to have had a moderately positive impact on the quality of the discussion (Delgado, 2019). Other news websites have reportedly achieved moderately positive results with alerts triggered by the API (Simon, 2020). I note that this approach arguably preserves open participation at the cost of effectiveness in reducing incivility.

Moving to other UI-based solutions, the Norwegian Broadcasting Corporation has incorporated custom-built quizzes to confirm that the user read the news article before commenting (Grut, 2017). It is reported that the good thing about this approach is that it reduces uncivil commenting. However, the bad thing is that it reduces brief civil commenting and brief responses to comments, where users do not bother to fill out the quiz. Additionally, building quizzes that are hard enough but not too hard to complete is tricky.

Another example of a UI-based solution comes from (Seering et al., 2019), who demonstrated that the tone of commenting could be manipulated toward positivity by having the user complete a CAPTCHA with positive images before commenting. CAPTCHAs are tests where the user needs to click the images containing, for example, school buses to prove they are not a robot. According to the researchers, the good thing about this approach is that it is simple and can be easily deployed on many websites. However, the researchers state that the problem with this approach is that the covert manipulation of users seems unethical.

UI-based approaches to incivility are also being studied outside the context of news commenting. For example, Y. Wang et al. (2014) developed a web-browser-plugin to prevent users from making impulsive disclosures on Facebook by reminding them of the audience. Based on findings from a six-week field trial, their participants tried to minimize the chances that they would offend others. Web-browser-plugins, however, need to be installed by the users, which is arguably a significant drawback to their widespread use.

While the above suggests many potential UI interventions in uncivil online news commenting, little is known about how the users perceive any of them and what users consider to characterize a good UI intervention. Presumably, attempts to influence user behavior should match the user's level of motivation and ability to act (Caraban et al., 2019; Fogg, 2009) and be transparent (Bovens, 2009). However, what achieves them is poorly understood.

Furthermore, as subtly reflected in the above paragraphs, the technological interventions are typically considered universally applicable, one-size-fits-all solutions. The same interventions would be done in all circumstances where impolite and insulting words or expressions are detected. However, online discussion uses a rich linguistic repertoire, and other users' civility assessments are contextually defined. Prior messages serve as a contextual resource for making sense of new messages (Arendholz, 2013; Linell, 2001). For example, Kluck & Krämer (2020) found that those who admitted to insulting or mocking others usually explained that they did so in the context of criticizing others for misbehavior. Shmargard et al. (2022) found that repeated incivility (i.e., several "messages that are unnecessarily disrespectful to the discussion or its participants") receives fewer up-votes if nearby comments are civil compared to when they are uncivil. To my knowledge, there are no studies on the relevance of previous comments on UI intervention design.

3.3 Supporting users' emotion regulation with computational affect labeling

In the thesis publications, I critically explore one approach to mitigating uncivil online news commenting: supporting users' emotion regulation with the help of automatic identification of emotional elements. Based on research in emotion psychology, many issues with digital media discussion culture are related to emotions and emotion regulation. Regulating one's emotions and mood is necessary practically for every area of life (Gross, 1998), but it is challenging in computer-mediated textual communication (Syrjämäki et al., 2022). Furthermore, the lack of nonverbal cues in textual communication may deteriorate the ability to control emotions and empathize with others (Syrjämäki et al., 2022; Walther, 1993).

The concept of implicit emotion regulation has recently been discussed in the literature (Syrjämäki et al., 2023; Torre & Lieberman, 2018). In contrast to explicit emotion regulation, which requires a conscious effort to suppress emotional responses, implicit regulation is effortless and potentially automatic. Therefore, implicit emotion regulation appears promising as a design concept (Syrjämäki et al., 2023). The ability to regulate emotions might be enhanced by affect labeling: for example, simply making the emotionally loaded elements in a message more perceivable (Syrjämäki et al., 2023). In the thesis publications, I consider labeling as an inspiration for design rather than a boundary. I explore various tactics to make the users more aware of the emotional elements in the messages.

Apart from the thesis publications, few studies explore affect labeling interventions in uncivil commenting. Linhares de Carvalho et al. (2021) interviewed 18 university students about their perceptions of four proposed UI mechanisms for guiding users to emotional self-reflection when reading and commenting on online news articles. The interviewees commented about the ease of use, usability, usefulness; feeling of control, censorship, intrusion; an unintended consequence of angering users; and level of trust towards the service. The study concluded that users do not want an intervention to interfere with fast-paced interaction in online news commenting. Syrjämäki et al. (2023) investigated the perceived effects of a UI intervention aiming to support online news commenters' emotion regulation using the experimental vignette methodology. The researchers found that the labeling intervention was assessed to evoke positive emotions and mitigate uncivil behavior when compared to a control condition.

3.4 Summary and opportunities for critical design

To summarize the above, mitigating uncivil online news commenting is difficult. Both the existing machine learning-based moderation approaches and changing the commenting system have potential drawbacks. Further, little is known about what the users think about technological interventions and what they require of them. This represents a large research gap.

There are several opportunities for CD here, many of which I explore in this thesis. Perhaps most obviously, CD could question why certain social behaviors and conventions are (not) supported, afforded, or instructed in commenting. Additionally, CD could ask if interventions should lie beyond what is familiar to designers and users.

Still, it is crucial to recognize that no CD project can question everything, as this is not feasible. Even critical designers must assume that many things are true and desirable. For example, based on the literature, I believe that having humans guide commenters is expensive and that technological solutions should be explored. Further, I believe that open participation in commenting should be protected, and I do not intend to propose that only some people should be allowed to comment on the news.

4 RESEARCH PROCESS AND METHODOLOGICAL SUMMARY OF ARTICLES

This chapter starts with a discussion of the general research approach and overall structure of the research process. The chapter then presents an overview of the studies, research methods, and key results. The chapter also presents the methodology for answering the thesis research questions.

4.1 Research approach

I follow the Critical Design (CD) methodology to address and study the research questions empirically. I subscribe to the view that CD is a Research through Design (RtD) methodology that aims to foreground the ethics of design practice, uncover potential hidden agendas and values, and explore alternative design values (J. Bardzell & Bardzell, 2013). RtD is about using design methods, practices, and processes to generate knowledge (Zimmerman & Forlizzi, 2014). I also subscribe to the notion that critical designs can simultaneously critique and function practically (S. Bardzell et al., 2012; Ghajargar & Bardzell, 2021). I speculate that news websites could implement my designs. Further, the research is primarily explorative and expansive (see (Krogh et al., 2015)). I focus on mapping out the design space, but not solely. I also develop some of my designs one step at a time.

Besides the CD methodology, I use a qualitative and quantitative online survey methodology to find statistical associations between design ratings and design and background variables (e.g., commenting frequency and views on comment moderation). Measuring how practical an audience expects a design to be and what background variables are associated with the expectation may provide insight into the design's potential to engender discussion with a similar audience. This is because a design's critical function can sometimes be connected to its practical function (Ghajargar & Bardzell, 2021). People may not bother to think about a critical design if they believe it is useless (J. Bardzell et al., 2014).

Further, the survey methodology allows me to gather numerical design ratings and written comments and reactions to each design. Qualitative analyses of

comments and reactions can provide an understanding of the effects of specific design features (J. X. Chen et al., 2021; Van Kleek et al., 2016; Wong et al., 2017). Qualitative analyses can also provide insight into how people approach and discuss a design (based on all the previously cited studies).

The critical design artifacts I create and use in the research are pictures that illustrate possible UI interventions in uncivil online news commenting. I use 18 illustrations in total. I use the illustrations to elicit reactions and comments from study participants. This was similarly done by (J. X. Chen et al., 2021; Van Kleek et al., 2016; Wong et al., 2017).

In creating the designs I utilize the theory of affect labeling (Torre & Lieberman, 2018) as a source of inspiration. The core idea of affect labeling, naming emotions, is adapted in the designs as follows: a machine or other users indicate for the user the presence, placement, quality, or strength of emotions in their comments. The exception is the *Evaluate* design presented in Publication II and III, where the user is asked to indicate their own emotional state. Further, I also draw inspiration from other emotion regulation strategies (e.g., avoidance, raising self-awareness, suppressing expressions) mentioned by (Gross, 1998; Yoon et al., 2019). I use these strategies in Study II to help to ensure that the designs I select to study represent different ways of supporting emotion regulation. However, please note that the designs are not intended to be as simple and clear representations of the theories as possible.

4.2 Overview of the research process

In Table 4. below, I illustrate the relationship between the studies, publications, and research questions. Study I was a deep dive into affect labeling UI interventions in uncivil online news commenting and the meaning of CD in the context. It was the most practice-based and emergent of the studies; design first, ask questions later, like a “voyage” and “return” (Gaver et al., 2022). The study involved analysis of the criticality of designs created and analysis of what Finnish journalists thought about the designs. In the two subsequent studies, I used more advanced versions of designs I created during the first study and entirely new designs. The two subsequent studies were more like quests for answers to questions than the first study.

Table 4. The relationship between the studies, publications, and research questions.

Study I: Analysis of critical designs based on designer reflection and interviews with Finnish news media experts	Study II: International online survey with online news commenters on critical UI intervention designs		Study III: International online survey with online news commenters to find how discussion context affects alert design evaluation
Publication I: Applying Critical Voice in Design of User Interfaces for Supporting Self-Reflection and Emotion Regulation in Online News Commenting	Publication II: Online Survey on Novel Designs for Supporting Self-Reflection and Emotion Regulation in Online News Commenting	Publication III: User-centred quality of UI interventions aiming to influence online news commenting behaviour	Publication IV: Evaluating Alerts to Impolite Online News Commenters: The Impact of Previous Commenter's Politeness and the Form and Amount of Guidance
RQ1: What characterizes the design space for critical design of user-interface interventions aiming to influence online news commenting behavior			
		RQ2: What characterizes a high-quality user-interface intervention aiming to influence online news commenting behavior?	

Study II followed a mixed methods approach in the analysis of online survey data. I compared survey respondents' ratings and expectations of specific designs and observed the effect of respondents' backgrounds. In Study III, I qualitatively analyzed responses to the same online survey to identify notions of quality across the designs.

Study III expanded into a territory I had not touched on in the previous studies: the effect of the application context of UI interventions on their perceived quality. While this jump was not strictly based on the previous findings, it is an example of thinking critically about the designs and exploring the design space.

I describe the studies and illustrate the designs in Sections 4.3-4.5.

4.2.1 Research ethics

The research adhered to the guidelines outlined by the Finnish Advisory Board on Research Integrity. As per the guidelines, conducting an external ethical review of the research plans before implementation was considered unnecessary for the following reasons. The physical integrity of participants was not interfered with, and the participants were not exposed to powerful stimuli. Taking part in the research did not constitute a departure from the principle of informed consent. No psychological harm exceeded the limits of everyday daily life was identified. All participants were at least 18 years old during the studies. (Kohonen et al., 2019)

In all of the studies, informed consent was obtained from all participants before participating. In addition, the participants were informed that participation was voluntary, and they were free to withdraw at any moment. The participants were also provided information regarding the research aims, procedures, project funders, data management, pseudonymization and anonymization, and how the research results will be published and disseminated. Additionally, the participants who were recruited from Prolific (an online platform for participant recruitment) were compensated +£6.7/hr (+\$8/hr) on average and they were informed about the rate beforehand.

Digital materials such as survey responses and interview recordings were stored on computers provided by Tampere University and on network drives managed by Tampere University. Only researchers affiliated with the project were given access to the research materials. Written consent forms were stored on campus in a locked office, in a locked cabinet.

4.3 Study for publication I: An analysis of critical designs based on designer reflection and interviews with Finnish news media experts

To understand what kind of AL UI intervention designs to uncivil online news commenting are slightly strange, in the first study, I studied existing designs, and conducted design generation and selection, and analysis of the criticality of the selected designs. While Publication I covers the details, the following summarizes the methodology and key findings of Study I.

4.3.1 Preparation work

As a first step, before beginning to ideate designs, I examined in 2019 what the current online news commenting UIs were like. I looked at the commenting systems used in the 15 most popular—by traffic—news websites in the U.S. at the time. Further, as the research took place in a Finnish university, I examined the commenting systems used in the four most popular by traffic (according to Alexa Internet analytics in 2019) Finnish news websites at the time (tabloids *Ilta-Sanomat* and *Ilta-lehti*, national newspaper *Helsingin Sanomat*, and Finland’s national broadcaster *Yle*). (In retrospect, also other European, and Asian and African news websites might have been good to examine). Based on the examination, I generated

lists of cultural conventions, UI conventions I frequently encountered, and UI features I came across only occasionally. The following are updated and clarified versions of the lists I created. Please note that the listed items represent only my observations and are not featured in previous peer-reviewed publications.

Cultural conventions. Commenters do not appear to customarily

- Keep face (i.e., like at a workplace).
- Ensure that their comments are well-written and clear.
- Reply to those who reply to them.
- Cite sources for their claims.

Additionally, concerning news organizations and commenting

- Journalists do not usually participate in commenting in any way.
- Comment moderators very rarely participate in commenting.
- Commenting sections are sometimes open when the news article covers controversial issues and sometimes closed when the topic is seemingly uncontroversial.

UI and interaction design conventions, or things that I saw on most of the news websites:

- There is a text area for writing.
- The comments are only plain text and do not contain pictures, videos, or attachments.
- The structure of the comments is tree-like.
- The style of the comments (i.e., color, font, border-width, padding, margin, etc.) does not reflect what the comments are about and does not change between news articles or topic areas.
- Usernames (some form of) are shown.
- The date a comment was posted is shown.
- A comment counter is near the comment section (i.e., counting how many comments commenters have posted).
- There is an option to sort the comments from latest-oldest, most liked-least liked, or by replies.
- There is an option to reply to a comment.
- There is an option to expand the comment thread when a thread has received many replies.

- A comment policy (aka commenting guidelines) or link to a comment policy is placed close to the comment section.

Additional UI features on news commenting sections:

- The comment section is not shown by default but can be opened by pressing a button or by clicking a link.
- There is an option to show the new comments that were posted after the comment section was first opened.
- The news organization's or users' favorite comments are shown above others or in a dedicated section.
- A character limit is enforced in commenting.
- The commenters can use rich text (bold, underline, italics, strikethrough, font-styles, lists, indents, spoilers, quotes) and graphic smileys.
- The commenters can post links.
- Profile pictures are attached to comments.
- There is an option to block a commenter.
- Each commenter has a public profile that shows their commenting history.
- There is an option to like or dislike a comment (or equivalent).
- There is a list of the users who pressed like (or equivalent) on a comment.
- The country, state, or city where the commenter is from is shown.
- A comment counter is shown under the news article's title.
- There is a list of active commenters above the comment section.

Based on my observations, I concluded the largest U.S. and Finnish news websites used basically the same commenting UI—and they still seem to do in 2023. While there are some differences between the UIs, I would not classify them as major. I am, however, uncertain whether or not the similarity is caused by the fact that this is simply the best way to allow news readers to comment on the news. Nevertheless, I interpreted the similarity to mean that creating unfamiliar or strange-looking UI and interaction designs will not take much. Additionally, there appeared to be much room for design exploration, and that CD or any other creative design process could easily result in innovations.

4.3.2 Critical designs

I sketched approximately 60 concept ideas based on several idea-generation sessions. Next, based on an iterative selection process I conducted with my colleagues, four

design artifacts were selected to be analyzed in the study. More details on the idea generation and selection process can be found in the publication.

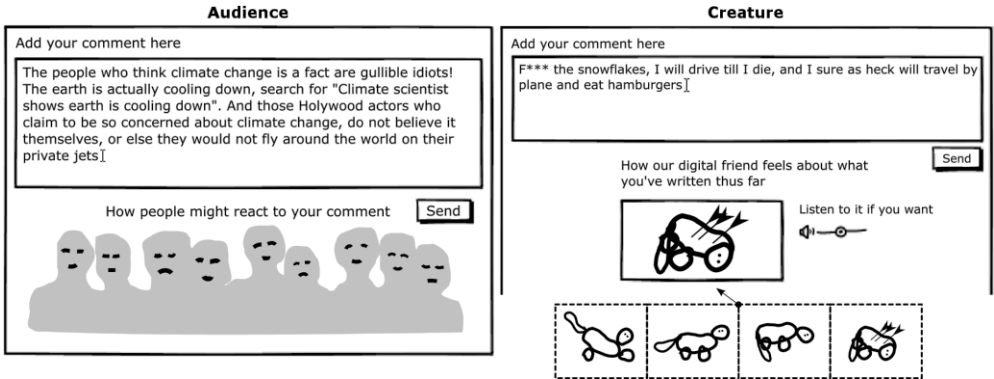


Figure 9. Left: The *Audience* as it appears for a comment writer. The audience shows the anticipated emotional reactions to the user's writing. Right: The *Creature* as it appears for a comment writer. The design features a virtual animal that thrives or suffers according to the user's writing.

The *Audience* is an animated graphical element illustrating probable emotional reactions to a comment or discussion thread. As the user writes a comment, an array of abstract animated anthropomorphic figures would begin to form as emotional elements are identified (see Figure 9. left). With the *Audience* design, commenters can predict how readers might feel about their comments. Additionally, the *Audience* would appear above the comment section to display responses to all published comments.

The *Creature* (see Figure 9. right) would work much like the *Audience*, but with a more direct take on emotions as it reduces the scale of emotions to one dimension (troubling–pleasant) and intends to represent it through the well-being of the creature. The *Creature*, like *Audience*, would also be placed above the commenting section (i.e., visible to comment readers).

Regret

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

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

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!

Figure 10. Left: The *Regret* mockup. Top: After publishing a seemingly uncivil comment, the user can regret their words. 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.

The *Regret* proposes to change the dynamics of discussion for the better by allowing the writer to regret their choice of words explicitly and publicly. In Figure 10. (top-left), John Smith has just published a nasty comment; a notification appears, allowing him to regret his words. Clicking the button would also cause the other users to see the writer's regret (Fig 10., bottom-left). The *Regret* proposes 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 might not be desirable.

The *Promise* proposes to force the user to make an explicit promise to control their emotions. In Figure 10. (right), the user is forced to promise good behavior before commenting based on predefined text and a large checkbox. If the user writes nastily after promising, a note will appear under the text area, "Are you sure you are keeping what you promised?" The design addresses the issue of users not taking the time to consider what they are about to write and how.

4.3.3 Criticality in the designs

Publication I outlines and discusses the various manifestations of criticality in the four designs, based on three of the four criticality dimensions identified by Bardzell et al. (2014). Hence, apart from the dimension, "Reflectiveness", the following is only a summary of what is written in the publication.

The "Reflectiveness" criticality dimension (J. Bardzell et al., 2014) focuses on triggering a reflective response, moving the user to a more self-aware stance. However, the dimension also includes the idea of the design itself embodying reflectiveness, for example, by revealing itself as a rhetorical device by containing obvious "twists" on the standard or expected design.

The first part of the "Reflectiveness" dimension is evident in all the designs presented in Publication I. They all aim to move the user to a more self-aware stance. However, the second part is only evident in the *Creature* and *Promise* designs. The *Creature* and *Promise* clearly contain satire (i.e., the use of humor, irony, exaggeration, or ridicule to expose and criticize people's stupidity or vices). The *Creature* and *Promise* may be seen to ridicule the user, telling them, "look at what you did!" The *Audience* and *Regret* do not reveal themselves as rhetorical devices as clearly. Still, they embody irony, in the form of internal conflict regarding implementation: the designs ask the user, "am I a real solution or not?"

The "Enhancing Appreciation" criticality dimension (J. Bardzell et al., 2014) focuses on highlighting the role of design in addressing socio-cultural issues. This dimension is most evident in the *Audience* design. The design emphasizes the distinction between text-based commenting and in-person discussions, and the anonymity of the audience underscores the presence of a silent majority.

The "Proposals for Change" criticality dimension (J. Bardzell et al., 2014), involves suggesting alternative perspectives. It is most evident in the *Regret* design, which proposes the users should publicly regret negative comments. The design proposes change in the commenting culture. Additionally, there is a proposal for change in the sense of user-publisher power dynamics in the *Promise* design. In the design, 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.

Considering "Changing perspectives" criticality dimension (J. Bardzell et al., 2014), the *Creature* design represents the emotional tone of text through the well-being of a virtual animal, offering a new perspective on what can be used to represent the emotional tone of text. The design raises questions about ethical implications and the portrayal of suffering due to hurtful comments.

Overall, the designs aim to improve understanding of the role of design in addressing uncivil commenting, proposing changes that challenge conventional perspectives and cultural norms.

4.3.4 Study details

The criticality of the four selected designs was analyzed in two stages, before and after interviews of Finnish journalists about the designs.

Analysis of the criticality of the designs before the interviews: The analysis was based on the framework proposed by Bardzell et al. (2014) and was conducted by me. The analysis is presented in its section in the publication. The interviews were carried out in May/June 2019.

Interview participants: 10 Finnish journalists (two females, eight males) with experience in moderating online discussions in news media or who were involved in developing solutions or policies for moderation and maintaining appropriate online discussion quality. All the interviewees represented Finnish news media organizations. The gender imbalance of the interviewees is regrettable.

Interview agenda: To understand what thoughts, feelings, and ideas the journalists have about the designs. To understand thoughts on if the designs would trigger users to reflect about their behavior or about the designs. Also, to understand concerns over the designs and other potential effects on user behavior.

Interview procedure and data gathering: The interviews were conducted by Heli Väättäjä, the third author. The interviews were audio-recorded and transcribed for analysis. The paper only covers the interview data related to the four designs. The interviewees were presented with the four selected designs in a randomized order. 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 the thoughts on the design. Brief design evaluation forms were used to aid the thinking aloud.

Analysis of the criticality of the designs after the interviews: The analysis followed a bottom-up approach. First, themes were identified in transcribed interview recordings, and themes were refined. Second, the data was used to analyze the designs' potential to have critical functions for online news commenters.

Contributed to: RQ1 and RQ2.

4.3.5 Publication I: key findings

The first study provided many perspectives on how critique could be manifested in this problem area and offered valuable insights into the social acceptance and possible effects of the designs. In addition, the artifacts provoked reflection. The following are some of the key findings and insights reported in the publication:

Highlighting the positive instead of removing the negative. Despite the negative connotations of critique in design, critical designs could also focus on positive perspectives. For example, discomfort with the status quo could be displayed in positive or fun ways. One of the interviewees considered that *Creature* had a positive dimension in potentially rewarding the excellent writer.

The risks in accurately predicting what the majority thinks. Based on the comments on the *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. Researchers have raised a similar concern that an algorithm may suppress the voices of minorities (Davidson et al., 2019; Lu, 2019).

Showing what is uncivil can support trolling. One aspect that the design process failed to recognize is that trolls might abuse especially the *Audience*, the *Creature*, and the *Regret*. Modeling and visualizing how badly one writes would help to optimize the text for malicious purposes. The option to add a label of regret could be used ironically or to annoy other users. Therefore, these designs could only be applied in limited contexts and under careful supervision or be supported by other mechanisms.

While the previous key findings were about the designs, the following is about the interviewed Finnish journalists and the media organizations they represented. The interviews reveal an ambivalent position toward the designs: while the Finnish news media experts desired to prevent trolling and uncivil discussion, they did not wish to limit the commenters' freedom of expression. This may be partly explained by the fact that the experts are journalists whose fundamental values include freedom of expression. According to the literature, journalists tend to have an ambivalent position toward uncivil commenting in general (G. M. Chen & Pain, 2017; Løvlie et al., 2018; Wolfgang, 2018). However, the experts appeared also to value and guard the publisher's reputation. They were concerned that the solutions presented might lead to the publisher being questioned in public discussion. This indicates that publishers are not known for experimenting with solutions. Furthermore, publishers' desire to preserve journalistic brands and general conservatism may hinder the identification of solutions that challenge conventions.

4.4 Study for publications II-III: An international online survey with online news commenters on critical UI intervention designs

Study II was an international online survey with online news commenters on UI intervention designs. Publication II presents the findings on design ratings and respondents' written explanations for the ratings. Additionally, the effect of various background variables (e.g., preference for comment moderation) on design ratings is presented. Publication III explored what characterizes good quality in UI interventions. The publication analyzes written first reactions to the designs and written justifications for design choice in a design choice task.

4.4.1 Designs in publications II-III and their criticality

The design work for Study II builds upon the previous study. I developed eight ideas generated during Study I for this study and made them more presentable. While the following describes the designs, more details on the idea generation and selection process can be found in Publication II and III. However, *as the ways the designs manifest criticality were not discussed in the publications, I do so here*. Later, in Section 5.2 I discuss which of the designs are the most likely to have successful critical function.

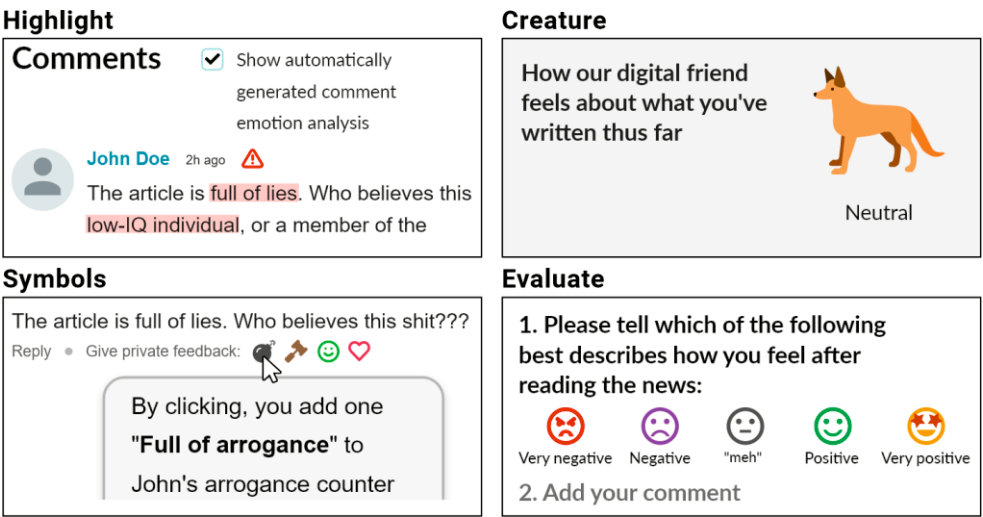


Figure 11. The Highlight, Creature, Symbols, and Evaluate designs in short.

In the *Highlight* design (see Figure 11. top left), the user can view an analysis of the emotions in comments. Upon checking a checkbox, the user can see any negative emotional expressions highlighted in red. The alert symbol is also displayed on comments with strong negative expressions. The design is strongly inspired by the theory of affect labeling (Torre & Lieberman, 2018) and speculates that highlighting negative emotional expressions in comments could calm the users. The proposal is novel and intended to present a new way of reading comments.

I argue criticality is present in the *Highlight* design in the form of “Reflectiveness”, “Enhancing appreciation”, “Proposals for Change”, and “Changing perspectives” as defined by Bardzell et al. (2014) and discussed above in chapter 4.3. The design is intended to trigger the user to reflect and to embody irony (it is not intended to appear as a fully credible, honest solution to commenters). However, it is not humorous or satirical the way the *Creature* is. The design is intended to enhance appreciation about the complexity of negativity in online news commenting and defining when negativity is uncivil. Relating to “Proposals for change”, the design is intended to present a non-sci-fi but somewhat unusual future for the user: doing emotion analysis while reading comments. Lastly, relating to “Changing perspectives”, as the design was introduced as a potential solution for mitigating incivility, the design framed the problem of incivility as an emotion expression and regulation problem, offering a new perspective.

In the *Creature* design (see Figure 11. top right), an animated dog reacts to the emotional tone of a comment as the user writes the comment. The design aims to encourage change by creating an emotional link between the user and a virtual pet dog. In the design, the pet dog is displayed below the text area and described as “our digital friend.” The dog appears happy and ready to play when the user writes positively. When writing neutrally, the dog appears neutral. When the user writes negatively, the dog lays on the floor, keeps its head down, tail between its legs, and faces away.

Criticality is arguably present in the *Creature* design like it was in the previous unpolished version of the design. The difference is it is now intended to satirize impolite commenting *gently*; to paint the user’s impolite commenting gently and humorously in a new light. In contrast, the earlier *Creature* design version was intended to strongly satirize user’s behavior. I changed the design to no longer feature the death of the creature because I estimated that it would make the design too off-putting to online news commenters and that they would not bother to think about the design because of it. After all, in critical designs, “slight strangeness” is key (Dunne & Raby, 2013).

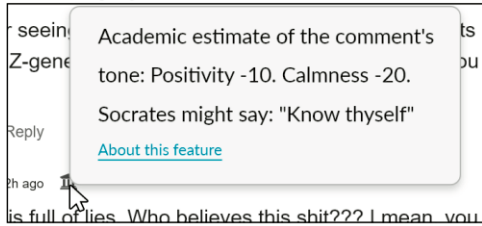
In the *Symbols* design (see Figure 11. bottom left), the user can provide anonymous, private feedback to any of the previous commenters. This is intended to decrease the likelihood of written personal attacks toward other commenters. The design has buttons depicting a bomb, a gavel, a smiling face, and a heart next to every comment. The bomb symbolizes “Full of arrogance”; the gavel “False claim/s”; the smiling face “Well said”; and the heart “Love it!” Additionally, every user’s profile contains a section entitled “Overview of the feedback from other users,” which displays the same symbols and the number of times the user has received these feedback types. The section is intended to be visible only to the user. The design is intended to highlight the possibility of private displays of aggression and agreement in the comments and explore if they could replace public ones.

I argue that criticality is present in the *Symbols* design but perhaps not to as great extent as in some of the other designs. The design is intended to look so humorous and function in such an amusing and witty way that it triggers a reflective response (“Reflectiveness” (J. Bardzell et al., 2014)). The design may be an example of how common UI features and functionality may be exaggerated to the point of critique. Considering “Proposals for change”, while the design proposes a change in user behavior, the proposed future is seemingly not alternative enough for the design to be counted to have this critical function.

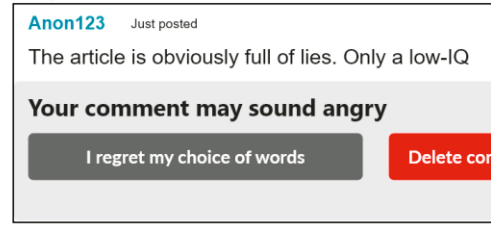
In the *Evaluate* design (see Figure 11. bottom right), the users must first indicate how they feel before they can add their comments. Users can do this by clicking on a smiley face representing their emotional state. It is proposed that naming the emotion could have a calming effect on an angry user. The design is firmly based on the theory of affect labeling (Torre & Lieberman, 2018). Unlike the other designs, *Evaluate* and *Symbols* do not propose that the website publicly evaluates comments for their quality. Hence, if this is an issue, it might display itself in findings on how online news commenters rate the designs.

I believe the *Evaluate* design features less criticality than most of the other designs, or that it is harder to argue that it has criticality. The smiley expressions are intentionally humorous and may reveal the designer is not completely serious with the proposal. Also, it is left ambiguous how the design is supposed to mitigate incivility, and if users’ emotion data is collected for some purpose. This may lead people to speculate about the design, given that the design is first framed as a solution to mitigating incivility.

Philosophy



Regret



Warning



Audience

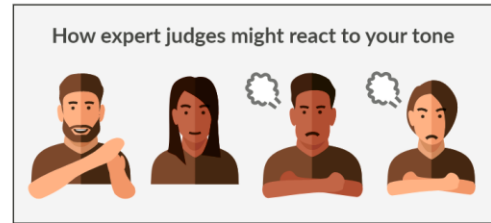


Figure 12. The Philosophy, Regret, Warning, and Audience designs, in short.

In the *Philosophy* design (see Figure 12. top left), problematic comments and comment threads are marked with a university icon. By pressing this icon, a box containing the emotion score for the comment or comment thread and a quote from Socrates, “Know thyself!” is revealed. The emotion score has two dimensions, positivity and calmness. In this design, the automatic evaluation of comments is proposed to be accomplished in a relatively subtle, inconclusive, and ambiguous way.

I argue criticality is present in the *Philosophy* design in the form of “Reflectiveness”, “Enhancing appreciation”, “Proposals for Change”, and “Changing perspectives” (see (J. Bardzell et al., 2014)). Considering “Reflectiveness”, the design features the use of gentle satire to criticize commenters and may trigger the user to reflect on their behavior. Considering “Enhancing appreciation” and “Proposals for Change”, the design aims to enhance appreciation about the complexity of interpreting comments and emotions in them and proposes a slightly strange future where users analyze emotions in the comments. Considering “Changing perspectives”, the *Philosophy* design, like the *Highlight* design, frames the problem of incivility as an emotion expression and regulation problem.

In the *Regret* design (see Figure 12. top right), users’ comments are automatically evaluated immediately following posting. For example, if a comment sounds very angry, the user is notified and offered various follow-up actions below the published comment and by email. The first offered follow-up action is to regret the choice of words, the second is to delete the comment, and the third is to edit it. If the user

selects the regret option, a notification is attached, indicating that the user has regretted their angry words.”

I argue criticality is present in the *Regret* design much like in the previous version of the design. The difference is, I added the delete and edit comment options to make the design less strange. The design is intended to trigger reflection and embody an internal conflict of whether it is a real solution (“Reflectiveness” (J. Bardzell et al., 2014)), propose a somewhat strange future where users regret their choice of words (“Proposals for change”), enhance appreciation about design’s role in the commenting behavior (“Enhancing appreciation”), and offer the perspective that the problem is that commenters do not show regret (“Changing perspectives”).

In the *Warning* design (see Figure 12. bottom left), a notification is shown above the comment section, indicating a description of the argumentation within the comment section (e.g., “10% Hatefulness”). It is proposed that labeling the emotional content of the comment section will assist the user in dealing with overly negative comments. In addition, it is suggested that the design would assist news readers in deciding whether or not to read the comments. The design may be interpreted to ask if even a little hatefulness is enough to warrant a warning.

The *Warning* design arguably features criticality, but it is not as visible or obvious as in the other designs. There is “Reflectiveness” in the design: it features exaggerated graphical elements and curiously specific percentages. The design may be read as critical if it is noticed that it dishonestly claims that it is possible to accurately measure the comments hatefulness etc. Additionally, the use of the term hatefulness in the design implies that the designer thinks news sites should accept at least some hateful comments. However, I do not; I am being dishonest here.

In the *Audience* design (see Figure 12. bottom right), when a user writes their comment, a virtual audience of expert judges reacts to its tone in real-time and is displayed below the text area. The design is intended to create a sense of having a live audience, encouraging commenters to consider their self-presentation through writing. The *Audience* would function as follows: If the user writes in a moderately positive way, some members of the audience appear glad, and others have a neutral expression. If the user writes in a rather negative way, most members of the audience appear angry or frustrated. The design is a new version of the *Audience* design featured in the first study.

I argue that criticality is present in this *Audience* design like in its previous version. The design is intended to gently satirize impolite commenting and appear to have internal conflict regarding if it is supposed to be a solution or not (“Reflectiveness” (J. Bardzell et al., 2014)). The design is intended to propose that commenting is a

more substantial performance than is usually thought (i.e., maybe the user should feel like they entered a stage, instead of sitting alone in some room) (c.f., “Changing perspectives”). The design is intended to enhance appreciation about the complexity of commenting the way it is done today (e.g., one does not know who are in the audience) (“Enhancing appreciation”). Lastly, it proposes a new, slightly strange way of writing comments with the help of the virtual audience (“Proposals for change”).

4.4.2 Study details shared by publications II-III

How the designs were presented. The eight designs were presented using storyboards (or rather low-fi mockups), that is, using series of pictures and text illustrating how a user could use them. The storyboards are included in the publications as appendices. Storyboards were used because they were judged more cost effective than videos and interactive prototypes, which are harder to make and harder to embed in an online survey.

Survey respondents: Recruited from Prolific, with criteria: fluency in English, normal or corrected to normal vision, and a minimum of 70% previously submitted studies approved. The survey received 439 valid survey responses (e.g., click-throughs were discarded).

Survey procedure: Each respondent was shown two pseudo-randomly selected designs in randomized order. The survey questions included a question on the respondent’s first reaction to the design and a broad array of scales (e.g., on desirability, familiarity). Additionally, the survey included questions about the respondent’s background and views on commenting and comment moderation.

4.4.3 Study details for publication II

Statistical analyses were conducted on background statements (e.g., “I tend to reply to others’ comments) and design scale answers (e.g., I feel that the solution is sarcastic or a spoof). A thematic analysis was conducted on written first-reactions to the designs. To increase the validity of the design comparisons, the dimensions in the data were extracted using exploratory factor analyses. This resulted in background factors: view on the situation (on the news site), interest in debate, toleration of incivility, and preference for moderation. The resulting factors on the designs were: instrumental quality (i.e., the degree it serves as a crucial tool) and inappropriateness. Following this, Kruskal-Wallis tests were used to compare the

ratings of the various designs (based on the factor-based scores). Significant effects ($\alpha = .05$) were followed with pairwise comparisons, with Bonferroni correction being used to correct the family-wise Type-I error rate. Then, to investigate background variables' effects on design ratings, univariate linear regression analyses were conducted. Lastly, a thematic analysis of the respondents' first reactions to the designs was conducted to gain insight into the reasoning behind the numerical ratings.

Contributed to: RQ1 and RQ2.

4.4.4 Publication II: key findings

I found that online news commenters did not expect *Audience (V2)* to be useful and were unsure if it is appropriate. This suggests that the design's user experience would begin on the wrong foot if the design was deployed on a commenting platform. However, whether it would improve noticeably when using the design is unknown. *Audience (V2)* was rated significantly worse than *Evaluate*, *Regret (V2)*, and *Warning* in terms of expected instrumental quality (i.e., the degree to which it is expected to serve as a crucial tool) (see Figure 13. below). *Audience (V2)* was also rated significantly less appropriate than *Symbols*, *Evaluate*, and *Warning*. The instrumentality was a factor-based variable based only on responses to positively worded statements. The inappropriateness was also a factor-based variable, but it was based only on responses to negatively worded statements.

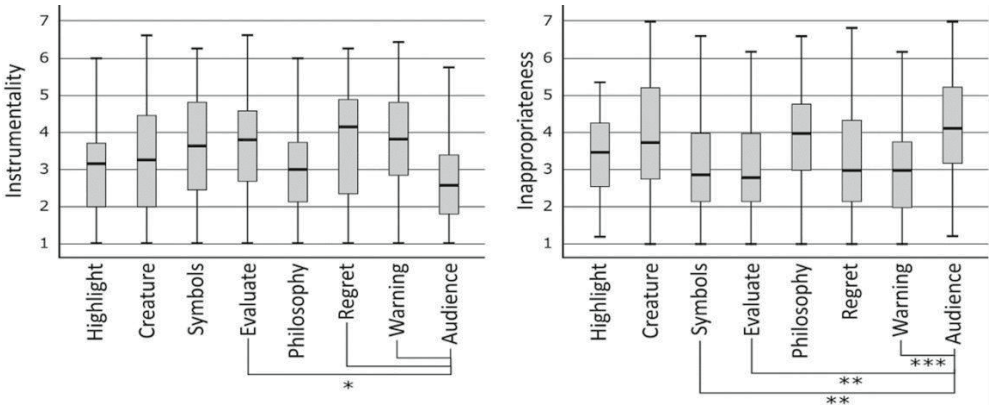


Figure 13. The expected instrumental quality and inappropriateness ratings of the eight designs (V2 designs). The asterisks indicate significant differences according to p-values adjusted with Bonferroni correction: * $p < .05$, ** $p < .01$, *** $p < .001$.

The other statistical key findings are that several background factors predict perceived instrumental quality ratings (see Table 5. below). However, the background variables were not found to significantly predict the perceived inappropriateness of the designs (p -values $> .069$).

Table 5. Results of regression analyses investigating associations between background variables and instrumental quality ratings.

Background variable	R ²	B (95% CI)	F (1, 438)	p
Preference for moderation	.031	0.17 [0.08, 0.26]	13.8	<.001
Not tolerating incivility	.034	0.18 [0.09, 0.27]	15.3	<.001
View on the situation	.014	0.14 [0.03, 0.26]	6.4	.012

Besides the statistical findings, the respondents written first-reactions to the designs are briefly reported in Publication II. In addition, the written first reactions are used to give insight into the ratings. In the following, I demonstrate what the respondents wrote about the *Audience* and *Regret*, as I believe comments on them to be the most interesting of all:

Considering the *Audience*, several respondents expressed that giving the commenter feedback using the virtual audience of experts would cause the commenter to feel overly anxious or annoyed. For example: “I do not want to instantly know that I am being judged before the comment is even posted” and “I would be concerned that it would encourage me to write comments that make the virtual experts happy rather than helping me concentrate on what I am thinking about the news issue.” Further, some respondents noted that “[the feedback] may only serve to encourage some people to carry on their comment further [into negativity].” That said, some expected they would find the feedback useful when composing.

Considering *Regret*, some respondents saw value in the option of adding a label indicating that one regretted their words, for example: “I feel like it would be a good way to redeem the person who sends his angry thoughts as an impulse reaction upon reading an article, but then gets the chance to show other people than although he stands by his opinion, he admits that he could have worded it better.” Most respondents thought using this option would lead to disrespect by others, for example: “It feels rather sanctimonious. People do not like admitting they were wrong, and it could cause other users to disrespect them.” Nonetheless, notifying users after posting and providing them with the option to edit or delete their posts was considered acceptable by most respondents.

4.4.5 Study details for publication III

Publications II and III were both based on the same online survey, and thus shared the same basic methodology (see subsection 4.4.3). This subsection shows how the survey data was analyzed for Publication III.

Qualitative analyses of the responses to two open-ended questions (first reactions to designs and explanations for the better design choice) were conducted. MS Excel was used for coding and organizing the data. A data-driven explorative analysis was conducted informed by the socio-cognitive analytical lens of technological frames (users' assumptions, expectations, and knowledge) (Lin & Silva, 2005; Orlikowski & Gash, 1994). It was kept in mind that people generally choose to emphasize some aspects of reality so that certain problem definitions, causal interpretations, moral evaluations, and outcomes are favored and promoted (Entman, 1993; Lin & Silva, 2005; Orlikowski & Gash, 1994). An open and axial coding approach, informed by the lens, was used to highlight themes from the data and build a hierarchy of categories. The approach was grounded theory *-oriented*. The coders paid particular attention to the following aspects of the responses: (a) how the responses described the designs, (b) how the respondents described their reactions to the designs, and (c) what kind of vocabulary was used in the responses (e.g., style, tone, length of the response).

Contributed to: RQ1 and RQ2.

4.4.6 Publication III: key findings

The data analysis found several interesting categories and characteristics of high-quality UI interventions in uncivil online news commenting. Simple and vague characteristics mentioned by the respondents, such as ease of use or familiarity, were omitted. A detailed discussion of the identified characteristics of high-quality can be found in the publication, and most of them are discussed in the thesis Chapter 6. Hence, to avoid repetition, they are not mentioned here.

4.5 Study for publication IV: An international online survey with online news commenters to find how discussion context affects alert design evaluation

The third study was a mixed-methods factorial quasi-experiment. It examined the impact of three factors on the perceived quality and expected effectiveness of alert designs: the level of politeness of the alerted user's intended recipient and the form and amount of guidance in alert designs. Alerts are brief pop-up messages on the user's screen in the context of their current task, informing them of something they must immediately realize. While I present the study's findings in Publication IV, I illustrate the key findings below.

4.5.1 Critical and conventional designs

To test how the amount and form of guidance influences alert evaluation, I created alerts that vary within these dimensions (see Figure 14. below). I refer to the alerts with less guidance later as text-less-guidance and figures-less-guidance. Later, I refer to the alerts with more guidance as text-more-guidance1, text-more-guidance2, figures-more-guidance1, and figures-more-guidance2.

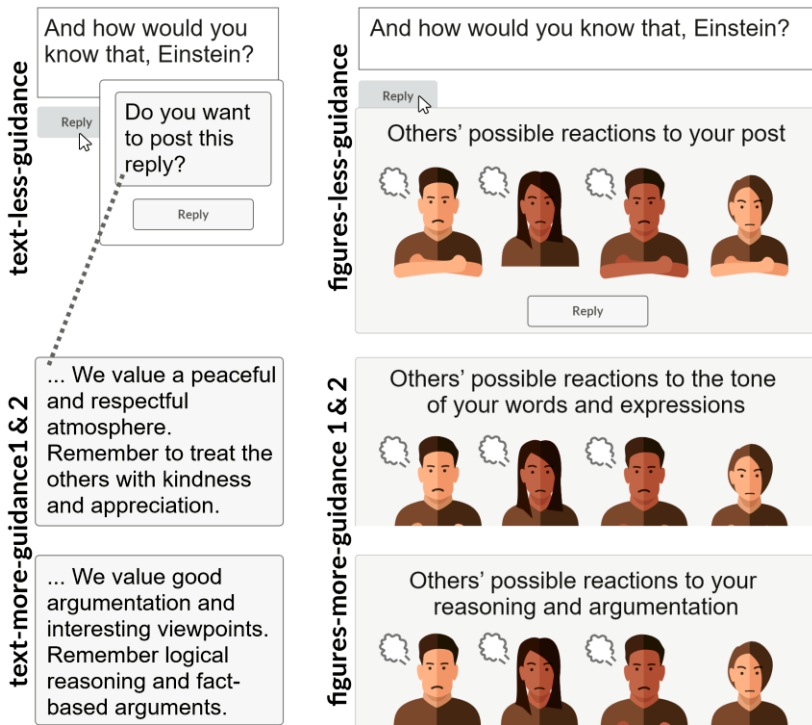


Figure 14. Illustration of the six alert designs explored in the study.

I created a new version of the *Audience* design for this study. In this version, the *Audience* does not continuously watch the commenter. The *Audience* shows possible reaction to the user's comment after they press "Reply." If the user writes somewhat impolitely, some audience members might appear neutral, while others have an angry expression. If the user writes very impolitely, most audience members would appear angry. The design aims to motivate commenters to consider the spectrum of readers who might see their comment. In other words, by making the *Audience* appear only when the comment might be impolite, I have made the design potentially more acceptable, and potentially more likely to have a critical function. It was of interest if this makes a noticeable difference.

The conventional, text-based alert designs (see the "text" designs in Figure 14. above) were inspired by the Toxic Comments plugin by Vox Media's Coral Project (Coralproject, 2017). The designs' conventionality rests primarily on the use of common user-interface elements like buttons and text. I do not consider it worthwhile to analyze this design's criticality.

4.5.2 Study III details

How the designs were presented. The eight designs were presented using storyboards, that is, series of pictures and text illustrating how a user could use them or how they would encounter them in commenting.

Survey respondents: Recruited from Prolific, with criteria: fluency in English, normal or corrected to normal vision, and a minimum of 70% previously submitted studies approved. First, 970 people were asked how often they comment, where, and if they would like to participate in another survey. Second, those who commented at least occasionally on online news sites and confirmed they wanted to participate in another survey were invited to another survey. A pre-study survey that was used to select suitable comments for the primary survey received 169 valid responses (e.g., click-throughs were discarded). The main survey received 248 valid survey responses (e.g., click-throughs were discarded). 25.8% of the 248 respondents were from South Africa, 14.5% from Italy, 13.3% from Poland, and other countries $\leq 10\%$ per country.

Survey procedure: First, each survey respondent was shown a chain of comments and asked to evaluate a new, impolite reply to the chain. The recipient of the new reply was varied impolite/polite. Second, the survey respondent was asked to evaluate how the replier would react to a UI alert. This was asked about a text-based and an *Audience* alert (i.e., one of the three versions of both). Third, the survey respondent was asked to indicate which one of the alert designs they saw was better and to explain why. The survey questions included various closed-ended statements and open-ended questions. The questions on the designs' effects included questions on what the replier would do to their comment upon seeing the alert and how desirable the survey respondent believed it to be to alert the replier/user using the design.

Data analysis: ANCOVAs controlling for various socio-demographic variables were conducted to analyze if offensiveness and appropriateness ratings of the replier's reply differed depending on whether the recipient's comment was impolite or polite. Exploratory factor analyses were conducted to extract dimensions from the design ratings data. A 2 (Recipient: impolite/polite) \times 2 (Alert's form: text-based/human figures) \times 3 (Amount of guidance: more1/more2/less) mixed-design ANCOVA was conducted on factor-based design scores. Significant interaction effects were broken down with one-way ANCOVAs and Bonferroni-corrected post hoc pairwise t-tests. Design preference counts were compared with Chi-Square

Goodness of Fit tests. Lastly, a brief qualitative analysis was conducted on the written explanations for the choice of design.

Contributed to: RQ1 and RQ2. Results presented in Publication IV.

4.5.3 Publication IV: key findings

I found that online news commenters preferred text-based alerts over alerts containing expressive human figures when both contained explicit guidance about what should be posted (see Figure 15. below). However, when both alerts were inexplicit about what should be posted, the alert with the human figures was preferred. This highlights that online news commenters prefer alerts that contain guidance.

This finding suggests that the expected practical usefulness of all the designs could be slightly increased by adding or specifying the guidance in them. For example, if guidance text were to be added under the dog animation in *Creature V2*, the design might be expected to be slightly more practical. Similarly, in the *Regret V2*, if some guidance on how to improve the comment were to be offered rather than only the edit and remove options, the design could seem slightly more practical.

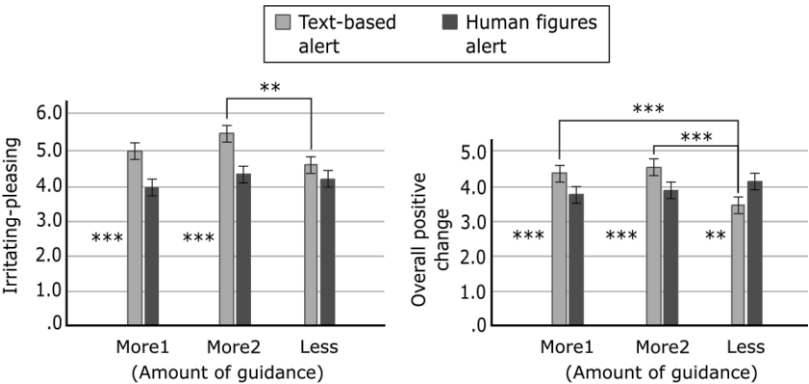


Figure 15. Designs' irritating-pleasing and expected overall positive change in user behavior scores. The asterisks indicate significant differences according to p-values adjusted with Bonferroni correction: **p < .01, ***p < .001.

Besides statistical findings, qualitative explanations for the respondents' alert design evaluations are briefly reported in Publication IV. Based on them, the “flip” seen in the overall positive change score (i.e., the darker bar representing the audience design in the figure is higher than the lighter one representing the text-based design) is

explained by four arguments. The arguments are: the website should tell the user what is wrong with their comment, and the text alert with less guidance does not; it would be useful to see how the other users reacted; it would be beneficial to bring a sense of human presence into commenting; and the users would pay more attention to the images than the text.

4.6 Analysis for answering to the thesis research questions

As the thesis includes new analysis to answer the research questions, in this section I explain how it was done.

4.6.1 Analysis for answering to research question 1.1

To answer RQ1.1 “What design dimensions and aspects may reasonably be used to constrain CD in this context?”, I created a design space schema (Biskjaer et al., 2014), consisting of design dimensions (i.e., measurable extents of particular kinds, for example, size, emotionality) and aspects (i.e., categories of parts or features, for example, functionality: different functions). This was similar to what Bardzell et al. (2014) did. The challenge was as Bardzell et al. state, “to understand the particular ways that a critical design can differ from the more conventional designs that they simultaneously embody and critique.” The following details the analysis process:

1. The starting point for creating the design space schema were the high-level interaction design aspects proposed by Bardzell et al. (2014) (topic, purpose, functionality, interactivity, form, materiality) and questions “why”, “how”, and “what.” I used these to help to think about different aspects and to ensure that I would not focus too much in some area (e.g., pertaining only to “form”).
2. I created an initial map of manipulable and goal related components of the design space based on three things in my memory: (1) the design spaces I used in creating the designs. (2) the findings of the publications. (3) factors that literature mentioned are important to consider in designing interventions or comment moderation.
3. I re-read the findings and the relevant literature and refined the schema. I added sources and explanations for each of the dimensions and aspects that I identified.

4. I asked my thesis supervisors to comment on the design space schema and refined and clarified it based on their comments. Also, I further refined it based on thesis pre-examiners comments.

As my description illustrates, creating the design space schema was a highly iterative process. This aligns with what Biskjaer et al. (2014) noted about design spaces: “it changes not only according to [conditions of the design project], but also when designers learn more about the situation they as designers address, and examine new approaches while discarding old ones.”

After having created the design space schema, I compared several of the designs against it, and created tables that illustrate how the designs filled the design space. I present the design space schema for criticality in Chapter 5 Section 1 and how different designs filled the design space in Chapter 5 Section 2.

4.6.2 Analysis for answering to research question 1.2

To answer RQ1.2 “What kind of designs and design features are likely to provoke reflection, discussion, imagination, appreciation of complexity, and consideration of new perspectives among people who are knowledgeable about online news commenting?”, I estimated which of my designs are the likeliest to have successful critical functions based on the participants' design evaluations and comments. Bardzell et al. (2014) suggested arguments about “successful critical function” can be strengthened with empirical component: observing how people react to critical designs. “Successful critical function” refers to the design's ability to engender discussion, reflection, imagining, appreciation of complexity, and consideration of new perspectives by the intended audience (ibid.). It requires the design to appear sufficiently provocative, plausible, appropriate, or fascinating to its target audience (ibid.). Here the target audiences are people who know what it is like to comment. Designs that appear too provocative to them are less likely to have a critical function than designs that appear somewhat provocative. Similarly, designs that appear acceptable to them are less likely to have a critical function than designs which acceptability is uncertain. The same is true for specific design features.

For the above reasons, the first part of the analysis was about noting the participants' initial reactions to the designs. The second part of the analysis was about noting whether discussion, reflection, etc., occurred. The third and last part of the analysis was about comparing these findings on different designs and design features. I also created a figure; a map of which designs appear more likely to have successful

critical function and practical function. I present the map and the analysis in Section 5.2.

4.6.3 Analysis for answering to research question 2

To answer RQ2 “What characterizes a high-quality user-interface intervention aiming to influence online news commenting behavior?”, I combined the findings of Publication III (which was focused on this question) with the findings from the other publications. I used the categorization of the quality characteristics in Publication III as the starting point and constructed a set of characteristics and requirements that takes the other publications’ findings into account. Then I iteratively clarified and explicated the characteristics, partly based on my thesis supervisors’ comments. This also included creating a summary of how the characteristics relate to the design space schema I created to answer RQ1.1. Lastly, I estimated how the online news commenters and journalists’ standpoints might have impacted their quality-related comments. The findings of the analysis are presented in Chapter 6.

5 RESULTS: DESIGN SPACE FOR CRITICAL DESIGN

The previous chapter presented some of the studies' key findings, and the research questions in the studies were narrow. This chapter takes a more broad and discursive approach to the data. This chapter looks at all the studies, theoretical background, and answers to **RQ1**: “What characterizes the design space for critical design of user-interface interventions aiming to influence online news commenting behavior?” The chapter begins by focusing on the overall design space, answering RQ1.1 “What design dimensions and aspects may reasonably be used to constrain CD in this context?” After this, the chapter focuses on specific design spaces, answering to RQ1.2 “What kind of designs and design features are likely to provoke reflection, discussion, imagination, appreciation of complexity, and consideration of new perspectives among people who are knowledgeable about online news commenting?”

5.1 Dimensions, aspects, and guiding ideas of the design space

Back in 2019, I used a design space consisting of the creativity constraints “UI,” “critical,” and “support emotion regulation,” and not much else. However, my understanding of the design space grew and evolved as my work and studies progressed. I argue in Tables 6-8 below that specific design dimensions and aspects are relevant for CD that aims to engage online news commenters, readers, or journalists. Table 6 includes design aspects and dimensions related to the intention of the design, while tables 7 and 8 feature design aspects related to achieving the intentions. Table 7 includes system-related aspects; in other words, aspects related to an intervention's operation and application rules. These aspects are partially or entirely “behind the scenes” from the user's point-of-view. In contrast, Table 8 includes interaction design aspects, which refer to the visible, concrete side of design. Later, I describe which dimensions and aspects I focused on in the research.

Table 6. Intention-related design dimensions and aspects.

	Meaning	Sources and reasoning
Practical—Critical	To what extent the design should present a practical or critical function? Note that some people might not be interested in discussing a design they believe could not have a practical function.	Inferred based on Sections 1.1, 2.2-2.3, and Publication I.
Reflectiveness	Should the design trigger a reflective response, moving the user to a more self-aware stance? Also, should the design itself embody reflectiveness, for example, by revealing itself as a rhetorical device?	Based on (J. Bardzell et al., 2014). In this context, the critical functions are what is intended to be achieved by unconventional system and interaction design as discussed in the following tables.
Changing perspectives	Should the design present “a framing or point of view that is new, coherent, and interesting enough to help the user perceive the particulars of a domain according to a new schema”?	
Enhancing appreciation	Should the design contribute “to the user’s appreciation of or judgment on design’s role(s) in a sociocultural issue of significance, by making the user more perceptive, imaginative, or aware of the complexity (political, symbolic, etc.) of a domain”?	
Proposals for change	Should the design embody a provocative proposal for an alternative way of being; grounded in possibility, not easily dismissed as ‘science fiction,’ and fairly easy to imagine?	
Target behavior	What conscious or automatic behavior the intervention should encourage, mitigate, or prevent. In CD, this could be something unexpected (but it does not have to be).	All the publications and literature on interventions and critical designs broadly (J. Bardzell & Bardzell, 2013; Fogg, 2009).
Empowerment	Who should end up receiving more opportunities, control, and power. They could be moderators, media organizations, users, bots, or trolls. CD could propose to empower a group of people.	Relevance inferred based on discussion and findings in Publication I and III.
Experience	How is the intervention intended to feel? For example, serious, funny, friendly, or honest. CD could propose somewhat unusual experiences.	A basic design dimension. Based on all thesis publications

Table 7. System-related design aspects related to achieving the intentions in Table 6.

	Meaning	Sources and reasoning
Intervention context	Should the intervention occur under a specific discussion topic or everywhere; on what devices; at what time of day; when the user is of certain age or personality, or in a certain state of mind such as motivated, happy, or agitated? CD could propose a somewhat strange context for the intervention.	Based on the discussion in Publication IV. Additionally, Fogg (2009) argued that interventions work best when the user is motivated.
Preconditions	What exactly causes the intervention to be activated? Alternatively, if the system observes the users constantly, what causes changes? CD could explore unusual or somewhat strange intervention preconditions.	Relevancy is inferred based on findings in Publication III and discussion in Section 3.1.
Timing	When does the intervention occur on the user action level (e.g., before reading, before writing, during writing) and on the interaction level (e.g., upon opening a webpage, scrolling, clicking a button)? CD could explore unexpected timings.	The “user action level” was discussed in Publication II. “Interaction level” or similar is discussed by (Cooper et al., 2014; Silver, 2007).
Integration	How is the UI intervention connected to comment moderation tools or moderation actions? For example, what if the intervention is ignored? CD could highlight ethical and privacy questions here. For example, CD could propose that successful interventions are recorded, joyfully celebrated, and the developers are given bonuses.	Inferred based on findings in Publication III and discussion in (Rantasila et al., 2022).

Table 8. Interaction design aspects and dimensions related to achieving the intentions in Table 6.

	Meaning	Sources and reasoning
Form	How is the intervention made available to human awareness via the user interface? Appearance, structure, colors, etc. CD could propose an unusual form.	A basic design dimension. Based on all thesis publications.
Content	What text, audio, and visual content there is? CD could propose unusual content.	A basic design dimension. Based on all thesis publications.
Design's behavior	How should the design "behave" (from user's point of view)? CD could propose unusual "behavior."	A basic interaction design dimension (Cooper et al., 2014). Based on all thesis publications.
Theoretical claims	What theoretical claims does the design make? There are at least three types of theoretical claims: category claims (e.g., incivility is undesirable), event claims (e.g., this user was uncivil), and explanatory claims (e.g., the user was uncivil because they failed to regulate their emotions). Traditional designs do not tend to make weird or speculative theoretical claims, but critical designs do.	Based on (J. Bardzell & Bardzell, 2013) discussion of how CD uses speculative theory, discussion in Section 2.4, and Kuhn & Pearsall's (2000) discussion of different types of theoretical claims.
Literary devices	What literary devices are used? In traditional design, for example, ethos, pathos, kairos, humor, allegory, pun, hyperbole, and metaphor might be used. However, CD commonly also uses irony, ambiguity, and satire.	Based on Publication I and discussion in thesis Section 2.4.
Elements of simplicity	What is used to make the behavior easier for the user? For example, time, money, guidance, and reminding of the social price of deviation. CD could propose unexpected elements of simplicity.	Based on Publication III and (Fogg, 2009).
Interactivity	The input/output functionality. What input the user needs to give or can give, and why. For example, what buttons must be clicked and what input the user can give voluntarily. Also, is the intervention gamified? Also, are there settings? CD could propose strange input/output functionality.	A basic design dimension. Based on (J. Bardzell et al., 2014). See also findings in Publication III.
Elements of transparency	What features and elements make the purposes and means of the intervention explicit or hidden from the user? For example, written explanations or an option to contact administrators when there are questions. CD could propose something unexpected, like a video of an engineer explaining the system very technically.	Based on the discussion in Publication III, and in (Caraban et al., 2019).

To illustrate the design dimensions, constraints, and aspects in the previous tables, I explain how they relate to one of my critical designs, the *Creature* (see Figure 11. in section 4.4.1):

Intention. The *Creature* is a critical UI intervention which is proposed because people have trouble regulating their emotions online and leave uncivil comments on

the news (c.f., *Target behavior*). The *Creature* is intended to help the user to make quick improvements on their writing as well as trigger the user to reflect on commenting, emotions, and commenting systems (c.f., *Target behavior*, *Practical—Critical*). The *Creature* is intended to have three critical functions (*Reflectiveness*, *Changing perspectives*, and *Proposals for change*). The *Creature* was not designed to empower anyone in particular or to highlight issues related to power for discussion (c.f., *Empowerment*). As a whole, the design is intended to feel innocent and playful (c.f., *Experience*).

System. The *Creature* is proposed to be a one-size-fits-all solution; it is not proposed to be applied in a specific context (c.f., *Intervention context*). The *Precondition* that is proposed to be used to trigger, or rather cause changes in the *Creature* is the tone of user's writing. The user is proposed to see the *Creature* when they begin to write a comment (c.f., *Timing*). How the *Creature* is connected to moderation or other interventions is left unspecified (c.f., *Integration*).

Interaction. The *Creature* is a virtual dog and reacts to the tone of writing like a dog might react to shouting (c.f., *Form*, *Content*, *Design's behavior*, *Interactivity*). The dog concept is used to communicate that there is not a strong theoretical claim about the user's behavior or writing (c.f., *Theoretical claims*). A computer may not understand what the user is writing, and neither do dogs—but they can still make people look foolish in comparison. In other words, the design features gentle satire (c.f., *Literary devices*). The *Creature* may make it easier to watch one's tone (c.f., *Elements of simplicity*). The design, however, does not explicitly explain any of the previously mentioned things to the user (c.f., *Elements of transparency*).

5.1.1 Key design aspects

Having illustrated the dimensions and aspects, I believe some are more important than others. A critical UI intervention design must have a *critical function* (along a practical one or solely), use a *literary device* typical to CD (e.g., irony, ambiguity, satire, or humor), have a *form*, and at least a hint of *precondition/s* and *target behavior/s*. Further, as I discussed in Section 2.2, it is not necessary for the *form* of the critical design to be unusual or strange; the critical function may be tied to unusual *preconditions* or *target behavior*, for example.

However, suppose the goal is to explore the design space further, find a set of good critical UI intervention designs, or improve an existing design. In that case, all the listed dimensions and aspects may be necessary to consider. Additionally, if the intention were to explore aspects beyond the UI, which was not my intention, it

might be good to consider also possible physical and offline aspects. Exploring what physical materials, devices, hardware, and actions could be used or involved in an intervention is a future opportunity for CD. Alternatively, if the intention were to provoke discussion inside a media organization (again, not my intention in the thesis), then branding and marketing-related aspects could be wise to focus on. I speculate those aspects could include, for example, the desired effect on brand and long-term business strategy, visual design vis-à-vis brand, and functionality vis-à-vis brand.

5.1.2 Path relationships between the aspects

Continuing to examine the aspects and dimensions, it is worth noting that they may have path-dependent relationships. That is, choosing one thing can necessitate or prevent choosing another thing. For example, if the purpose of a UI intervention should be clear to the user (c.f., *Elements of transparency*), then it could not be ambiguous (c.f., *Literary devices, Form*).

That said, it is also worth noting CD can break path relationships that traditional design cannot. CD can use ambiguity (as discussed in Section 2.4.), meaning that its proposed combinations and theoretical claims can be nonsensical. For example, CD could propose that one design aspect is another (e.g., timing is form), akin to proposing an alternative universe or dreamworld. Additionally, CD may use combinations of things that are too offensive from the perspective of traditional design. For example, in traditional design, blood cannot be used as material for decoration. However, in CD one could use what one claims to be blood as a decoration, to communicate some point, or to provoke discussion, as is seen in Dunne & Raby's Teddy Bear Blood bag Radio (J. Bardzell et al., 2014; Dunne & Raby, n.d.).

5.1.3 Design dimensions and aspects the thesis focused on

Having explained what design aspects and dimensions I identified for CD, I now consider which ones I focused on in the thesis studies. See Table 9.

Table 9. The CD dimensions, aspects, and guiding ideas in the thesis studies.

		Highlighted or studied?
Intention	Practical—Critical	Yes, I explored what is a practical or critical design.
	Reflectiveness	Yes, particularly the Audience, Creature, Highlight, Philosophy, Promise, and Regret intend to trigger a reflective response and reveal themselves as unusual design proposals.
	Changing perspectives	Yes, particularly the Audience, Creature, Highlight, Philosophy, and Regret are intended to propose new perspectives to uncivil commenting or mitigating it.
	Enhancing appreciation	Yes, particularly the Audience, Highlight, and Philosophy are intended to enhance appreciation about emotions in comments and the complexity of the problem of uncivil commenting.
	Proposals for change, Target behavior	Yes, particularly the Audience, Creature, Highlight, Philosophy, Promise, and Regret intend to propose somewhat unusual behaviors for the users.
	Empowerment	Little. I did not specifically highlight (potential) power imbalance or inequality, except a little with the Promise design (see Publication I).
	Experience	Somewhat. For example, I used the Creature V2 design to explore having a fun and warm experience.
System	Intervention context	Yes, however, I highlighted contextual factors only in Publication IV.
	Preconditions	No. Or a little bit in the Highlight design.
	Timing	Yes, I proposed several different timings for reflection.
	Integration	No.
Interaction	Form	Yes. Several of my designs have provocative or unusual forms, content, and behavior to achieve critical functions.
	Content	
	Design's behavior	
	Theoretical claims	Yes. The designs made speculative explanatory claims. For example, the Audience design made the claim that the commenters need to see their audience.
	Literary devices	Yes. I used ambiguity in most of the designs. With the exception of Evaluate, Highlight, and Symbols I used satire. They contain only humor, irony (being contrary to expectations), and ambiguity.
	Elements of simplicity	Yes, for example, I proposed novel means to help the user reflect on how they are commenting.
	Interactivity	Yes, somewhat. I explored novel user input mechanisms in a couple of my designs.
	Elements of transparency	No.

Considering *Integration*, I did not focus on how the UI interventions I designed could be connected to comment moderation actions. This represents a future opportunity for CD of the UI interventions. For example, future work could present an octopus-like artificial intelligence controlling multiple UI interventions like tentacles.

Alternatively, CD could raise the question of what the moderators should do if the users ignore automatic UI interventions. Should those users be banned, for example?

Considering the *Preconditions*, only in the *Highlight* did I focus on the problem of deciding what exact behavior should trigger an intervention. Exploring different triggers represents a future opportunity for CD.

Considering the *Elements of transparency*, I do not believe having specifically highlighted or played with elements that could make for increased transparency of UI intervention. Future work could highlight transparency issues. However, some of the designs did lead study participants to reflect on what counts and does not count as manipulating user behavior (Publication I).

5.2 Reactions to the designs

Previously, I focused on the design space *aspects* and *dimensions* for CD of UI interventions in uncivil online discussion. However, the findings did not focus on what works critically or practically in the real-world, or what could be improved so that it does. In other words, what are some real-world constraints? Consequently, this section addresses RQ1.2 “What kind of designs and design features are likely to provoke discussion among people who are knowledgeable about online news commenting?”

In Figure 16. , I estimate which of my designs are the likeliest to have successful critical functions and which are the likeliest to have practical functions. The designs’ positions along the two axes are based on the journalists’ and online news commenters’ reactions to the designs, as shown in the publications.

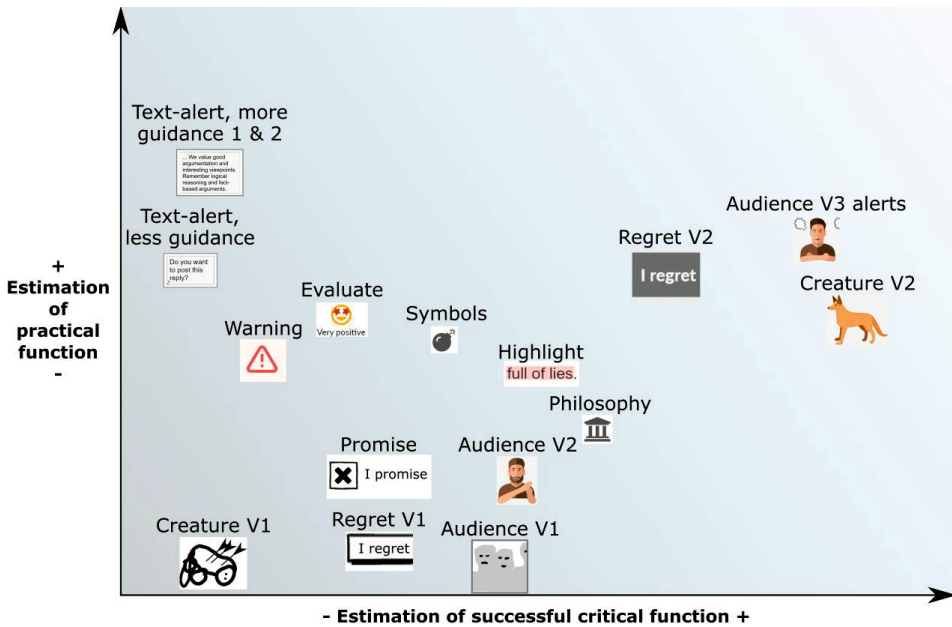


Figure 16. My estimation of the designs' successful critical function vs. practical function. This is based on my subjective analysis of the statistical and qualitative findings in all publications.

As previously explained, “successful critical function” refers to the design's ability to engender discussion or reflection by the intended audience (J. Bardzell et al., 2014). It is a construct of the design appearing sufficiently provocative, plausible, appropriate, or fascinating to its target audience, as evidenced by the audience's reactions and discussion (ibid.). Here the target audiences are people who know what it is like to comment. In order to determine where the designs fall on this dimension, I analyzed the participants' design evaluations and comments. However, as there is randomness in any audience's reaction and the analysis of reactions is subjective, the positions of the designs are approximate. J. Bardzell et al. (2014) note that “empirical results are unlikely to conclusively resolve debates [about a design's successful critical function].”

“Practical function” in the figure refers to the design's ability to support or improve online news commenting. The positions of the designs along this dimension are also based on my analysis of the participants' design evaluations and comments. However, the vertical positions may be slightly more trustworthy than the horizontal ones because some can be traced back to the statistical findings. For example, *Audience V2* received significantly worse ratings than the *Evaluate* (Publication II),

and the *text alerts with more guidance* received better ratings than *Audience V3* alerts (Publication IV).

Additionally, it is noteworthy that the two dimensions can correlate more or less strongly. For example, I estimate that some participants wanted to comment on the *Regret V2* in depth because they felt the design was a mixed bag. This is based on the fact that many participants commented that they were fine with the proposed edit and remove options but not with the proposed regret option (i.e., adding a label showing others that one regrets one's words). They feared that using the regret option would lead to the user being ridiculed by others. The edit and remove options, which were perceived practical by many, probably support discussing the regret option, which was not perceived practical by many.

5.2.1 Regret

I now describe people's reactions to some of my critical designs and how the designs might be improved. However, rather than describing reactions to every one of the designs, I focus on three designs that may be the most relevant for discussing CD, beginning with the *Regret*.



Figure 17. The Regret V1 and V2 designs in short.

As I already discussed above, the findings on the *Regret* designs (see Figure 17. above) show that asking the user to attach the label “Username regretted their choice of words” on their comment is a provocative idea. In Publication I, I found the interviewed Finnish journalists thought the users would use the regret option in *Regret V1* for rhetorical purposes or fun, ‘regretting’ all over the place. They also thought that providing only the regret option would feel limiting to the user. On the other hand, the survey respondents appeared more concerned about the regretting user ending up bullied by the other users. Nevertheless, the findings agree that the regret option is a provocative idea. The design would likely have a much less critical function and appear more practical if the regret option was removed.

In Table 10. below, I further describe what makes the design critical. I also describe what opportunities for future CD there are concerning the design.

Table 10. The CD dimensions and aspects and the Regret V1 & V2.

	Dimensions and aspects	Highlighted or studied?
Intention	Practical—Critical	Yes, the design is intended to have both practical and critical functions.
	Reflectiveness	Yes, the design is intended to trigger reflection and reveal it has purposes other than problem solving.
	Changing perspectives	Yes, the design proposes speculatively that the problem is the users do not regret.
	Enhancing appreciation	Yes, the design is intended to enhance appreciation about design's role/s in the commenting culture and behavior.
	Proposals for change, Target behavior	Yes, the design proposes a novel, unusual, but not unimaginable behavior of publicly regretting comments.
	Empowerment	No, the design is not intended to empower anyone in particular. This is a future opportunity.
	Experience	Yes, the experience is intended to be like making a confession, which is unusual.
System	Intervention context	No. Should the intervention occur under a specific discussion topic or everywhere, etc., is not highlighted. This is a future opportunity.
	Preconditions	No. The specific preconditions of the intervention are not really focused on. This is a future opportunity.
	Timing	No, the timing is not strange or unusual.
	Integration	No. How the design relates to, for example, moderator actions, is not focused on. This is a future opportunity.
Interaction	Form	Not really, as the form is not strange or unusual.
	Content	Yes, asking specifically to regret is unusual.
	Design's behavior	Yes, the system showing to other users that the user regretted is unusual.
	Theoretical claims	Yes, the design claims that regretting publicly in commenting is desirable, which is an unusual claim.
	Literary devices	Yes, it could be seen to feature irony (be contrary to what one expects) and humor.
	Elements of simplicity	Yes, this is a provocative proposal to make regretting easier.
	Interactivity	Yes. It proposes new input and output: asking to leave the regret label.
	Elements of transparency	No. This is a future opportunity.

5.2.2 Audience

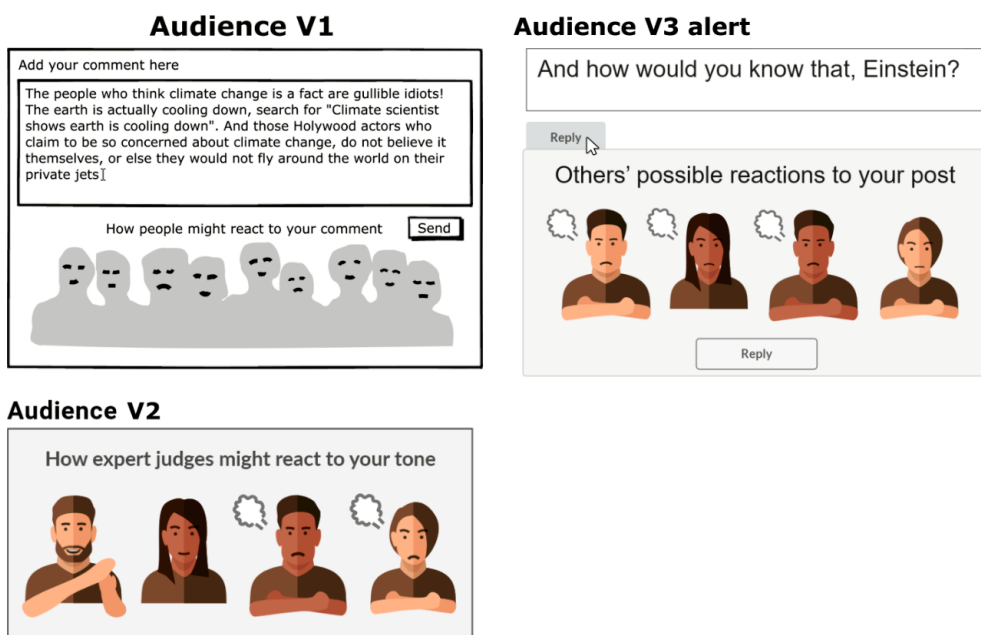


Figure 18. The Audience V1, V2 and V3 designs in short.

Overall, the findings on the *Audience* designs (see Figure 18. above) show that expressive human figures that watch the comment writer and where the expressions reflect the quality of the writing as it progresses are controversial. As reported in Publication I, the interviewed Finnish journalists feared the *Audience V1* design would make the user too anxious. They also feared that some users would use the design as a guide to say as nasty things as possible. As reported in Publication II on *Audience V2*, many of the respondents (online news commenters) took issue with the concept of judging the commenter before they post a comment, while some did not find the design troubling at all.

However, comparing Publication II and IV findings suggests that if the audience does not watch the writing as it progresses but is shown to the user upon pressing “Reply,” the design is likely less troubling. The use of cartoonish human figures, however, seemingly remains controversial. The respondents (online news commenters) who preferred a text-based design over the audience design tended to do so only because they hated the human figures in *Audience V3* (Publication IV).

Furthermore, comparing the findings on *Audience V2* to findings on *Creature V2* (Publication II) suggests that it may not be provocative to propose that an algorithm

monitors the writing as it progresses and shows an incivility score to the user. While the *Creature V2* monitors the writing, it was not rated significantly worse than the designs that do not, unlike *Audience V2*. Besides, Bossens et al. (2022) found that online news commenters considered a monitoring algorithm acceptable.

In Table 11. I further describe what makes the design critical. I also describe what opportunities for future CD there are in relation to the design.

Table 11. The CD dimensions and aspects and the Audience V1, V2 & V3.

	Dimensions and aspects	Highlighted or studied?
Intention	Practical—Critical	Yes, the design is intended to have both practical and critical functions.
	Reflectiveness	Yes, the design is intended to trigger reflection and reveal that it has purposes other than problem solving.
	Changing perspectives	Yes, the design proposes speculatively that the problem is the users do not see their audience.
	Enhancing appreciation	Yes, the design is intended to enhance appreciation about the complexity of commenting well, when one does not know who reads the comments.
	Proposals for change, Target behavior	Yes, the design proposes a novel, unusual, but not unimaginable way of commenting with the help of a virtual audience.
	Empowerment	No. This is a future opportunity.
	Experience	Yes. The design proposes the user should experience presentation anxiety or enjoyment (depending on the person).
System	Intervention context	Somewhat. The Audience V3 is proposed to intervene in a more specific discussion context (replying) than Audience V1 and V2, but the context still is not very specific. Exploring specific contexts is a future opportunity.
	Preconditions	No. The specific preconditions of the intervention are not really focused on. This is a future opportunity.
	Timing	No, the timing is not strange or unusual.
	Integration	No. How the design relates to, for example, moderator actions, is not focused on. This is a future opportunity.
Interaction	Form, Content	Yes, the use of human figures to show a measurement of the user's comment is unusual. Nevertheless, future work could explore AI-generated realistic looking human audiences.
	Design's behavior	Yes, it is a unconventional idea to have an audience react to writing. In Audience V3 this is less strange, however, because it does not react during writing but after the user believes their comment is ready.
	Theoretical claims	Yes, the design claims the user should see their audience or a virtual representation of it. This is unusual.
	Literary devices	Yes, the design features satire and humor.
	Elements of simplicity	Yes, I propose using social pressure to make it seem easier not to write anything offensive.
	Interactivity	Yes. The users write like before but see the audience react. The commenting process may be gamified in a sense.
	Elements of transparency	No. This is a future opportunity.

5.2.3 Philosophy

Philosophy

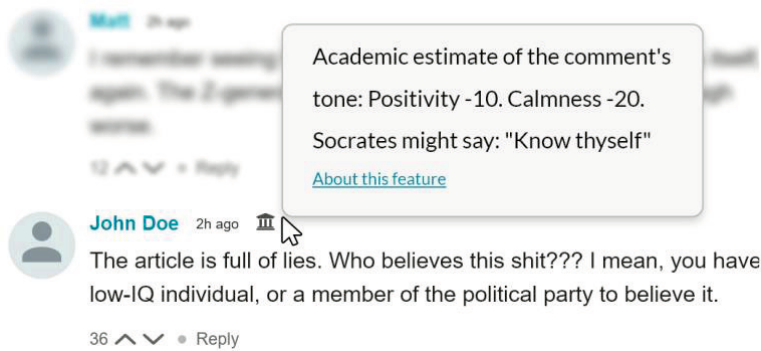


Figure 19. The Philosophy design in short.

The findings on the *Philosophy* design (see Figure 19. above) show that marking the published comments that may be problematic with a small icon is a controversial idea. As reported in Publication II, some respondents (online news commenters) expected the icon to draw other users to attack the user who got the icon and that some users would try to get the icon. At the same time, the respondents commented that the icon might help avoid reading some of the nasty comments. Further, as reported in Publication III, some respondents expected the quote from Socrates to cause the users to write negative comments about the system itself.

In Table 12. I further describe what makes the design critical. It also describes what opportunities for future CD there are concerning the design.

Table 12. The CD dimensions and aspects and the Philosophy.

	Dimensions and aspects	Highlighted or studied?
Intention	Practical—Critical	Yes, the design is intended to have both practical and critical functions.
	Reflectiveness	Yes, the design is intended to trigger reflection and reveal that it has purposes other than problem solving.
	Changing perspectives	Yes, the design frames the problem of incivility as an emotion expression and regulation problem.
	Enhancing appreciation	Yes, the design is intended to enhance appreciation about the complexity of interpreting comments and emotions in them.
	Proposals for change, Target behavior	Yes, the design proposes a slightly strange but imaginable future where the readers analyze emotions in the comments.
	Empowerment	No. This is a future opportunity.
	Experience	Somewhat. The design proposes to bring some fun to commenting.
System	Intervention context	No. Should the intervention occur under a specific discussion topic or everywhere, etc., is not highlighted. This is a future opportunity.
	Preconditions	No. The specific preconditions of the intervention are not really focused on. This is a future opportunity.
	Timing	Yes, the timing is unusual. Published comments are not commonly marked to be potentially problematic.
	Integration	No. How the design relates to, for example, moderator actions, is not focused on. This is a future opportunity.
Interaction	Form	A little. The university icon is unusual in the commenting context.
	Content	Yes, the quote from Socrates is unusual.
	Design's behavior	Yes, the system marking potentially problematic comments is unusual.
	Theoretical claims	Yes, the design claims that it is useful to consider the emotions in the potentially problematic comments. The claim is unusual.
	Literary devices	Yes, it features gentle satire.
	Elements of simplicity	Yes, it proposes to use the threat of being marked to make it seem easier to not write anything offensive.
	Interactivity	Yes. It is a novel proposal to score the comments. This could also be interpreted as gamifying the commenting.
	Elements of transparency	No. While the design features a link "About this feature", I would not call this highlighting transparency issues.

6 RESULTS: CHARACTERISTICS OF HIGH-QUALITY USER-INTERFACE INTERVENTIONS

This chapter examines all the studies and answers to **RQ2**: “What characterizes a high-quality user-interface intervention aiming to influence online news commenting behavior?” As this is based on users’ and journalists’ opinions, the chapter differs from the previous, which discussed design from the designer’s perspective. Additionally, the chapter considers who perceives high-quality and why, adding some detail to the answer to RQ2.

6.1 Discovered characteristics and requirements of high-quality UI interventions in uncivil commenting

In the studies, I identified several characteristics and requirements of high-quality UI interventions from participating journalists’ and online news commenters’ thoughts on what was good, bad, or could be improved in the critical designs. In Figure 20. below, I summarize the found characteristics and requirements. These interrelate, and I do not intend them to replace general UI and interaction design considerations. In the following subsections, I explain the contents of the figure in detail.

A high-quality UI intervention to uncivil commenting helps the user to avoid regrettable situations.

This also entails accounting for the following characteristics and requirements:

<i>System-related:</i>	<i>System and interaction-related:</i>	<i>Interaction-related:</i>
- Has appropriate intervention preconditions and threshold.	- Provides evidence of objectivity.	- Uses informative metaphors, language, and concepts.
- Considers what might happen if the user does change their behavior.	- Stops trolling or at least does not help the trolls.	- Is unprovocative and does not distract users.
	- Does not tarnish or enable others to tarnish the user's reputation.	

Figure 20. A summary of the discovered characteristics and requirements of high-quality UI interventions in uncivil online news commenting.

6.1.1 Helps the user to avoid regrettable situations

According to online news commenters, a high-quality UI intervention helps them to avoid doing something they might soon regret (Publication III). This seemingly relates to Nielsen’s usability heuristic, “Error prevention” (Nielsen, 2020). The commenters did not believe that UI interventions should be limited to targeting only a small minority of troublemakers (i.e., trolls). This finding is also present in the interviewed journalists’ comments (Publication I) but less explicitly than in the commenters’ survey responses (Publication III). As an illustration of the kind of regrettable situations I found the users would like to avoid, I provide the following three examples: “Uh oh, I should not have done that,” “Ugh, why did I read that?”, “Arghh, why did I engage them?”

The example of “Uh oh, I shouldn’t have done that” reflects that some of the survey respondents expected the *Audience V2*, *Regret V2*, and *Creature V2* would help them avoid accidentally or ‘half-accidentally’ posting something offensive. The example of “Ugh, why did I read that?” reflects that some respondents expected the *Philosophy*, *Warning* and *Highlight* to help them avoid reading some of the comments. Finally, the “Arghh, why did I engage them?” indicates that some respondents expected the *Symbols* to help them avoid engaging with users who would only drag them down. This is because, in the *Symbols*, I proposed that instead of writing a negative reply, the user could anonymously give negative private feedback to a commenter.

Additionally, helping to avoid regrettable situations implies that the intervention should not be something the user would regret encountering, nor should the intervention lead other users to create regrettable situations. In other words, this characteristic of high-quality has many implications for design. The following subsections expand on them.

6.1.2 Has appropriate intervention preconditions and threshold

This characteristic of high-quality relates to the preconditions of intervention and hence to system-related design aspects (see section 5.1). The interviewed Finnish journalists (Publication I) and surveyed international online news commenters (Publication III) expressed that UI interventions should not restrict the expression of opinion. In other words, opinions alone should not trigger an intervention. I also found interventions should not evaluate the popularity of opinions. Making users anxious to voice unpopular opinions is perceived as restricting opinion expression

(Publication I). I learned this from the journalists' comments about *Audience V1*, which proposed showing the user how the other users might react to the comment they were writing.

Further, I found that the users want help when needed, and helping them when they do not need help would feel patronizing to at least some of them (Publication III). I also found that users could feel patronized if they interpreted the UI interventions to attempt to moderate the comments to sound positive or to discourage criticism. Some of the proposed UI interventions could make the users think, 'Do they really think we cannot handle negativity?!' These findings imply that the users expect the threshold of intervention to be relatively high.

6.1.3 Considers what might happen if the user does change their behavior

This requirement is perhaps best conveyed with an example: recommending the user mark that they regret their comment could result in the other users bullying them for showing that they regret it (Publication III). In general, encouraging the user to take a positive action that puts them in the spotlight could be detrimental to the user. Others may question the user as to why they turned the spotlight on themselves, and they may be drawn in to attack the user. The requirement suggests that an intervention system like the *Regret V2* should have predictive capabilities and context awareness that inform whether the intervention should be triggered. Therefore, the requirement relates to the system-related design aspects (see section 5.1).

6.1.4 Provides evidence of objectivity

This characteristic concerns how transparent the system should be and what evidence of objectivity of the intervention is provided; hence it relates to the system and interaction-design aspects (see section 5.1). I report in Publication III that high-quality is indicated by having actors that the users can trust as those who judge and moderate the comments. Some survey respondents appeared to indicate that they were doubtful that the proposed evaluators (other users or an algorithm) would evaluate the comments objectively. The responses implied a need for evidence that the intervention is objective in assessing the tone of the comments. From their perspective, why and how the intervention was done and who made the call to intervene should not be opaque—even if the intervention did not technically prevent posting a comment. The interviewed Finnish journalists also appeared skeptical

about objectively evaluating the tone of the comments (Publication I). The journalists also worried that some of the proposed UI interventions, like the *Creature V1*, might be perceived as manipulation attempts.

6.1.5 Stops trolling, or at least does not help the trolls

This requirement concerns the intervention's triggers and how accurately and long the user is shown what triggered it. Hence it relates to the system and interaction design-related design aspects (see section 5.1). I report in Publications I and III that helping the user to avoid doing something they would later regret could also accidentally help the trolls to troll. I found that showing the user what is uncivil could encourage and help the trolls to aim toward incivility (Publications I and III). This was feared to happen if the user is shown how uncivil their comment is estimated to be as they write the comment, when some of the published comments are marked as potentially problematic, and when the level of civility of all published comments as a whole is shown.

That said, some users think that it is not enough to help the users while avoiding helping the trolls. Some users think UI interventions should also deal with the trolls, that is, to stop the trolls (Publication III). This, however, may be difficult to do without accidentally harming those who do not troll (see next subsection).

6.1.6 Does not tarnish or enable others to tarnish the user's reputation

This requirement is about not punishing the users harshly for misbehaving and ruining their reputation or enabling users to tarnish some users for perceived misbehavior (Publication III). The requirement relates to the system and interaction design-related design aspects (see section 5.1). I illustrate the requirement with the following two examples. First, if the users can punish those who troll, they can also attempt to harm those who only have a different opinion. Second, designs where a user who trolls is forever marked as a troll are problematic because if they change their ways, other users could still treat them like a troll.

6.1.7 Uses informative metaphors, language, and concepts

This characteristic of high-quality relates mainly to the interaction design-related aspects (see section 5.1). In Publication III, I report that metaphors and concepts used in UI interventions should harmonize with commenters' values and media preferences. For example, some commenters expected seeing human figures react to their writing (Audience V2) would help them write better comments. In contrast, others did not appear to think so. Further, when alerting a user who may write uncivilly, the alert should provide some guidance on improving the comment (Publication IV). The survey respondents in Publication IV preferred the *Audience V3 less guidance* alert over the *text-less guidance* alert because the text alert contained less guidance on improving the comment. Based on this finding, a good UI intervention also contains clear, concise, and informative language and guidance.

6.1.8 Is unprovocative and does not distract users

Like the previous, this characteristic of high-quality relates to the interaction design-related aspects (see section 5.1). I found some complications associated with using powerful metaphors and concepts in UI interventions to 'drive the point home.' First, I found the participants were concerned that the metaphors and concepts could be anxiety-inducing and distracting (Publications I and III). For example, in Publication I, I report that the Finnish journalists expected *Audience V1* to discourage some users from posting well-reasoned comments that they know other users would not appreciate. In Publications I and III, I also report that the participants believed that the dog and audience metaphors could distract the user from writing a comment and cause them to forget what they were about to write.

Second, I found that the participants feared the designs would become a point of discussion or focus for the users, distracting them from commenting on the news. For example, the journalists feared that some users would try to make the *Audience V1* show expressions that they wanted to other users (Publication I).

That said, I discovered that the users could also be distracted by the core functionality of the studied UI interventions. In Publications I and III, I report that some users could be misdirected to aim for a positive analysis score for their comment if one was shown to them. In Publication III, I report that designs that show which published comments might be problematic or which contain negative expressions could direct the users to focus on negativity. In such a situation, the users would be diverted from the positive comments.

6.1.9 Summary: practicality matters to users

The participants' comments suggest that users perceive UI interventions with practical functionality as high-quality. A UI intervention that does not help the user to avoid regrettable situations (or, even worse, creates those situations) is not high-quality. Other functions are not considered equally important, such as pleasing the senses, engaging the imagination, or triggering reflection on assumptions and unchallenged modes of thought (i.e., critical function).

6.2 Background factors that may affect the perception of quality

In the following subsections, I report findings on factors that may influence how desirable UI interventions in uncivil online news commenting appear to people. First, I report if journalists and online news commenters see things differently. Second, I report if commenters' backgrounds and beliefs influence the perception of quality.

6.2.1 Journalists' vs. commenters' differing stances

Comparing what the ten interviewed Finnish journalists (Publication I) said about the critical designs to what the surveyed 439 international online news commenters (Publications II and III) said about other critical designs shows that the groups characterized high-quality similarly. For example, both journalists and commenters valued commenters' freedom of expression. However, unfortunately, no one from either group expressed that 'it would be great if the news media experimented with solutions, and the world would forgive them for not getting it right from the get-go.' Instead, both groups seemed to expect high-quality UI interventions to be developed behind the scenes and deployed. Both groups seemed to have unrealistic expectations of designers of UI interventions and to underestimate how difficult the problem is to solve without infringing on some stakeholders' privacy or freedom of expression in the process. Additionally, neither group explicitly commented that 'it would be nice if the commenters could discuss with the comment moderators or people in charge of the intervention.' Instead, both appeared stuck in the present world, where such interaction rarely happens (Rantasila et al., 2022). However, I did find one difference between the journalists' and commenters' answers. All the journalists appeared to believe that more or better comment moderation work is

necessary (Publication I), while around a fifth of the participating commenters did not believe so. A fifth of the commenters responded strongly disagree, disagree, or somewhat disagree onto the statement “The news site should moderate the discussion more than currently” (Publication II).

6.2.2 Confirmation bias and the third-person effect

The statistical findings show that some online news commenters were more likely to believe the proposed critical designs would work (Publication II). They were the commenters who hoped for more moderation, did not tolerate incivility, or viewed the commenting situation on their favorite news site as dire. In other words, when people pray for an answer, they are less critical of one. This is a form of confirmation bias.

Next, the findings I reported in Publication III indicate that approximately three in five online news commenters feel that they would not be influenced or helped by an intervention design, while other users would be. This suggests that online commenters may underestimate how helpful UI interventions could be to them and, thus, overall. Moreover, the fact that they believed they would not be influenced but others would be may be an instance of the “third-person effect,” where people believe others are more suggestible than they are (Davison, 1983). It is worth noting that this effect can cause people to ignore messages advocating for them to take positive actions because they wrongly believe that others will be influenced to take positive actions (Wei et al., 2008). This effect might cause designs that attempt to persuade users to take some action to help improve the comments to fail even if they agree it is a good idea. Future CD work could raise the point, ‘If you do not down-vote the trolls, help the moderators, or write civil comments, nobody does.’

7 DISCUSSION AND CONCLUSIONS

This chapter discusses the meaning of the results, reflects on the methodology, revisits the research problems, and points out the next steps.

7.1 The uses of critical UI interventions

Chapter 5 examines criticality as a design aspect (designer perspective) and Chapter 6 focuses on the user's perception of high-quality. The most pressing question based on the chapters is, should people seek to deploy UI interventions perceived as both critical and high-quality (practical)? The question is seemingly comparable to asking, whether users should reflect deeply on their assumptions rather than only make a reflective choice about their behavior. If the answer is yes to the first question, then the answer to the second question also appears to be yes. Criticality does not equal provocativeness, complexity, criticizing the users, or general negativity (J. Bardzell & Bardzell, 2013; Dunne & Raby, 2001). As Jeffrey Bardzell pointed out in a conference presentation, criticality in critical design refers to critique (J. Bardzell, 2018). A critique points out the assumptions, the familiar, unquestioned, unconsidered modes of thought that underlie the practices we accept (Foucault, 1981). Criticality can trigger a reflective response in objects that otherwise serve practical functions (Ghajargar & Bardzell, 2021).

However, it is not evident that the users should be triggered to use their brainpower to reflect on their assumptions. Based on Fogg's Behavior Model (Fogg, 2009), people are less likely to be influenced by UI interventions that require many "brain cycles" than by ones that do not require many "brain cycles." However, it is also probably true that people sometimes use plenty of brain power to self-critique their assumptions about commenting in response to reading comments or when faced with moderation. Thus, trying to trigger reflection on assumptions about commenting culture and technology is not asking for the moon. Based on this reasoning, the answer to the above question changes from "yes" to "sometimes." Sometimes a deployed UI intervention probably should be both critical (to an apparent extent) and practical.

Further, if the question is not the deployment of the designs but their use in design research, then critical UI intervention designs seem quite desirable based on the findings. As stated in Publication I, the critical designs provided the participants with substance to reflect on and provoked the very act of reflection through nonobvious features. This resulted in insights that probably 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.

7.2 Significance and contributions

I now take a step back and reflect on the significance of this thesis for society. First, I reflect on the work against the research problems and the ethical awakening in HCI. I reflect on the work against information systems increasingly displaying agency and intelligence. Finally, I summarize the contributions of the thesis.

This thesis addresses a major, perhaps timeless problem: people focus on designed products and begin to engage with design as production-oriented activity, while it is actually about shaping society (Pierce, 2021; Tromp et al., 2011). For example, through design of technology, designers shape the way we live, work, and interact with one another, even the way we see one another (Horton et al., 2022)—design is much more than design of products. However, designers are often not expected to deliver outcomes other than those which are clearly in the service of production (Pierce, 2021). Designers are often discouraged from foregrounding the ethics of design, problem-finding, and exploring alternatives (Dunne & Raby, 2013; Pierce, 2021). Unfortunately, excessive productional expectations align with unethical design and design that supports prevailing ideologies and trends even when it should not (Dunne, 1999; Jakobsone, 2017; Pierce, 2021). The thesis is significant for society because it contributes to literature that seeks to find a balance between progression and friction in the design of digital technology. Balancing these in design practice can enable the development of more sustainable and appropriate digital products and services.

Furthermore, concerns about inadequate designing may be exacerbated by information systems increasingly displaying agency and intelligence. In other words, as technology moves fast, can design keep up? For background, in 2023, artificially intelligent systems seemed to take a giant leap forward (Lund & Wang, 2023). At the time of writing, it will probably soon be possible (or might already be) for a machine to intelligently intervene in uncivil online discussions by applying tailored and

context-aware UI interventions and sending personalized messages. While this might be a good thing, I fear there is also potential for privacy violations, manipulation, and censorship. Hence, I argue that designers and researchers should use critical and speculative designs to provoke both themselves and citizens to reflect on, discuss, and imagine what these technologies ought to do and what the desired effects would be. Supporting and calling designers and researchers to do this is another reason this thesis is significant for society. More research is urgently needed.

This thesis is also significant for contributing toward “*mitigating uncivil commenting on online news sites*” (the second research problem). To recap, uncivil commenting on news websites causes harm to all parties involved, including moderators and media companies (G. M. Chen & Ng, 2017; Prochazka et al., 2018; Rantasila et al., 2022). Simultaneously, traditional moderation work carried out by journalists is slow and costly (Rantasila et al., 2022). Therefore, the need to guide commenters with automated solutions has been recognized, and such solutions have been explored (see Section 3.2). Nevertheless, the thesis represents a significant leap forward regarding how many automated solutions have been explored and how much they have been studied by researchers. The thesis provides perspectives and knowledge that can help find sustainable automated solutions to mitigating uncivil online news commenting.

7.2.1 Guidance and inspiration for the critical design of user-interface interventions in socio-technical systems

The following summarizes the scientific contributions of the thesis.

The thesis contributes the following knowledge to address the first research problem, “Embracing critical design of socio-technical systems is difficult”:

An interpretation of CD. I explain what CD is in my view and how designers can do it in this context. The discussion may be helpful for people with design backgrounds; in particular, the figures in Chapter 2 may be helpful to interaction, UI, and user experience designers and design researchers broadly. The thesis is a unique, condensed package of many written accounts of CD.

Knowledge about the uses of critical UI interventions in STS. The thesis shows that critical designs may provoke STS “insiders” to consider out loud what social behaviors might emerge after a design is deployed, which may be difficult to speculate on by “outsiders.” Also, the thesis suggests that critical designs can be used to challenge assumptions relating to technological and social factors important for STS

performance, such as ease of use and freedom to participate. Designers may use this knowledge to help justify future CD projects in the area of STS. Researchers who might be most interested to read about using critical designs in these ways include researchers who have studied nudges or intervention designs in social media misbehavior or bullying (e.g., (Difranzo et al., 2018; Wang et al., 2011)), interventions in political polarization in social media or online communities (e.g., (Jhaver et al., 2017; Nelimarkka et al., 2019)), and toxicity in online multiplayer gaming (e.g., (Reid et al., 2022)).

Knowledge of the characteristics of the design space for CD of behavioral UI interventions in uncivil online discussion. The thesis contributes findings on what design dimensions and aspects may reasonably be used to constrain and enable CD in this context. The findings open and reveal the unexplored design space; they should not be interpreted to suggest that only specific aspects and dimensions exist. The findings can both guide and inspire future CD work as well as designing for criticality in the UI interventions. The findings might prove useful for design researchers who do critical or speculative design in this area in the future.

Knowledge of what UI intervention designs are and are not perceived as usual by online news commenters. Achieving “slight strangeness” in CD can involve much trial and error if one does not already know what is perceived as usual (S. Bardzell et al., 2012). Therefore, the thesis findings on expectations of online news commenters may make it easier to design UI interventions that are intended to have a critical function. This knowledge might be interesting to the same as the previous type of knowledge.

7.2.2 Knowledge to help find automated solutions for mitigating uncivil online news commenting

The following discusses the thesis contributions from the perspective of progress toward practical solutions to mitigating uncivil online news commenting.

The thesis may help developers, moderators, journalists, and media researchers see that more UI interventions might be possible than previously realized. As the thesis demonstrates, there are numerous possible UI interventions, even when they are aimed at supporting emotion regulation. Other strategies might more than double the number of possibilities, for example, supporting good argumentation or user-moderator cooperation. Hence, it is clear that not everything has yet been tried to mitigate incivility on commenting platforms. More studies and testing are needed.

Simultaneously, the thesis may help developers and moderators think critically about UI interventions, improve them, and choose new ones to test. This thesis provides the first comprehensive overview of what characterizes high-quality UI interventions from commenters' and journalists' perspectives. The overview highlights things to consider to mitigate incivility while minimizing undesirable side effects like driving many users away. Additionally, the thesis provides the first map of the design space (though primarily intended for CD) and knowledge about individual UI intervention designs.

7.3 Limitations and future work

The CD and the survey methodologies I employed have strengths and weaknesses that affect the validity of the results. These limitations are also often opportunities for future work. While there is no consensus on how to judge the quality of Research through Design (RtD) projects, traditional research validity indicators such as replicability and internal validity may have applications in RtD (Prochner & Godin, 2022). Evaluation of survey methodologies, however, is not unclear. In most of the following subsections, I discuss the thesis studies in relation to traditional research validity indicators and those Prochner & Godin point are applicable in RtD. I also provide suggestions for future work.

7.3.1 Exploring criticality with participants

One of the limitations of the thesis is that criticality was not explained to participants and explored with them. Hearing their thoughts directly on criticality could have enabled making stronger arguments that some of the designs are more likely to have critical function than others (see Section 5.2). It could also have led to greater understanding about criticality in this design context overall.

The main reason why Study I did not explain criticality to the Finnish journalists is that at the time I thought that it is enough to show critical designs to participants and study their reactions. After all, literature on CD suggests critical designers often do not tell people that the critical designs are critical (Pierce, Sengers, Hirsch, Jenkins, Gaver, & Disalvo, 2015), and there are studies where criticality was seemingly not explained to participants (e.g., (S. Bardzell et al., 2012)).

However, at the time I planned Study II I wanted to explore criticality with online news commenters. There are two reasons why I did not. The first reason was COVID-19, which forced me to use online surveys, where one cannot answer participants' questions and explain complex theories. The second reason was I did not believe the survey respondents would understand questions about the designs' critical function. I was also unaware, and still am, of any survey studies that asked specifically about designs' critical functions.

Future work could explore the designs' critical functions with online news commenters, journalists, and designers. Designers might also understand survey questions about designs' critical functions.

7.3.2 Strengthening the findings on the designs and design space by replicating the studies

Moving on to consider research quality indicators vs. my research, the first one I discuss is replicability, and the second is transparency. Replicability refers to steps to reproduce the experiment and get the same results. Transparency refers to an explanation of how and why the research was done. Translated to RtD, these mean explaining the project to ensure recoverability and transparency (Prochner & Godin, 2022).

Attention should be paid to the questions of replicability and transparency when reading the results on the design space (Chapter 5) and characteristics of high-quality (Chapter 6). When describing relevant design dimensions, aspects, and characteristics, it should be clear why they are relevant, how one arrived at that conclusion, and why something might be missing (i.e., potential blind spots). Therefore, I described why they are relevant and what reasoning and study I used to arrive at each conclusion. Regarding the design dimensions and aspects, I state that the list may be incomplete as I did not work with engineers or marketers. I believe it would be ideal if multiple designers and projects were involved in creating a list of relevant design dimensions and aspects. Future work could do precisely that. Future work could consider what relevant design dimensions and aspects in interventions in misbehavior are in other STS contexts (e.g., social media and instant messaging).

Regarding the characteristics of high-quality, they are based on comments that online news commenters and a handful of Finnish journalists knowledgeable about comment moderation had on the designs. Future work could ask opinions of people

who currently work as comment moderators about the designs. It could seek the opinions of people who read comments on news but never comment themselves.

Additionally, the questions on replicability and transparency concern the design generation and selection process and the online surveys. Cockton (2018) proposed that explaining an RtD project well requires explaining five things:

- All significant insights and ideas.
- All primary and secondary sources of information that strongly influenced the exploration and choice of design options.
- All associations between the previous two that were used to integrate and coordinate the design moves.
- All craft knowledge, expertise, and production values that were vital to the design research outcomes.
- How deliberation, understanding, and judgment were used to holistically integrate all of the previous.

I believe I described the issues above reasonably well in Publication I, which was design-focused and most theoretical, but not as well in the other publications that focused on surveying user reactions. While the other publications include descriptions of the design and selection process, the descriptions are brief compared to those in Publication I. Therefore, if Publications II, III, and IV are read without also reading Publication I, it may be unclear how the designs were created and selected.

Considering the online surveys, I included the first online survey in an appendix in Publication II to allow researchers to replicate it. I also included the designs and scenarios shown to survey respondents in an appendix in the publication. However, I am uncertain if I explain the second survey methodology thoroughly enough in Publication IV to allow replication. Similar surveys could be conducted in the future, perhaps even with the same designs but with different people, to understand better how people perceive the designs.

7.3.3 Studying other background factors' influence and developing better measures

Researchers must account for extraneous factors that may influence the research outcome (McDermont, 2011). This relates to internal validity, the extent to which a research design can accurately measure what it claims to measure (*ibid.*). In RtD, considering the extraneous factors seemingly enables researchers to differentiate the

consequences of design variables from socio-demographic and other variables (e.g., commenting frequency, preference for more/less moderation). In thesis studies II and III, I measured the effects of many such variables. However, I did not measure the effects of, for example, political stance and personality on design preference. Measuring them represents an opportunity for future work.

Additionally, researchers must use appropriate measures to ensure internal validity in research (McDermont, 2011). In the thesis research, most of the measures were ad hoc because suitable sets of validated measures did not exist. However, this is not worrisome as traditional, reductionist measurement and evaluation techniques are often not helpful in studying how novel products would be perceived and experienced (Suri, 2002). Due to their reliance upon knowledge of what is relevant to measure, many aspects of perceived quality are likely to be missed.

It may be possible to develop questions or even measures based on the findings reported in Chapter 6. These could then be used in future work when new UI interventions in uncivil online discussion are evaluated with users.

7.3.4 Involving participants who are less interested in scientific research

Representativeness of the sample is a part of the external validity of the research (McDermont, 2011). In contrast to internal validity, external validity refers to the degree to which researchers can generalize a study's findings to other contexts (*ibid.*). External validity is, however, a somewhat unnatural quality indicator for RtD projects as they are often concerned with contextual relevance rather than with extrapolation to overarching theories (Prochner & Godin, 2022; Zimmerman & Forlizzi, 2007). Nevertheless, in this research, I attempted to increase the external validity by carefully selecting participants. I only selected people who would be directly affected by the UI interventions if they were deployed and had experience with commenting or comment moderation.

However, the external validity of the online survey studies may be challenged by noting that the survey respondents, recruited from Prolific, may not represent the majority of online news commenters. Prolific users, or people participating in scientific experiments in general, might be more attentive and open-minded than others. As a result, the UI intervention designs could appear more provocative to online news commenters less interested in participating in scientific experiments. Showing the designs to such people and studying their reactions represents an opportunity for future work.

7.3.5 Using more ecologically valid study setups

Ecological validity is a particular aspect of external validity that enables researchers to transfer findings from experimental to real work situations (Hoc, 2010). I took steps to ensure that the findings would be relevant for the design of critical UI interventions in uncivil online discussion. I created artifacts that looked like UI interventions (i.e., not ambiguous artifacts or ones that allegorically refer to UI interventions) and illustrated how they would work in realistic scenarios. However, what decreases the ecological validity here is that the study settings did not reflect real commenting or group discussion settings. The artifacts were not interactive prototypes but a series of pictures. People might have more or fewer things to say and different things to say about interactive artifacts they use in a real-world context. Studying interactive designs in a real-world context is a clear opportunity for future work. Additionally, the thesis studies did not involve group interviews or similar, but people discussing the artifacts together might have different things to say, and the artifacts might more or less provoke them. This, too, is an opportunity for future study.

Nevertheless, the fact that my designs were not interactive and used by a group of people may have had a positive side. As a result of the lack of interactivity and group setting, the participants were forced to reflect on what they were looking at and develop their own interpretations of it. This might have led to more affluent and more thoughtful reflections.

7.3.6 Iterating on the designs

Iterations can increase ecological validity in RtD (Krogh et al., 2015). I believe the findings on the Audience designs have the highest external validity on the question of which designs have critical and practical functions. This is because of the three iterations of the design (V1, V2, V3), where I evaluate different versions of the design with different people in different settings. In general, iterations allow researchers to understand the reasons for the findings better and identify patterns and trends that may be applicable more broadly (Krogh et al., 2015). Findings concerning designs studied only once are less likely to have external validity (e.g., findings about the Symbols design). In the thesis studies, I sacrifice the external validity of findings on single designs to explore many designs and design options. Future work could iterate on the designs to increase their acceptability and criticality and to better understand the effects of specific design features or combinations of design features.

7.3.7 Future work of a different nature

Having suggested future work that could address the limitations of the thesis research, I now ideate some future work that is more different from mine. First, I propose studying best practices for using critical UI intervention designs in graduate education. I speculate that teachers could use my designs as teaching devices to facilitate discussion among computer science, HCI, or journalism students on sociotechnical questions and the ethics of technology. Teacher and student experiences on the incorporation of the designs in teaching could be recorded and qualitatively analyzed. This proposal came to my mind after reading that ethics courses are now being introduced to CS teaching (Horton et al., 2022). Could critical designs be used to make such courses more engaging and memorable?

My second proposal is to conduct a meticulous study on what constitutes a minimum viable critical UI intervention design (or other UI design) for research purposes. A study on the matter could explore, for example, the following variables: the framing of the design (e.g., “research suggests many perceive it as provocative” vs. “research suggests many perceive it as appropriate”), stated and visualized design features, and the level of polish (e.g., wireframe, mockup, or finished looking). The inspiration for this proposal comes from the fact that while I use rough/quick mockups in Publication I and mockups which look finalized in the other publications, I do not know if such mockups were much better when it came to their critical function. On the one hand, critical designs which look finalized may look more threatening, thus inviting participants to comment on what is wrong with them. On the other hand, rough mockups may invite participants to comment about what they want from the designs, as the designs may appear unset in stone. Nevertheless, since unpolished designs are faster to create, there should probably be significant advantages associated with polished designs to justify their use in research.

8 REFERENCES

- Arendholz, J. (2013). *(In)Appropriate Online Behavior: A pragmatic analysis of message board relations*. John Benjamins Publishing Company.
- Auger, J. (2013). Speculative design: Crafting the speculation. *Digital Creativity*, 24(1), 11–35. <https://doi.org/10.1080/14626268.2013.767276>
- Badham, R., Clegg, C., & Wall, T. (2000). Socio-technical theory. In *Handbook of Ergonomics*. John Wiley.
- Bardzell, J. (2018). *What Rhymes with Critical Design? - Presentation at Critical by Design Conference, Basel*. <https://criticalbydesign.ch>
- Bardzell, J., & Bardzell, S. (2013). What is critical about critical design? *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, 3297. <https://doi.org/10.1145/2470654.2466451>
- Bardzell, J., Bardzell, S., & Blythe, M. (2018). *Critical Theory and Interaction Design*.
- Bardzell, J., Bardzell, S., & Stolterman, E. (2014). Reading critical designs. *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems - CHI '14*. <https://doi.org/10.1145/2556288.2557137>
- Bardzell, S., Bardzell, J., Forlizzi, J., Zimmerman, J., & Antanitis, J. (2012). Critical design and critical theory: The challenge of designing for provocation. *Proceedings of the Designing Interactive Systems Conference, DIS '12*. <https://doi.org/10.1145/2317956.2318001>
- Beheshtian, N., Moradi, S., Ahtinen, A., Väänänen, K., Kähkönen, K., & Laine, M. (2020). GreenLife: A Persuasive Social Robot to Enhance the Sustainable Behavior in shared Living Spaces. *ACM International Conference Proceeding Series*. <https://doi.org/10.1145/3419249.3420143>
- Biskjaer, M. M., Dalsgaard, P., & Halskov, K. (2014). A constraint-based understanding of design spaces. *Proceedings of the Conference on Designing*

- Interactive Systems: Processes, Practices, Methods, and Techniques*, DIS, 453–462.
<https://doi.org/10.1145/2598510.2598533>
- Bleecker, J., Foster, N., Girardin, F., Nova, N., & Viadest, I. (2022). *Design Fiction — Near Future Laboratory*. <https://www.nearfuturelaboratory.com/design-fiction>
- Blythe, M., Andersen, K., Clarke, R., & Wright, P. (2016). Anti-Solutionist Strategies: Seriously Silly Design Fiction. *CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*.
<https://doi.org/10.1145/2858036.2858482>
- Blythe, M., & Encinas, E. (2016). The co-ordinates of design fiction: Extrapolation, irony, ambiguity and magic. *Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work, 13-16-Nov*, 345–354.
<https://doi.org/10.1145/2957276.2957299>
- Bossens, E., Geerts, D., Storms, E., & Boesman, J. (2022). RHETORiC: an Audience Conversation Tool that Restores Civility in News Comment Sections. *CHI '22 Extended Abstracts*.
- Bossens, E., Storms, E., & Geerts, D. (2021). Improving the Debate: Interface Elements that Enhance Civility and Relevance in Online News Comments. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 12935 LNCS*. Springer International Publishing. https://doi.org/10.1007/978-3-030-85610-6_25
- Bostrom, R. P., & Heinen, J. S. (1977). MIS Problems and Failures: A Socio-Technical Perspective. Part I: The Causes. *MIS Quarterly*, 1(3), 17.
<https://doi.org/10.2307/248710>
- Bovens, L. (2009). The Ethics of Nudge. In T. Grune-Yanoff & S. O. Hansson (Eds.), *Preference Change* (pp. 207–219). Springer.
https://doi.org/10.1007/978-90-481-2593-7_10
- Bowen, S. J. (2009). *A critical artefact methodology: using provocative conceptual designs to foster human-centred innovation*. [Doctoral thesis, Sheffield Hallam University].
<http://shura.shu.ac.uk/3216/>
- Brey, P., Gauttier, S., & Milam, P.-E. (2019). *Harmful internet use Part II: Impact on culture and society Study*. European Parliamentary Research Service.

<https://doi.org/10.2861/391152>

- Caraban, A., Karapanos, E., Gonçalves, D., & Campos, P. (2019). 23 Ways to Nudge. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–15. <https://doi.org/10.1145/3290605.3300733>
- Chen, G. M. (2017). *Online incivility and public debate: Nasty talk*. Springer.
- Chen, G. M., & Ng, Y. M. M. (2017). Nasty online comments anger you more than me, but nice ones make me as happy as you. *Computers in Human Behavior*, 71, 181–188. <https://doi.org/10.1016/j.chb.2017.02.010>
- Chen, G. M., & Pain, P. (2017). Normalizing Online Comments. *Journalism Practice*, 11(7), 876–892. <https://doi.org/10.1080/17512786.2016.1205954>
- Chen, J. X., Vitale, F., & McGrenere, J. (2021). What happens after death? using a design workbook to understand user expectations for preparing their data. *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/3411764.3445359>
- Cockton, G. (2018). Way Back to Some Design Futures: Aristotle's Intellectual Excellences and Their Implications for Designing. In J. Bardzell, S. Bardzell, & M. Blythe (Eds.), *Critical Theory and Interaction Design* (pp. 287–309). MIT Press.
- Coe, K., Kenski, K., & Rains, S. A. (2014). Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication*, 64(4), 658–679. <https://doi.org/10.1111/jcom.12104>
- Cooper, A., Reimann, R., Cronin, D., & Noessel, C. (2014). *About Face 4th ed.* Wiley.
- Coralproject. (2017). *Toxic Avengeance - Coral by Vox Media*. <https://coralproject.net/blog/toxic-avenging/>
- Davidson, T., Bhattacharya, D., & Weber, I. (2019). Racial Bias in Hate Speech and Abusive Language Detection Datasets. In: *Proceedings of the Third Workshop on Abusive Language Online, Florence; 2019.*, 25–35. <https://doi.org/10.18653/v1/w19-3504>
- Davison, W. P. (1983). The Third-Person Effect in Communication. *Public*

- Delgado, P. (2019). *How El País used Perspective API to make their comments section less toxic*. <https://www.blog.google/outreach-initiatives/google-news-initiative/how-el-pais-used-ai-make-their-comments-section-less-toxic/>
- Diakopoulos, N., & Naaman, M. (2011). Towards Quality Discourse in Online News Comments Human Factors. *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, 133–142. <https://doi.org/10.1.1.188.3516>
- Difranzo, D., Taylor, S. H., Kazerooni, F., Wherry, O. D., & Bazarova, N. N. (2018). Upstanding by design: Bystander intervention in cyberbullying. *Conference on Human Factors in Computing Systems - Proceedings, 2018-April*. <https://doi.org/10.1145/3173574.3173785>
- Dufva, M. (2020). *MEGATREND 4: Technology is becoming embedded in everything*. Sitra. <https://www.sitra.fi/en/articles/megatrend-4-technology-is-becoming-embedded-in-everything/>
- Dunne, A. (1999). *Hertzian tales: Electronic products, aesthetic experience, and critical design*. RCA Computer Related Design Research.
- Dunne, A., & Raby, F. (n.d.). *Dunne & Raby Projects*. Retrieved January 31, 2020, from <http://dunneandraby.co.uk/content/projects>
- Dunne, A., & Raby, F. (2001). *Design Noir: The Secret Life of Electronic Objects*. Birkhäuser.
- Dunne, A., & Raby, F. (2013). *Speculative Everything: Design, Fiction, and Social Dreaming*. MIT Press.
- Entman, R. M. (1993). Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication*, 43(4), 51–58. <https://doi.org/10.1111/j.1460-2466.1993.tb01304.x>
- Ferri, G., Bardzell, J., Bardzell, S., & Louraine, S. (2014). Analyzing Critical Designs: Categories, Distinctions, and Canons of Exemplars. *Design of Interactive Systems (DIS) 2014*. <https://doi.org/10.1016/j.inoche.2012.03.033>
- Fogg, B. (2009). A behavior model for persuasive design. *Proceedings of the 4th*

International Conference on Persuasive Technology 2009 Apr 26, 350, 1–7.
<https://doi.org/10.1145/1541948.1541999>

Foucault, M. (1981). Practicing criticism (A. Sheridan, Trans.). In L. D. Kritzman (Ed.), *Politics, Philosophy, Culture: Interviews and Other Writings, 1977-1984* (pp. 152–158). <https://doi.org/10.4324/9780203760031>

Fraser, N. (1990). Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy. *Social Text*, 25/26, 56.
<https://doi.org/10.2307/466240>

Gaver, W. W., Krogh, P. G., & Boucher, A. (2022). Emergence as a Feature of Practice-based Design Research. *Designing Interactive Systems Conference (DIS '22)*, 517–526.

Ghajargar, M., & Bardzell, J. (2021, May 6). Synthesis of forms: Integrating practical and reflective qualities in design. *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/3411764.3445232>

Gross, J. J. (1998). The Emerging Field of Emotion Regulation: An Integrative Review. *Review of General Psychology*, 2(3), 271–299.
<https://doi.org/10.1037/1089-2680.2.3.271>

Gross, J. J. (2015). Emotion Regulation: Current Status and Future Prospects. *Psychological Inquiry*, 26(1), 1–26.
<https://doi.org/10.1080/1047840X.2014.940781>

Grut, S. (2017). *With a quiz to comment, readers test their article comprehension*. <https://nrkbeta.no/2017/08/10/with-a-quiz-to-comment-readers-test-their-article-comprehension/>

Habermas, J. (1991). *The Structural Transformation of the Public Sphere: An inquiry into a category of bourgeois society* (Vol. 68). The MIT Press.

Hardaker, C. (2010). Trolling in asynchronous computer-mediated communication: From user discussions to academic definitions. *Journal of Politeness Research*, 6(2), 215–242. <https://doi.org/10.1515/JPLR.2010.011>

Hoc, J. M. (2010). Towards ecological validity of research in cognitive ergonomics. *Theoretical Issues in Ergonomics Science*, 2(3), 278–288.
<https://doi.org/10.1080/14639220110104970>

- Horton, D., McIlraith, S. A., Wang, N., Majedi, M., McClure, E., & Wald, B. (2022). Embedding Ethics in Computer Science Courses: Does it Work? *SIGCSE 2022 - Proceedings of the 53rd ACM Technical Symposium on Computer Science Education*, 1, 481–487. <https://doi.org/10.1145/3478431.3499407>
- Iivari, N., & Kuutti, K. (2017). Critical Design Research and Information Technology. *Proceedings of the 2017 Conference on Designing Interactive Systems*, 983–993. <https://doi.org/10.1145/3064663.3064747>
- ISO. (2019). *ISO 9241-210 : 2019 : Ergonomics of human-system interaction*.
- IxDF. (2023). *What is User Experience (UX) Design?* <https://www.interaction-design.org/literature/topics/ux-design>
- Jakobsone, L. (2017). Critical design as approach to next thinking. *The Design Journal*, 20(1), 253–262. <https://doi.org/10.1080/14606925.2017.1352923>
- Jhaver, S., Vora, P., & Bruckman, A. (2017). Designing for Civil Conversations: Lessons Learned from ChangeMyView. *GVU Technical Report*. www.reddit.com/r/changemyview
- Jigsaw(*google*) *Perspective tool*. (n.d.). Retrieved January 28, 2019, from <https://www.perspectiveapi.com/#/>
- Jigsaw. (2017). *Perspective API*. <https://www.perspectiveapi.com/#/home>
- Johannessen, L. K. (2017). *The Young Designer's Guide to Speculative and Critical Design*.
- Johannessen, L. K., Keitsch, M. M., & Pettersen, I. N. (2019). Speculative and Critical Design-Features, Methods, and Practices. *Proceedings of the Design Society: International Conference on Engineering Design*, 1623–1632. <https://doi.org/10.1017/dsi.2019.168>
- Kang, R., Brown, S., & Kiesler, S. (2013). Why do people seek anonymity on the Internet? Informing policy and design. *Conference on Human Factors in Computing Systems - Proceedings*, 2657–2666. <https://doi.org/10.1145/2470654.2481368>
- Kluck, J. P., & Krämer, N. C. (2020). It's the aggression, stupid! *International Conference on Social Media and Society (SMSociety '20)*.

<https://doi.org/10.1145/3400806.3400826>

- Kohonen, I., Kuula-Luumi, A., & Spoof, S.-K. (2019). Guidelines for ethical review in human sciences. *Publications of the Finnish National Board on Research Integrity* 3/2019, 3.
- Kolko, J. (2009). Thoughts on Interaction Design. In *Thoughts on Interaction Design*. <https://doi.org/10.1016/C2009-0-61347-7>
- Krogh, P. G., Markussen, T., & Bang, A. L. (2015). Ways of drifting—Five methods of experimentation in research through design. *Smart Innovation, Systems and Technologies*, 34, 39–50. https://doi.org/10.1007/978-81-322-2232-3_4/TABLES/1
- Kuhn, D., & Pearsall, S. (2000). Developmental Origins of Scientific Thinking. *Journal of Cognition and Development*, 1(1), 113–129. https://doi.org/10.1207/S15327647JCD0101N_11
- Lapidot-Lefler, N., & Barak, A. (2012). Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. *Computers in Human Behavior*, 28(2), 434–443. <https://doi.org/10.1016/J.CHB.2011.10.014>
- Lawson, B. (2004). *What designers know*. Architectural Press. <https://doi.org/10.4324/9780080481722>
- Lawson, B., & Dorst, K. (2009). *Design expertise*. Taylor and Francis. <https://doi.org/10.4324/9781315072043>
- Lin, A., & Silva, L. (2005). The social and political construction of technological frames. *European Journal of Information Systems*, 14(1), 49–59. <https://doi.org/10.1057/palgrave.ejis.3000521>
- Linell, P. (2001). *Approaching Dialogue: Talk, Interaction and Contexts in Dialogical Perspectives*. John Benjamins Publishing Company.
- Linhares de Carvalho, M., Olsson, T., & Kiskola, J. (2021). Exploration of user interface mechanisms with affect labeling to enhance on-line discussion. *AcademicMindtrek '21: Proceedings of the 24rd International Conference on Academic Mindtrek*.
- Løvlie, A. S., Ihlebæk, K. A., & Larsson, A. O. (2018). User Experiences with

- Editorial Control in Online Newspaper Comment Fields. *Journalism Practice*, 12(3), 362–381. <https://doi.org/10.1080/17512786.2017.1293490>
- Lowry, P. B., Zhang, J., Wang, C., & Siponen, M. (2016). Why Do Adults Engage in Cyberbullying on Social Media? An Integration of Online Disinhibition and Deindividuation Effects with the Social Structure and Social Learning Model. *Information Systems Research*, 27(4), 962–986. <https://doi.org/10.1287/ISRE.2016.0671>
- Lu, D. (2019). Google’s hate speech AI may be racially biased. *New Scientist*, 243(3243), 7. [https://doi.org/10.1016/s0262-4079\(19\)31505-2](https://doi.org/10.1016/s0262-4079(19)31505-2)
- Lund, B. D., & Wang, T. (2023). Chatting about ChatGPT: how may AI and GPT impact academia and libraries? *Library Hi Tech News*, ahead-of-print(ahead-of-print). <https://doi.org/10.1108/LHTN-01-2023-0009>
- Malpass, M. (2017). *Critical Design in Context: History, Theory, and Practices*. Bloomsbury Academic.
- Masullo Chen, G., Muddiman, A., Wilner, T., Pariser, E., & Stroud, N. J. (2019). We Should Not Get Rid of Incivility Online. *Social Media + Society*, 5(3), 205630511986264. <https://doi.org/10.1177/2056305119862641>
- Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. <https://doi.org/10.1080/02699930802204677>, 23(2), 209–237. <https://doi.org/10.1080/02699930802204677>
- Mazé, R. (2009). “Critical of what.” In *Iaspis forum on design and critical practice: The reader* (pp. 379–397).
- McDermont, R. (2011). Internal and External Validity. In J. N. Druckman, D. P. Green, J. H. Kuklinski, & A. Lupia (Eds.), *Handbook of Experimental and Political Science*. Cambridge University Press.
- Morozov, E. (2013). *To save everything, click here: technology, solutionism, and the urge to fix problems that don’t exist*. Allen Lane.
- Mujica, A., Crowell, C. R., Villano, M. A., & Uddin, K. M. (2022). ADDICTION BY DESIGN: Some Dimensions and Challenges of Excessive Social Media Use. *Medical Research Archives*, 10(2). <https://doi.org/10.18103/MRA.V10I2.2677>

- Nelimarkka, M., Rancy, J. P., Grygiel, J., & Semaan, B. (2019). (Re)design to mitigate political polarization: Reflecting Habermas' ideal communication space in the United States of America and Finland. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW). <https://doi.org/10.1145/3359243>
- Nelson, H. G., & Stolterman, E. (2012). *The Design Way: Intentional Change in an Unpredictable World* (2nd ed.). The MIT Press.
- Nielsen, J. (2020). *10 Usability Heuristics for User Interface Design*. <https://www.nngroup.com/articles/ten-usability-heuristics/>
- Nitschinsk, L., Tobin, S. J., & Vanman, E. J. (2022). The Disinhibiting Effects of Anonymity Increase. *Cyberpsychology, Behavior, and Social Networking*, 00(00), 1–7. <https://doi.org/10.1089/cyber.2022.0005>
- Noortman, R., Funk, M., & Eggen, B. (2021). What Would Margaret Atwood Do? Designing for Utopia in HCI. *AcademicMindtrek '21: Proceedings of the 24rd International Conference on Academic Mindtrek*.
- Norman, D. A. (2005). Human-centered design considered harmful. *Interactions*, 12(4), 14–19. <https://doi.org/10.1145/1070960.1070976>
- Norman, D., & Nielsen, J. (2023). *The Definition of User Experience (UX)*. <https://www.nngroup.com/articles/definition-user-experience/>
- Orlikowski, W. J., & Gash, D. C. (1994). Technological Frames: Making Sense of Information Technology in Organizations. *ACM Transactions on Information Systems (TOIS)*, 12(2), 174–207. <https://doi.org/10.1145/196734.196745>
- Paakki, H., Vepsäläinen, H., & Salovaara, A. (2021). Disruptive online communication: How asymmetric trolling-like response strategies steer conversation off the track. *Computer Supported Cooperative Work: CSCW: An International Journal*, 30(3), 425–461. <https://doi.org/10.1007/s10606-021-09397-1>
- Perrigo, B. (2023). *Elon Musk Signs Open Letter Urging AI Labs to Pump the Brakes*. TIME Magazine. <https://time.com/6266679/musk-ai-open-letter/>
- Pierce, J. (2021). In tension with progression: Grasping the frictional tendencies of speculative, critical, and other alternative designs. *Conference on Human*

- Pierce, J., Sengers, P., Hirsch, T., Jenkins, T., Gaver, W., & Disalvo, C. (2015). Expanding and refining design and criticality in HCI. *Conference on Human Factors in Computing Systems - Proceedings, 2015-April*, 2083–2092. <https://doi.org/10.1145/2702123.2702438>
- Prochazka, F., Weber, P., & Schweiger, W. (2018). Effects of Civility and Reasoning in User Comments on Perceived Journalistic Quality. *Journalism Studies*, 19(1), 62–78. <https://doi.org/10.1080/1461670X.2016.1161497>
- Prochner, I., & Godin, D. (2022). Quality in research through design projects: Recommendations for evaluation and enhancement. *Design Studies*, 78, 101061. <https://doi.org/10.1016/J.DESTUD.2021.101061>
- Rains, S. A., Kenski, K., Coe, K., & Harwood, J. (2017). Incivility and Political Identity on the Internet: Intergroup Factors as Predictors of Incivility in Discussions of News Online. *Journal of Computer-Mediated Communication*, 22(4), 163–178. <https://doi.org/10.1111/jcc4.12191>
- Rantasila, A., Kiskola, J., Olsson, T., Syrjämäki, A. H., Ilves, M., Isokoski, P., & Surakka, V. (2022). Outlining Approaches to Improve Online News Commenting with Computationally Aided Comment Moderation. In *Futures of Journalism* (pp. 195–210). palgrave macmillan.
- Redström, J. (2006). Persuasive design: Fringes and foundations. *International Conference on Persuasive Technology 2006*, 112–122. https://doi.org/10.1007/11755494_17
- Reid, E., Mandryk, R. L., Beres, N. A., Klarkowski, M., & Frommel, J. (2022). Feeling Good and In Control: In-game Tools to Support Targets of Toxicity. *Proceedings of the ACM on Human-Computer Interaction*, 6, 27. <https://doi.org/10.1145/3549498>
- Rösner, L., & Krämer, N. C. (2016). Verbal Venting in the Social Web: Effects of Anonymity and Group Norms on Aggressive Language Use in Online Comments. *Social Media and Society*, 2(3). https://doi.org/10.1177/2056305116664220/ASSET/IMAGES/LARGE/10.1177_2056305116664220-FIG2.JPEG

- Rowe, I. (2015). Civility 2.0: a comparative analysis of incivility in online political discussion. *Information Communication and Society*, 18(2), 121–138. <https://doi.org/10.1080/1369118X.2014.940365>
- Rynning, M. (2017). Speculative Graphic Design: Visual Identity Branding As a Catalyst for Change. *Nordes 2017: DESIGN+POWER*, 7(7), 1–6.
- Saresma, T., Pöyhtäri, R., Knuutila, A., Kosonen, H., Juutinen, M., Haara, P., Tulonen, U., Nikunen, K., & Rauta, J. (2022). Verkkoviha : Vihapuheen tuottajien ja levittäjien verkostot, toimintamuodot ja motiivit. *Valtioneuvoston Selvitys- Ja Tutkimustoiminnan Julkaisusarja*, 48. <https://julkaisut.valtioneuvosto.fi/handle/10024/164244>
- Seering, J., Fang, T., Damasco, L., Chen, M. C., Sun, L., & Kaufman, G. (2019). Designing user interface elements to improve the quality and civility of discourse in online commenting behaviors. *Conference on Human Factors in Computing Systems - Proceedings*, 14, 1–14. <https://doi.org/10.1145/3290605.3300836>
- Shmargad, Y., Coe, K., Kenski, K., & Rains, S. A. (2022). Social Norms and the Dynamics of Online Incivility. *Social Science Computer Review*, 40(3), 717–735. <https://doi.org/10.1177/0894439320985527>
- Silver, K. (2007). *What Puts the Design in Interaction Design*. <https://www.uxmatters.com/mt/archives/2007/07/what-puts-the-design-in-interaction-design.php>
- Simon, G. (2020). *OpenWeb tests the impact of “nudges” in online discussions*. OpenWeb. <https://www.openweb.com/blog/openweb-improves-community-health-with-real-time-feedback-powered-by-jigsaws-perspective-api/>
- Springer, N., Engelmann, I., & Pfaffinger, C. (2015). User comments: motives and inhibitors to write and read. *Information, Communication & Society*, 18(7), 798–815. <https://doi.org/10.1080/1369118X.2014.997268>
- Stroud, N. J., Van Duyn, E., & Peacock, C. (2016). *Survey of Commenters and Comment Readers. Center for Media Engagement*. <https://mediaengagement.org/research/survey-of-commenters-and-comment-readers/>
- Stuart, J., & Scott, R. (2021). The Measure of Online Disinhibition (MOD):

Assessing perceptions of reductions in restraint in the online environment.
Computers in Human Behavior, 114, 106534.
<https://doi.org/10.1016/J.CHB.2020.106534>

Suler, J. (2004). The Online Disinhibition Effect. *CYBERPSYCHOLOGY, BEHAVIOR*, 7(3), 321–326. <https://doi.org/10.1089/1094931041291295>

Suri, J. F. (2002). Designing experience: Whether to measure pleasure or just tune in. In W. S. Green & P. W. Jordan (Eds.), *Pleasure with products: Beyond usability* (pp. 161–174). CRC Press.

Syrjämäki, A. H., Ilves, M., Isokoski, P., Kiskola, J., Rantasila, A., Olsson, T., Bente, G., & Surakka, V. (2022). Emotionally Toned Online Discussions Evoke Subjectively Experienced Emotional Responses. *Journal of Media Psychology: Theories, Methods, and Applications*.

Syrjämäki, A. H., Ilves, M., Kiskola, J., Rantasila, A., Isokoski, P., Olsson, T., & Surakka, V. (2023). Facilitating Implicit Emotion Regulation in Online News Commenting—An Experimental Vignette Study. *Interacting with Computers*. <https://doi.org/10.1093/IWC/IWAD010>

Tharp, B. M., & Tharp, S. (2019). *Discursive design : critical, speculative, and alternative things*. MIT Press.

Tharp, B. M., & Tharp, S. M. (2013). Discursive Design Basics: Mode and Audience. *Nordic Design Research Conference*, 406–409.

Torre, J. B., & Lieberman, M. D. (2018). Putting Feelings Into Words: Affect Labeling as Implicit Emotion Regulation. *Emotion Review*, 10(2), 116–124. <https://doi.org/10.1177/1754073917742706>

Tromp, N., Hekkert, P., & Verbeek, P. P. (2011). Design for socially responsible behavior: A classification of influence based on intended user experience. *Design Issues*, 27(3), 3–19. https://doi.org/10.1162/DESI_a_00087

Van Kleek, M., Murray-Rust, D., Guy, A., O'Hara, K., & Shadbolt, N. (2016). Computationally mediated pro-social deception. *Conference on Human Factors in Computing Systems - Proceedings*, 552–563. <https://doi.org/10.1145/2858036.2858060>

Walther, J. B. (1993). Impression development in computer-mediated

- interaction. *Western Journal of Communication*, 57(4), 381–398.
<https://doi.org/10.1080/10570319309374463>
- Wang, Y., Komanduri, S., Leon, P. G., Norcie, G., Acquisti, A., & Cranor, L. F. (2011). “I regretted the minute I pressed share”: A Qualitative Study of Regrets on Facebook. *Proceedings of the Seventh Symposium on Usable Privacy and Security - SOUPS '11*. <https://doi.org/10.1145/2078827>
- Wang, Y., Leon, P. G., Acquisti, A., Cranor, L. F., Forget, A., & Sadeh, N. (2014). A field trial of privacy nudges for facebook. *Conference on Human Factors in Computing Systems - Proceedings*, 2367–2376.
<https://doi.org/10.1145/2556288.2557413>
- Wei, R., Ven-Hwei, L., & Lu, H. (2008). Third-person effects of health news: Exploring the relationships among media exposure, presumed media influence, and behavioral intentions. *American Behavioral Scientist*, 52(2), 261–277.
- Whitworth, B. (2009). The Social Requirements of Technical Systems. In B. Whitworth & A. De Moor (Eds.), *Handbook of Research on Socio-Technical Design and Social Networking Systems* (pp. 2–22). Information Science Reference. <https://doi.org/10.4018/9781605662640.ch001>
- Whitworth, B., & Zaic, M. (2003). The WOSP Model: Balanced Information System Design and Evaluation. *Communications of the Association for Information Systems*, 12. <https://doi.org/10.17705/1cais.01217>
- Wolfgang, J. D. (2018). Taming the ‘trolls’: How journalists negotiate the boundaries of journalism and online comments. *Journalism*, 146488491876236. <https://doi.org/10.1177/1464884918762362>
- Wong, R. Y., Mulligan, D. K., Van Wyk, E., Pierce, J., & Chuang, J. (2017). Eliciting values reflections by engaging privacy futures using design workbooks. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW). <https://doi.org/10.1145/3134746>
- Wu, S., Lin, T. C., & Shih, J. F. (2017). Examining the antecedents of online disinhibition. *Information Technology and People*, 30(1), 189–209.
<https://doi.org/10.1108/ITP-07-2015-0167/FULL/PDF>
- Yoon, J., Li, S., Hao, Y., & Kim, C. (2019). Towards Emotional Well-Being by

Design. *Pervasive Health' 19: 13th EAI International Conference on Pervasive Computing Technologies for Healthcare*, 351–355.
<https://doi.org/10.1145/3329189.3329227>

Ziegele, M., Weber, M., Quiring, O., & Breiner, T. (2018). The dynamics of online news discussions: effects of news articles and reader comments on users' involvement, willingness to participate, and the civility of their contributions. *Information, Communication & Society*, 21(10), 1419–1435.
<https://doi.org/10.1080/1369118X.2017.1324505>

Zimmerman, J., & Forlizzi, J. (2014). Research through design in HCI. In J. S. Olson & W. A. Kellogg (Eds.), *Ways of Knowing in HCI* (pp. 167–189). Springer New York. https://doi.org/10.1007/978-1-4939-0378-8_8

Zimmerman, J., & Forlizzi, J. (2007). Research through Design: Method for Interaction Design Research in HCI. *CHI 2007*, 167–189.

PUBLICATIONS

PUBLICATION

I

Applying critical voice in design of user interfaces for supporting self-reflection and emotion regulation in online news commenting

Kiskola, J., Olsson, T., Väättäjä, H., H. Syrjämäki, A., Rantasila, A., Isokoski, P., Ilves, M., & Surakka, V.

In Proceedings of the 2021 CHI conference on human factors in computing systems (pp. 1-13).
10.1145/3411764.3445783

Publication reprinted with the permission of the copyright holders.

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 *Ilta-lehti*, 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).

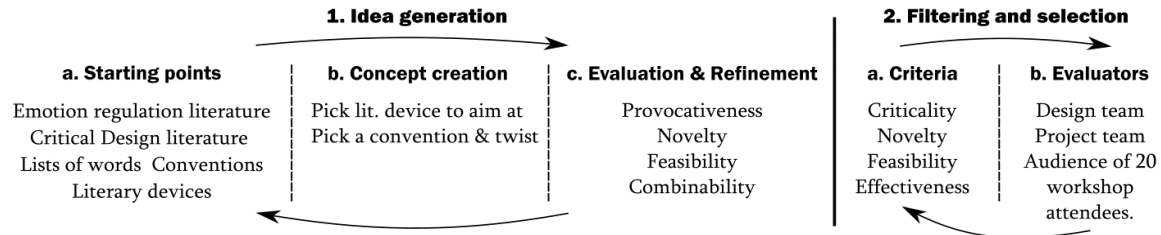


Figure 1. Approximation of the iterative idea generation and filtering and selection process.

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).

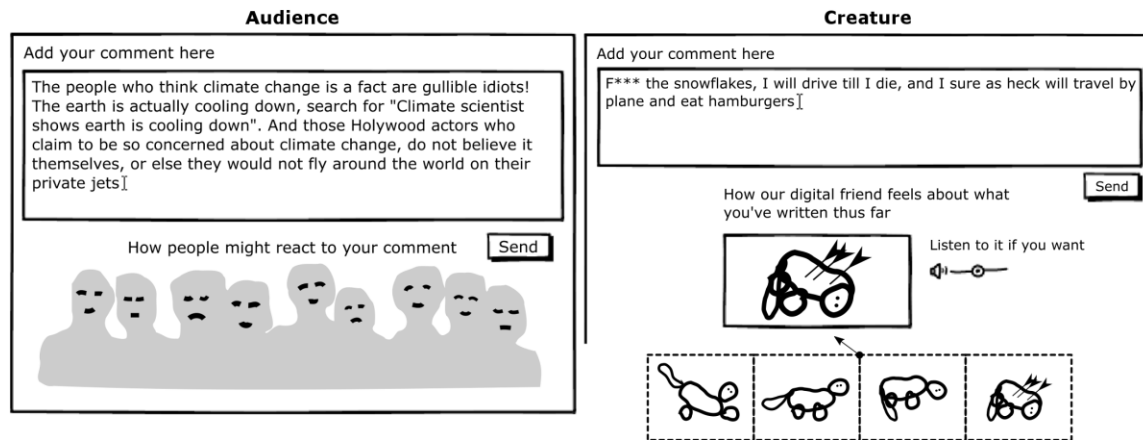


Figure 2. Left: The AUDIENCE as it appears for a comment writer. The audience shows the anticipated emotional reactions to the user's writing. Right: The CREATURE as it appears for a comment writer. The design features a virtual animal that thrives or suffers according to the user's writing.

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).

Regret

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

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

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

PUBLICATION II

Online Survey on Novel Designs for Supporting Self-Reflection and Emotion Regulation in Online News Commenting

Kiskola, J., Olsson, T., Syrjämäki, A. H., Rantasila, A., Ilves, M., Isokoski, P., &
Surakka, V.

In Proceedings of the 25th International Academic Mindtrek Conference 2022 (pp. 278-312).
10.1145/3569219.3569411

**Publication is licensed under a Creative Commons Attribution 4.0
International License CC-BY-NC-ND**



Online Survey on Novel Designs for Supporting Self-Reflection and Emotion Regulation in Online News Commenting

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

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

Aleksi H. Syrjämäki
aleksi.syrjamaki@tuni.fi
Tampere University
Tampere, Finland

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

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

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

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

ABSTRACT

Uncivil commenting on online news is regarded as a persistent and complex sociotechnical issue. Because commenting behavior is inherently conditioned by user interfaces (UIs) on news sites, HCI scholars may approach the issue by proposing alternative UI solutions and thereby potentially mitigating incivility. This paper explores eight novel UI design proposals that aim to support emotion regulation and self-reflection during commenting and reports how the designs are evaluated in an international online survey (N=439) among online news commenters. This exploratory study advances our understanding of what kind of UI solutions, from the end-user's perspective, appear desirable—and why—in terms of improving the quality of online news commenting. For example, desire for moderation was found to predict more favorable ratings of the design proposals in general.

CCS CONCEPTS

• Human-centered computing → Interaction design theory, concepts and paradigms.

KEYWORDS

Design Research, Design Fiction, Social Media, Emotional reflection, Design conventions, Online Survey

ACM Reference Format:

Joel Kiskola, Thomas Olsson, Aleksi H. Syrjämäki, Anna Rantasila, Mirja Ilves, Poika Isokoski, and Veikko Surakka. 2022. Online Survey on Novel Designs for Supporting Self-Reflection and Emotion Regulation in Online News Commenting. In *25th International Academic Mindtrek conference*



This work is licensed under a Creative Commons Attribution International 4.0 License.

Academic Mindtrek 2022, November 16–18, 2022, Tampere, Finland
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9955-5/22/11.
<https://doi.org/10.1145/3569219.3569411>

(*Academic Mindtrek 2022*), November 16–18, 2022, Tampere, Finland. ACM, New York, NY, USA, 35 pages. <https://doi.org/10.1145/3569219.3569411>

1 INTRODUCTION

The communication culture in digital media services has been widely problematized, with scholars referring to issues such as social media rage, use of uncivil language [15, 68], and increased hate speech [30]. This has motivated various approaches attempting to mitigate online incivility, ranging from human-based content moderation [24, 51] to the computational detection of hate speech [14] and toxic language [46]. In this paper, we focus on online news commenting as a specific form of interaction in digital media, where incivility has been found to harm both the readers of news articles, journalists, and moderators [3, 38, 39, 43, 69].

It is well established that online discussion is shaped and conditioned by the computer-mediated nature of communication [9, 53, 65]. From the perspective of emotion psychology, the current largely text-based interfaces may limit the ability to control one's emotions and empathize with other people [59, 64]. Emotion regulation refers to the process and strategies that influence the quality, intensity, and timing of the experienced emotion [28]. The need for emotion regulation arises when emotions are of strong intensity, duration, frequency or wrong type for a particular situation, or they maladaptively bias cognition and behavior [28, 42]. It is likely that especially the attenuation of emotion regulation online is associated with factors identified in the Online Disinhibition effect by Suler [58]. Accordingly, for example, anonymity, invisibility, asynchronicity, and minimization of authority in online communication may result in shift in processes of affect and cognition so that they function differently than in in-person interaction. Consequently, it has been proposed that improvement of communication culture in digital media could be approached also by rethinking how user interfaces can support individual users' emotion regulation [62].

While recent HCI literature features some design speculations of alternative UIs [27, 37], there is little understanding of the user-centric quality of the envisioned UIs. To understand which UI alternatives would be 'better' for a diversity of potential users we need

empirical studies that utilize multiple perspectives of evaluation, explore a variety of design alternatives, and involve an extensive representation of potential users. Also, it is necessary to study what the users anticipate would happen if designs were deployed. In general, what the users anticipate of products can play a central role in shaping their experience [36]. If the users first react negatively and anticipate the designs would not work, this likely affects an actual test of the effectiveness. Furthermore, it is not reasonable to test different alternatives the first time in realistic news commenting environments because of the risk to the news site's reputation and the risk that the design makes the situation worse [37]. To this end, the paper explores eight design proposals to support self-reflection and emotion regulation in the context of online news commenting and reports on an evaluation study of the designs. The designs build on the idea of affect labeling, that is, identifying and explicating the emotional elements in comments or by asking the user to name how they feel [63]. The designs apply different metaphors and design concepts. For example, a virtual audience is shown reacting to a comment as it is written; and potentially problematic published comments are marked with a symbol.

The evaluation study was implemented as an international online survey ($N = 439$) among people who comment on online news sites. We asked each respondent to evaluate two designs. Also, we asked the respondents for background details, for example, to rate their experiences regarding comment moderation. In the results, we first examine quantitative ratings of the designs. Second, we explore the possible reasons behind the ratings by investigating both quantitative associations between the ratings and background variables and the open-ended answers of the survey.

2 RELATED WORK

2.1 Uncivil Online News Commenting is a Difficult Problem to Approach

Online news commenting is a form of public discourse between strangers [15] that takes place around journalistic content on comment sections of online newspapers or broadcasters' websites. The negative aspects of online news commenting, and their consequences have motivated conceptual work and empirical studies with respect to both the reasons for regulating and tools with which to regulate the tone of discussions (e.g., [12, 15, 39, 70]). In addition to harmful effects for the involved commenters, uncivil comments on online news can hurt journalists and moderators, who cannot easily avoid them [22], harm the publisher's brand [49], and have negative effects on readers who do not participate in commenting [12].

Incivility in online news commenting platforms can be approached from many angles by moderators and designers. Ruckenstein and Turunen [51] identify two logics within content moderation on commercial platforms: the logic of choice focuses on finding and deleting uncivil or 'insufficiently neutral' messages, while the logic of care tackles disorder with moderator-writer interaction. The logic of choice is seen in action in the form of users flagging messages, publishers putting up paywalls, limiting the number of characters in posts, and using algorithmic moderation to quarantine or delete messages (see also [26]). However, Ruckenstein and Turunen argue that the logic of choice fails to encourage behavioral change. Within

the logic of care, moderators attempt to improve discussions by educating users or by persuading them to reflect on their commenting. The drawback is that human moderator-driven approaches are costly, difficult to scale, and emotionally stressful for moderators because they must confront emotionally troubling writing. This highlights the need for also technological and scalable approaches to this issue.

2.2 Technological Strategies for Preventing Incivility

Because of the difficulty and expense of human moderation, media companies and researchers have looked for potential technological and user-interface solutions to preventing incivility. The Norwegian Broadcasting Corporation has incorporated custom-built quizzes to confirm that a user has read an article before commenting on it [29]. While this is a relatively low-cost solution, it is time consuming to apply to each news article and discourages some forms of civil commenting, such as quick replies [29]. Another approach is the psychologically "embedded" CAPTCHAs (i.e., challenge-response tests used to determine whether or not the user is human) containing stimuli that prime participants' positive emotions [53]. The authors found that priming increased the positivity of the tone of texts in online commenting. However, as they point out, there are ethical issues involved in influencing users in a "stealthy, covert fashion". Bossens et al. [4, 5] studied the effect of interface designs on online news commenting civility. Their designs directed the users to comment and share their opinion on a particular statement (relevant to the news article). The researchers found that their designs caused the comments to be more civil compared to a control where the users were only asked to leave their comment on the news article. However, as the researchers noted, directing users to comment on a particular statement may not work or be reasonable for all news articles.

Solutions based on computational approaches, such as machine learning, are also being developed, particularly to address the issues of cost and the demand for scalability. For example, Perspective API, developed by Jigsaw [34], can detect "toxic" writing to some extent, and this can be shown to the writer as a score or an emoji or made to trigger a notification that attempts to persuade the writer to reflect on their writing. Reportedly, triggering a simple text-based nudge asking the user to edit their comment can increase the percentage of approved comments by 2.5–4.5% [54]. Thirty-four percent of users chose to edit their comment before sending it upon seeing the nudge, and 54% of them changed it in such a way as to render it "immediately permissible" [54]. In addition, there are solutions for monitoring the tone of writing that people can install as add-ons on their web browsers. For example, Grammarly [25] attempts to detect 19 different tones (e.g., excited, egocentric, and accusatory) with the help of machine learning. The add-on illustrates the detected tone of the writing with an emoji that is placed inside the text-input box.

Algorithmic approaches may also be used to show the readers a sentiment analysis of published posts and threads, which may make some users stop and think before commenting. Such approaches include sentiment analysis on Yahoo News [45] and Gremobot chatbot emotion regulator [48]. Yahoo News has used a row of three

small emojis and percentages to visualize the overall sentiment of comments (see also [45]). The GremoBot chatbot emotion regulator supports emotion regulation in group chats by interpreting the situation positively and visualizing group emotion [48]. The results of their study “suggest that a chatbot emotion regulator can enhance positive feelings and alert people of negative situations”.

Overall, while the previously mentioned algorithmic solutions and tools appear promising, we argue that the solution space remains unexplored. As the problem of uncivil commenting seems to persist regardless of various interventions, we argue for further exploration from the viewpoint of UI design.

2.3 Supporting Self-Reflection and Emotion Regulation

The following elaborates on the theoretical foundations of our design exploration. To complement the logic of choice and the logic of care [51] and to address the aforementioned limitations of the existing solutions, we have suggested a third approach [37]: supporting user self-reflection and emotion regulation with the help of the identification of emotional elements in comments or by asking the user to name how they feel.

Recently, the concept of affect labeling, as an implicit form of emotion regulation, has been discussed in psychology literature [63]. Several controlled laboratory studies have found that emotional experience can be attenuated by simply putting one’s own feelings into words or labeling the emotionally evocative aspect of a stimulus [63]. In addition, Fan et al. [19] analyzed the emotional content of the tweets of 74,487 Twitter users and found that emotional intensity decreased rapidly after their explicit expression in an “I feel” statement.

The present work continues our previous work [37] in taking the idea of affect labeling as an inspiration, rather than a boundary, and exploring various tactics to make users more aware of their own emotions and the emotional elements in the messages. Also, we have limited our exploration in the sense that we do not intend to make definitive judgments on comments’ civility or to argue that making passionate arguments is wrong. This aim arises from the knowledge of how difficult it is to accurately define the limits of (in)civility or “the freedom of expression” [51]. There is long-standing discussion on the (in)civility of public discourse [20, 31, 50], including debate on whether dispassionate deliberation is synonymous with the legitimate expression of public opinion [20].

To further position our work, we recognize that the idea of supporting self-reflection and emotion regulation relates to the theory of nudging [60]. The nudge theory proposes that peoples’ behavior can be influenced with indirect suggestions and positive reinforcement. In general, computational affect labeling could be an approach to nudging (toward emotion regulation) because it gently informs or guides the user while preserving their freedom of choice. However, we are aware that nudging has its risks. For example, nudging may feel patronizing in this context.

3 DESIGNS

The following presents the eight designs on a conceptual level and describes how they are intended to support self-reflection and emotion regulation. For brevity, the multi-stage process of producing

and selecting the designs for this study is only briefly reported in what follows. The full descriptions of the designs, as they were shown to the survey participants, can be found in Appendix 2.

3.1 The Process of Producing and Selecting the Designs for the Study

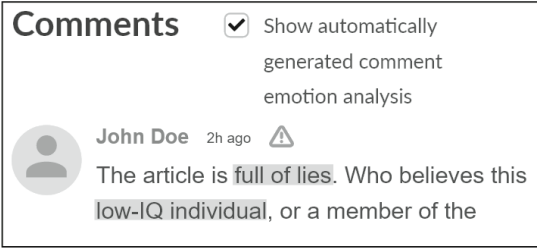
The design work for this study builds upon our earlier research-through-design exploration [37], in which we envisioned unconventional solutions to the problem of uncivil commenting with a critical voice. In the study, we unpacked this same problem area and outlined critical perspectives on potential solutions by describing and analyzing four designs that aimed to support emotion regulation by facilitating self-reflection. Next, to explain how the designs utilized in this study were created, we briefly recap the design process of the earlier study [37].

First, to create novel designs, we identified existing design conventions 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. in 2021. Further, as the research took place in a Finnish university, we examined them in four most popular Finnish news websites (tabloids *Ilta-Sanomat* and *Ilta-lehti*, national newspaper *Helsingin Sanomat*, and Finland’s national broadcaster *Yle*). This resulted in lists of existing UI conventions (e.g., an option to sort comments by recency) and cultural conventions (e.g., people are rarely specific about their intended audience). The lists were used in three ways: to find a convention to be tweaked slightly, to avoid reinventing existing solutions, and to reflect on what kind of solutions might fit various news websites.

Second, approximately 60 concept ideas were sketched based on several idea generation sessions. The idea generation was conducted by a design team consisting of the first author, who has a formal education in interaction design and industrial design, and two colleagues, who both have formal educations in user experience design and software engineering. While the idea generation was not guided by specific design creativity methods, such as fictional inquiry or brainstorming methods, two general strategies mentioned in the critical design literature were used: (1) the designer picks a literary device (e.g., irony, sarcasm, parody, or ambiguity) and attempts to implement it in designs [35] and (2) the designer picks a convention (cultural or UI) and tweaks it slightly, for example, by introducing a foreign concept, and then reflects on the result [2].

Third, 19 of the sketched ideas were subjectively evaluated by the design team as more promising in terms of perceived criticality, novelty, feasibility, and effectiveness. Following this, the first author created UI mock-ups of the 19 ideas. Also, four of the 19 mock-ups were pictured and analyzed in depth in the earlier study [37]. Then, in the present study, we further developed eight of the ideas and made them more presentable. To help ensure that the evaluated designs represent a rich breadth of approaches to support self-reflection and emotion regulation in online discussion, we categorized them by the timing of the intervention and by the design strategy for emotion regulation (more on this in the next section). In addition, we subjectively assessed the designs as conceptually different from one another.

Highlight



Symbols



Creature



Evaluate



Figure 1: Highlight, Creature, Symbols, and Evaluate designs in short.

3.2 The Designs in Brief

We first briefly describe main functionality of and the theory behind each design proposal, followed by an analysis and comparison of the emotion regulation strategies they manifest. Lastly, we briefly describe the basic motives present in the designs.

In the *Highlight* design (see Figure 1 top left), the user is offered an option to view an analysis of the emotions in comments. If the user checks a checkbox, negative emotional expressions are highlighted in red. Comments containing strong negative expressions are also marked with an alert symbol. The design is inspired by the theory of affect labeling [63], and speculates that highlighting negative emotional expressions in comments could calm the users. That said, while the idea of highlighting is straightforward, it is uncommon to show this type of analysis to users. We have not seen this in use on any website.

In the *Creature* design (see Figure 1 top right), an animated dog reacts to the emotional tone of a comment, as the user writes the comment. The design attempts to encourage change through an emotional attachment to a virtual pet dog. The benefits of using emotional attachment to pets to motivate behavior change have been documented in previous research (e.g., [16, 40]). In the design, the pet dog is displayed below the text-area, and it is described as “our digital friend.” If the user writes in a positive way, the dog appears happy, as if ready to play. If the user is writing in a neutral way, the dog appears neutral (see Figure 1 top right). If the user is writing in a negative way, the dog sits on the floor; keeps its head and ears down, with its tail between its legs; and faces away. We argue that the use of an animated dog for this purpose is a novel idea.

In the *Symbols* design (see Figure 1 bottom left), the user is offered a way to provide anonymous, private feedback to any of the previous commenters. This is intended to decrease the likelihood of written personal attacks toward other commenters. It has been demonstrated that uncivil comments (including replies) promote further incivility [11, 74], and that ad hominem attacks are a frequent type of incivility online [13, 41]. In the design there are buttons depicting a bomb, a gavel, a smiling face, and a heart next to every comment. The bomb symbolizes “Full of arrogance”; the gavel “False claim/s”; the smiling face “Well said”; and the heart “Love it!” Also, every user’s profile contains a prominent section entitled “Overview of the feedback from other users”, which displays the same symbols and the number of times the user has received these feedback types. The concept relates to comment up-voting tools seen on popular social media sites and commenting platforms. It is thus arguably less novel than, e.g., *Creature*.

In the *Evaluate* design (see Figure 1 bottom right), the user must first indicate how they feel before they can add their comment. This is done by clicking a smiley face that represents their emotional state. It is proposed that naming the emotion could have a calming effect on an angry user. The design is inspired by and applies the theory of affect labeling [63]. The proposed functionality is relatively like existing feedback tools (e.g., Facebook reactions), making the design appear as the least original of the eight designs. That said, unlike the other designs, Evaluate and Symbols do not propose that the website publicly evaluates comments for their quality. Hence, these designs are also included in the study out of interest for finding whether the difference in the evaluating party is a highly significant factor in acceptability.

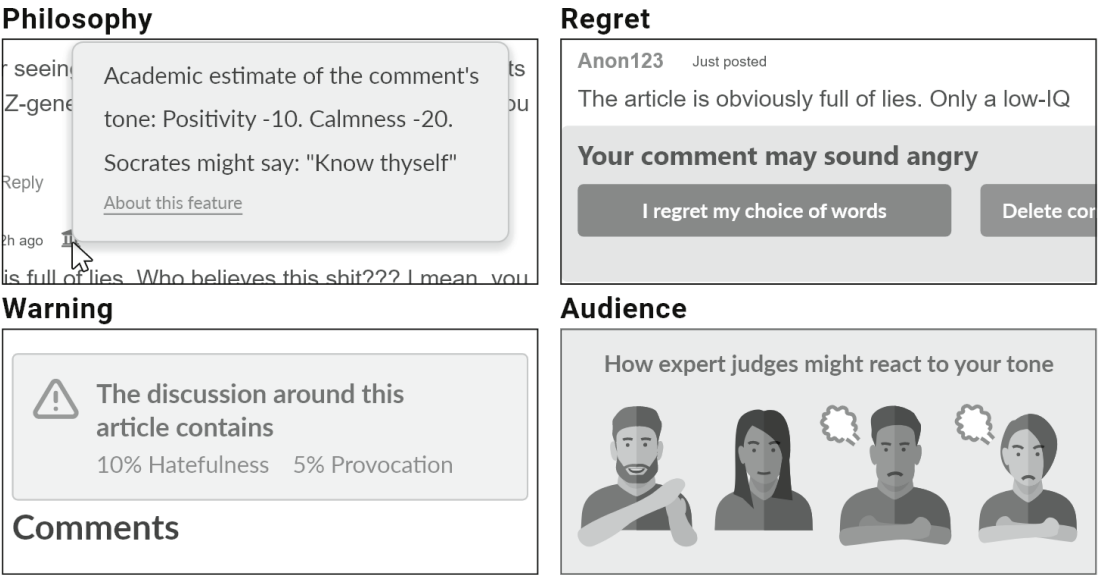


Figure 2: Philosophy, Regret, Warning, and Audience designs in short.

In the *Philosophy* design (see Figure 2 top left), problematic comments and comment threads are marked with a university icon. If the user presses the icon, a box with the emotion score for the comment or comment thread and a quote from Socrates, “Know thyself!” [72] is revealed. The emotion score has two dimensions, positivity, and calmness. The design proposes that automatic evaluation of comments should be done but in a relatively subtle, inconclusive, and ambiguous way. We argue that the use of the icon, the quote, and this type of analysis together are novel and uncommon.

In the *Regret* design (see Figure 2 top right), users’ comments are automatically evaluated directly after posting. If a comment sounds very angry, the user is notified and offered various follow-up actions below the published comment and by email. The first offered follow-up action is to regret the choice of words, the second is to delete the comment, and the third is to edit it. If the user chooses the regret option, a notification is attached to the comment, stating “username regretted their angry words”. While moderators often ask users to edit or delete their angry comments on social media sites (e.g., in Facebook groups), we argue that this emphasis on regret in online news commenting is novel. Previous research has found that postings with profanity or obscenity can be a cause of regret for Facebook users [66].

In the *Warning* design (see Figure 2 bottom left), a notification is shown above the comment section, indicating a description of the argumentation within the comment section (e.g., “10% Hatefulness”). The design proposes that labeling the emotional content of the comment section could help the user to deal with overly negative comments and decrease the likelihood of the user leaving an unconstructive comment. Also, it is proposed the design would

help news readers to decide if they want to read the comments. The concept is somewhat like what has been done in Yahoo News, as discussed above, and not as novel as some of the other designs.

In the *Audience* design (see Figure 2 bottom right), when a user is writing their comment, a virtual audience of expert judges reacts to its tone in real time and their reaction is displayed below the text area. The design intends to evoke the sense of having a live audience, which can make one consider their self-presentation through writing. Related to this intention, previous research has found that showing Facebook users profile pictures of people who will see (cf., judge) their posts can help some of them avoid regrettable disclosures [67]. Also, the *Audience* design utilizes the concept of being watched to induce self-awareness (e.g., [6, 10]). Previous research implies that designs that induce self-awareness might reduce abusive comments to news [55]. The *Audience* would function as follows: If the user writes in a moderately positive way, some members of the audience appear glad, and others have a neutral expression. If the user writes in a rather negative way, most members of the audience appear angry or frustrated. The audience’s appearance in the proposal is also intended to communicate that the audience is ethnically diverse. We argue that the proposal to use virtual audience in the context is, again, a novel one.

Next, we explain how we adopted Yoon et al.’s [73] framework that they created for designers to help them develop solutions that support users to better deal with their emotions. The framework contains 17 “emotion strategies”, which they propose might work in human-product interactions. We used five of the strategies, which we subjectively judged most applicable in this context, to help us select the eight designs to study. The strategy of avoidance relates

Table 1: Selected approaches to self-reflection and emotion regulation [73] adapted to this design context

Timing	Designs	Emotion regulation strategy
Before reading	Philosophy, Warning	Avoidance, Raising self-awareness
While reading	Highlight, Symbols	Problem-focused coping, Raising self-awareness
Before writing	Evaluate	Raising self-awareness
While writing	Audience, Creature	Suppressing expressions, Raising self-awareness
After writing	Regret	Reappraising events, Raising self-awareness

to “things one deliberately does before she/he experiences certain emotions as well as associated behavioral and expressive responses” [73]. *Philosophy* and *Warning* relate to the strategy because they intend to help the reader to avoid the negative emotions comments may cause. The strategy of problem-focused coping refers to finding “practical ways to deal with stressful situations” [73]. *Highlight* and *Symbols* intend to provide a way for the reader to investigate or deal with overly negative comments. It is hoped that these designs will reduce the chance that the reader will respond very negatively. Next, while the strategy of raising self-awareness can be said to be utilized in all the designs because they all have consequences for the comment writer and may induce the feeling of being observed, the strategy is at the forefront in *Evaluate*. *Evaluate* directly asks the user how they feel. Next, *Audience* and *Creature* relate to the strategy of suppressing expression because these designs intend to notify the writer that they are writing in an overly negative tone, enabling them to adjust their tone. Finally, *Regret* relates to the strategy of reappraising events because it intends to change how the writer and then, potentially, the reader perceive the situation.

Lastly, the design proposals may be read as critical or speculative [1, 2]. They are removed from commercial constraints, and they are intended to present new perspectives and encourage user reflectiveness.

4 METHODS

We ran an international online survey to collect a diverse sample of design evaluations by people who comment online news on news sites. The study was implemented with LimeSurvey and invitation to it was circulated at Prolific, a platform for online subject recruitment [47].

4.1 Participants and Recruitment

To select a diverse sample of participants, we first conducted a pre-survey regarding how often the candidate respondents read and commented on online news articles. It involved 2,000 participants who met the specified eligibility criteria: fluency in English, normal or corrected-to-normal vision, and a minimum approval rate of 70% in Prolific (percentage of total submitted studies minus returned).

The criteria for recruiting the pre-survey participants into the design survey were that the participant had provided complete answers and commented at least occasionally on online news sites (excluding social media sites and blogs). Altogether, 480 participants were recruited based on their commenting activity. Of the 480 survey responses, 41 were discarded as incomplete (i.e., missing answers), duplicates (i.e., the same person completing the survey twice), or click-throughs (i.e., two standard deviations faster

Table 2: Participants’ background information

Accepted responses	N = 439	%
Current residence		
UK	190	43.3
Poland	53	12.1
US	44	10
Portugal	39	8.9
Other countries	< 20 per country	23
Unspecified	12	2.7
Secondary education (e.g., GED / GCSE)	22	5
High School diploma / A-levels	69	15.7
Technical / community college	45	10.3
Undergraduate degree (BA / BSc / other)	166	37.8
Graduate degree (MA / MSc / MPhil / other)	127	28.9
Doctorate degree (PhD / other)	8	1.8
Did not know / not applicable	2	.5
Female	199	45.3
Male	240	54.7
Participants’ ages ranged from 18 to 75 years (average 33.5 years, SD = 11.98)		

than the average response time or nonsensical answers to open questions). Separate attention check questions were not used as meaningful answers to the open questions regarding the designs were thought to indicate commitment and attentiveness. For an overview of participants with accepted responses, see Table 2.

Lastly, we note respondents’ opinions of comment moderation were somewhat skewed in favor of greater moderation. In answering the question “The news site should moderate the discussion more than currently”, 3% of the respondents strongly disagreed, 9.6% disagreed, 9.8% somewhat disagreed, 19.6% neither agreed nor disagreed, 20.5% somewhat agreed, 24.6% agreed, and 13% strongly agreed.

4.2 Survey Procedure and Questions

Each participant was shown two pseudo-randomly selected designs. The presentation order of the two designs was randomized. The survey questions included various closed-ended statements and

open-ended questions so as to allow the researcher to holistically study the respondents' impressions and expectations. The questions on design evaluation included statements on various inherent design qualities, desirability, and the expected effects on emotion regulation and behavior. The same set of questions was presented for both designs, though in different, random order. The participants were asked to name their most frequently used news site and consider the presented designs in light of what the commenting is like in that particular context. In terms of background and contextual questions, the participants were asked about socio-economic factors and preferences regarding moderation strength, as well as to assess the commenting culture on the online news site that the respondent primarily uses. The full survey is provided in Appendix 1.

The questions on design evaluation were operationalized by us, except for three items we adopted from Hassenzahl et al. [32] (conventional–inventive, unimaginative–creative, and cautious–bold) and numerical version of the visual Self-Assessment Manikin – scale [7]. Researchers who ask participants to evaluate novel designs in an online survey must often invent new measures and/or pick and utilize parts of existing sets of measures [18, 33]. The same approach was justified in this study by the novel elements of the designs and the lack of suitable pre-existing sets of measures. In addition, we operationalized in Likert-scale items five design dimensions that may capture experienced “dissonance”: clarity, reality (similar to feasibility), familiarity, veracity (e.g., sarcasm or spoof in design), and desirability [61]. The background questions studied in this paper were also operationalized by us. Earlier research on commenting and comment moderation has also typically created new measures [15, 56, 57, 71].

4.3 Data Analysis

For statistical analyses, IBM SPSS Statistics2 Version 26 was used. To increase the validity of design comparisons, the dimensions in the data were first extracted using exploratory factor analysis (EFA). As stated by DiStefano et al. [17], “following an exploratory factor analysis, factor scores may be computed and used in subsequent analyses.” Principal axis analysis with oblique rotation (Promax) was conducted to identify and create sets of variables that explain the maximum amount of variability in the data (Tables 3 and 4). Notably, the EFA was based on 7-point scale items operationalized by us. Designs' emotional impact scores are not reported due to paper length limitations. Further, the EFA was based on the statements about the latter design the respondent saw because we assumed that the questions would be easier to answer when being answered for the second time. Also, all the factor loadings exceed 0.400 and are thus considered sufficient [23]. Then, sum variables (factor-based scores) were created based on found factors for use in subsequent analysis by averaging the individual variables in the factors.

Kruskal-Wallis tests were used to compare the ratings of the various designs (based on the factor-based scores). Significant effects (at $\alpha = .05$) were followed with pairwise comparisons, with Bonferroni correction being used to correct for the family-wise Type-I error rate.

To investigate background variables' effects on design ratings, we conducted univariate linear regression analyses. Separate analyses were conducted for each predictor–outcome variable pair due to multicollinearity between the background variables [44]. The predictors included the background variables extracted using EFA, and the outcome variables were the identified instrumental quality and inappropriateness constructs (see Section 5.1).

To gain insight into the reasoning behind the numerical ratings, we conducted thematic analysis [8] of the respondents' first reactions on the designs. Most of the analysis work was conducted by the first author, who was primarily responsible for creating the designs and thus most capable of understanding what the respondents referred to in their comments on the designs. The other authors offered additional viewpoints to the interpretations. The reactions were captured by an open-ended question, “How would you describe your immediate reaction to this solution? How do you feel about it?” The thematic analysis of the answers focused on explicit comments on design features and mechanisms that could help illuminate the design ratings. Therefore, quantifying the answers and reporting exact counts was not seen as reasonable. The analysis was conducted using MS Word. The respondent quotes are verbatim, except for corrected typos.

5 FINDINGS

5.1 Relevant Sum Variables Identified Using Explorative Factor Analysis

5.1.1 Design Quality Variables. The responses loaded into two key factors (Table 3). We interpret Factor 1 as indicating the perceived **instrumental quality** of the solution (i.e., the degree to which it is perceived to serve as a crucial tool). Factor 2 relates to negative impressions and risks and could be interpreted as referring to the perceived **inappropriateness** of the solution (i.e., the degree to which it is perceived as unsuitable or wrong in the context). The factors appear demarcated by valence (positive vs. negative). The included items, all on 7-point Likert-type scale, were averaged to create sum variables (factor-based scores) and thus represent the two factors in subsequent analyses.

5.1.2 Background Variables. The factor analysis identifies four relevant factors (Table 4). We interpret Factor 1 (Table 4) as reflecting the respondent's view of how desirable the commenting is on a given news site (we name this factor **view on the situation**). Factor 2 relates to behavioral tendencies regarding how likely a person is to engage in discussion, some of which may be heated or controversial (i.e., **interest in debate**). Factor 3 reflects emotional reactions in terms of the degree to which the person is not tolerant of uncivil commenting (i.e., **toleration of incivility**). Factor 4 concerns the user's attitude toward how comments should be moderated (i.e., **wish for moderation**). Because the factors seem meaningful in the given context, the negatively loading items in each factor were reversed, and the items of each factor were averaged to create sum variables (linear combination ignoring weights), or factor-based scores, to represent each factor and be used in subsequent analyses.

Table 3: Exploratory factor analysis of the design quality variables

	Factor	
	1	2
If this solution was implemented, I would take part in news commenting more actively	.882	
This solution would likely engage me in more active discussion on news articles	.870	
The solution would help me manage my emotional reactions	.832	
The solution would help me express my opinions more freely	.804	
The solution would have a calming effect on me	.766	
The solution matches what kind of solutions I wish for	.741	
Overall, I find the solution desirable	.690	
I feel that the designer who made this is trying to deceive or ridicule me		.760
The solution would violate my freedom of speech too much to be acceptable		.727
If I was angry, the solution would make me even angrier		.696
I feel that the solution is sarcastic or a spoof		.691
The risks that the solution introduces are higher than its benefits		.624
Note: Rotated factor solution (Promax with Kaiser Normalization). KMO = 0.908; Bartlett: $\chi^2 = 3522.2$; df = 66; p < .001. Coefficients < 0.3 suppressed. (N = 439; cut-off of eigenvalue ≥ 1 ; total variance explained: 59.74%; variance explained by Factor 1: 49.97%). Cronbach's alpha: Factor 1. 0.93; 2. 0.82.		

Table 4: Exploratory factor analysis of background statements concerning one's behavior, attitudes, and assessment of the commenting culture on a selected news site

	Factor			
	1	2	3	4
Inappropriate comments get quickly removed or are not published at all	.763			
The comments on news articles are respectful	.752			
The news site has moderation practices that ensure the quality of commenting	.744			
The comments on the news site are generally of high quality	.725			
Overall, the news site feels like a place where uncivil commenting simply does not belong	.703			
The news site does not encourage civilized commenting	-.581			
The comments on news articles include inappropriate language	-.495			
Trolling and other intentional misbehavior is common in the commenting section	-.494			
I tend to comment on news articles on topics that are controversial		.686		
I tend to participate in the discussion only when the discussion is heated		.682		
When reading others' inappropriate comments, I tend to write inappropriate responses		.603		
I tend to reply to others' comments		.594		
If I see inappropriate comments on the news site, it will bother me			.890	
If I see hateful speech in the comments, I will not be bothered			-.580	
If I see disrespectful comments on the news site, I will get anxious			.539	
Publishing inappropriate comments is a problem that should be taken more seriously on this news site				.848
The news site should moderate the discussion more than currently				.743
Note: Rotated factor solution (Promax with Kaiser Normalization). KMO = 0.818; Bartlett: $\chi^2 = 2664.756$; df= 136; p < .000. Coefficients < 0.3 suppressed. (N = 439; cut-off of eigenvalue ≥ 1 ; variance explained: 49.67%). Cronbach's alpha: Factor 1. 0.86; 2. 0.73; 3. 0.70; 4. 0.81.				

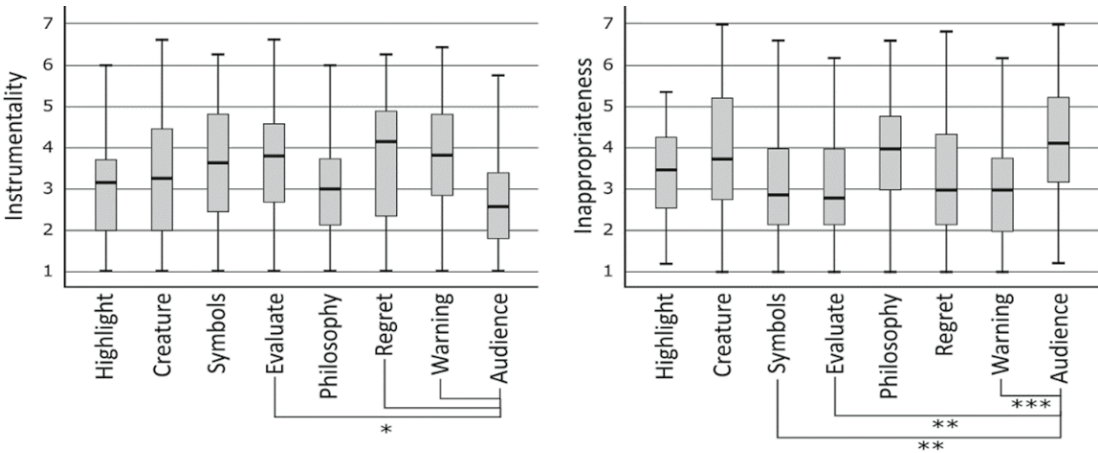


Figure 3: The instrumental quality and inappropriateness ratings of the eight designs (scale 1–7, 1: strongly disagree, 7: strongly agree). The asterisks indicate significant differences according to p-values adjusted with Bonferroni correction: * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 5: Results of regression analyses investigating associations between background variables and instrumental quality ratings

Background variable	Perceived instrumental quality		$F(1, 438)$	p
	R^2	$B(95\% CI)$		
Wish for moderation	.031	0.17 [0.08, 0.26]	13.8	<.001
Not tolerating incivility	.034	0.18 [0.09, 0.27]	15.3	<.001
View on the situation	.014	0.14 [0.03, 0.26]	6.4	.012
Interest in debate	.006	0.11 [-0.02, 0.25]	2.5	.115

Note: Instrumental quality ratings are on a 1–7 scale.

5.2 Design Quality Ratings

Figure 3 summarizes the design ratings for the **instrumental quality** and **inappropriateness** sum variables. Statistically significant differences between the designs were found for both variables (instrumental quality: Kruskal-Wallis $H(7) = 27.67$, $p < .001$; inappropriateness: $H(7) = 36.07$, $p < .001$). Further, post-hoc tests show significant differences in pair-wise comparisons (Fig. 3). Especially *Audience* was considered low in instrumental quality and high in inappropriateness, when compared to other designs. While the ratings do not imply any generally preferred design approach, *Regret* received the highest instrumental quality score, and *Evaluate*, *Symbols* and *Warning* received the lowest inappropriateness scores.

5.3 Associations between Background Variables and Design Ratings

Most of the identified background variable factors were found to significantly predict the instrumental quality rating (see Table 5). Only the variable of interest in engaging in debate was not statistically associated with instrumental quality. However, the background

variables were not found to significantly predict the perceived inappropriateness of the designs (p -values $> .069$; hence excluded from Table 5).

5.4 Respondents’ Reactions to the Design Features

To gain insight into the design ratings, we qualitatively analyzed the respondents’ first reactions on the designs. The analysis focused on explicit comments on the design’ features and mechanisms.

5.4.1 Philosophy and Warning. We proposed above that *Philosophy* and *Warning* would enable users to avoid reading uncivil comments. Considering *Philosophy*, some respondents noted that marking problematic comments with an icon could not only highlight comments for users to avoid but also comments to attack. A few respondents also expected some users to try to get the icon. Considering *Warning*, while many respondents liked the proposal as it would help them avoid reading negative comments, many also doubted the warning would be useful to users. For example: “People who tend to peruse the comments already know those figures, and those who won’t indulge in that, wouldn’t care about them.”

5.4.2 Highlight and Symbols. We proposed that *Highlight* and *Symbols* offer ways for the user to cope with negative comments. Considering *Highlight*, the respondents who saw it as useful thought the highlighting of negative words would help to avoid some comments altogether. No respondent commented that drawing more attention to the negative words could be helpful. Considering *Symbols*, many respondents seemed to believe the design would help to do something about an annoying comment while avoiding direct conflict. For example: “This is pretty intelligent way of expressing your opinion rather than getting personal and start attacking.” However, *Symbols* had another feature which was widely disliked: many respondents noted that enabling other users to leave a lasting, negative mark anonymously on another user’s profile for all users to see would be a bad idea.

5.4.3 Evaluate. While all the designs could raise the user’s self-awareness, *Evaluate* relies on it. However, the respondents were puzzled by the design. Only a few respondents commented that it would be helpful to the commenter to identify their emotion, for example: “it would help people reflect about how they are feeling which could moderate behaviors.” Many respondents speculated that other commenters or moderators could benefit from knowing how the commenter felt. Further, a few respondents commented that it would be annoying to indicate the emotion every time one comments, and a few commented that the emojis are not suitable for a news site.

5.4.4 Audience and Creature. We proposed above that *Audience* and *Creature* would provide the comment writer with the opportunity to adjust their tone. Considering *Audience*, several respondents were explicit that giving the commenter feedback on their writing using the virtual audience of experts would make the commenter feel overly anxious or annoyed. For example: “I don’t want to instantly know that I’m being judged before the comment is even posted” and “I’d be concerned that it would encourage me to write comments that make the virtual experts happy rather than helping me concentrate on what I’m thinking about the news issue.” Further, some respondents noted that “[the feedback] may only serve to encourage some people to carry on their comment further [into negativity].” That said, some commented they would find the feedback useful when composing. Considering *Creature*, while many commented the use of animated dog is childish, many also commented that it is clever as many people feel empathy with dogs. Further, while *Creature* would provide instant feedback like *Audience*, much fewer respondents commented it would make the writer feel anxious.

5.4.5 Regret. We proposed above that by enabling the comment writer to show regret, the design would change how the writer and then, potentially, the reader perceive the situation. Some respondents saw value in the option to add a label that one regretted their choice of words, for example: “I feel like it would be a good way to redeem the person who sends his angry thoughts as an impulse reaction upon reading an article, but then gets the chance to show other people than although he stands by his opinion, he admits that he could have worded it better.” However, most respondents thought using the option would lead to the user being disrespected by others, for example: “It feels rather sanctimonious. People don’t

like admitting they were wrong and it could cause other users to disrespect them.” At the same time, notifying the user after posting and providing the edit and delete options were perceived as fine by many respondents.

6 DISCUSSION

6.1 Reflection on the Findings

We first reflect on how the proposed designs differ from one another in terms of perceived user-centric quality. The findings showed that *Evaluate*, *Regret*, and *Warning* were rated significantly higher than *Audience* in terms of instrumental quality. Also, *Symbols*, *Evaluate* and *Warning* were rated significantly lower than *Audience* in terms of inappropriateness. The user reactions to the designs, as manifested by the scores, would likely affect a test of their actual effect (i.e., emotion regulation). The instrumental quality factor features measures related to positive valence and low arousal, while the inappropriateness factor features the opposite. Thus, for example, based on the scores, the *Audience* design is more likely to anger the user (high arousal, low valence) than *Warning*. That said, none of the design alternatives received particularly high ratings on average: on average, the designs were seen as neither particularly high in instrumental quality nor completely appropriate. At the same time, the variance in respondents’ evaluations is relatively high, which suggests that the participants’ preferences and/or viewpoints varied strongly.

Following this, we studied which background factors predicted the design ratings. While we found no associations between the background variables and inappropriateness ratings, several of the background variables predicted perceived instrumental quality. The results indicate that a decrease in toleration of incivility predicts increased perceived instrumental quality. In the same vein, an increase in instrumental quality was found to be predicted by desire for comment moderation and a decrease in view of how dire the situation is on the news site. Future research could elaborate on these differences. We speculate that a desire for moderation predicts a slight increase in all the ratings because those who wish for more moderation tend to agree with the stated goal of the designs to “help improve discussion around online news articles or help to keep it good”.

We also studied how respondents’ comments on the design features and mechanisms could illuminate the results. We discuss the findings on *Audience*, *Creature*, *Evaluate* and *Regret*, as we consider the responses to these the most illuminating in terms of reasons behind the relatively low ratings. Comparing respondents’ comments on *Audience* and *Creature* suggests that *Audience*’s low instrumental quality and high inappropriateness ratings are largely explained by the form and appearance of the feedback. The respondents did not appear to find the idea of receiving instantaneous feedback on their tone of writing disturbing in itself. This finding aligns with a recent study suggesting that providing users real-time feedback about the quality and language of their contribution in an online news commenting system is appreciated by users [4]. That said, also the novelty of the *Audience* and *Creature* designs may have contributed to their ratings, as people tend to prefer the environment to stay as it already is (i.e., status quo bias) [52].

Evaluate's mid-range instrumental quality ratings could be partially explained by the fact that most respondents were unaware that giving a label to one's emotion has a regulatory effect. This aligns with previous findings that people are mostly unaware of the regulatory effects of affect labeling [63]. That said, the appropriateness score implies that most users do not consider it inappropriate to ask a commenter to tell how they are feeling. Therefore, the design concept warrants further study and could probably be tested on a news site without major loss of users.

Regret's mid-range ratings could be explained by the fact that many respondents believed that using the regret option (to add a label that one regretted their choice of words) would lead to the user being disrespected by others. At the same time, the other features relating to the user regretting their post (the edit and delete option) were perceived more favorably; and a few respondents saw value also in the regret option. This might not be surprising as a significant portion of social media users have posted something they regret [21, 66]. Therefore, further study on possible ways to get some users to edit, remove, or otherwise show regret over their choice of words is warranted.

6.2 Reflection on the Research Process and Methodology

Considering the reliability of the findings, the explorative nature of the study and lack of well-established measurements creates challenges in terms of the reliability of the measurements and, hence, the validity of the statistical associations. In particular, the identified design quality factors are much simpler and fewer in number than we anticipated while operationalizing the various measures. This implies that it was difficult for the respondents to evaluate the proposed solutions. Further, considering the methodological approach in general, we acknowledge that the use of Prolific in recruiting participants for the survey resulted in the over-representation of participants from the UK and other western countries.

Despite these shortcomings, we argue that the methodological choices were justifiable vis-à-vis the set goals for the following reasons. First, the online survey enabled us to reach a large number and spectrum of people who actively comment on online news sites, offering an extensive overall picture of the potential end-users' views. The diverse sample and large number of respondents allowed us to recognize the variance in user perceptions more clearly than with an interview study, for example. Second, the self-operationalized measures managed to inquire about qualities and perspectives that go beyond conventional usability or task load measurements. For example, we obtained a deeper understanding of designs with respect to the potential of the solution in managing emotional reactions. Based on the findings, we will particularly consider studying in more detail designs like Regret, where the user is notified sometime after they have finished writing their comment. Also, we were able to form meaningful factors and factor-based sum variables based on the measures. Third, the qualitative analysis of respondents' reactions to the designs shed light on the ratings.

7 CONCLUSION

This paper provides a user-centric evaluation of eight unconventional design proposals that, through various mechanisms, aim to

support emotion-related self-reflection and emotion regulation in commenting on online news. The paper reports the findings of an online survey, analyzing differences in respondents' preferences across the designs, the respondents' comments on the designs, and the background factors that were associated with the evaluations. The key findings highlight that, while the preferences vary significantly, the participants rated four designs higher than a design where a virtual audience of experts would judge the tone of the writing. The analysis also shows that the perceived instrumental quality of the designs is associated with three background variables. For example, an increased desire for comment moderation was found to predict increased perceived instrumental quality.

All in all, the study advances our understanding of what kind of UI solutions, from the end-user's perspective, may be desirable in terms of improving the quality of online news commenting. We argue that this exploratory study is an important step toward the development of acceptable UI solutions that could effectively mitigate the issue of uncivil behavior in online news commenting. We expect the novel designs, the self-operationalized measures, and the findings to inspire new designs and studies on the role of UI design in mitigating online incivility.

ACKNOWLEDGMENTS

We thank all the participants in the survey study. The research was funded by the Academy of Finland (grants 320767, 320766).

REFERENCES

- [1] James Auger. 2013. Speculative design: Crafting the speculation. *Digit. Creat.* 24, 1 (2013), 11–35. <https://doi.org/10.1080/14626268.2013.767276>
- [2] Jeffrey Bardzell, Shaowen Bardzell, and Erik Stolterman. 2014. Reading critical designs. In *Proc. 32nd Annu. ACM Conf. Hum. factors Comput. Syst. - CHI '14*. <https://doi.org/10.1145/2556288.2557137>
- [3] Svenja Boberg, Tim Schatto-Eckrodt, Lena Frischlich, and Thorsten Quandt. 2018. The moral gatekeeper? Moderation and deletion of user-generated content in a leading news forum. *Media Commun.* 6, 4 New sand Participation through and beyond Proprietary (nov 2018), 58–69. <https://doi.org/10.17645/mac.v6i4.1493>
- [4] Emilie Bossens, David Geerts, Elias Storms, and Jan Boesman. 2022. RHETORiC : an Audience Conversation Tool that Restores Civility in News Comment Sections. In *CHI '22 Ext. Abstr.*
- [5] Emilie Bossens, Elias Storms, and David Geerts. 2021. *Improving the Debate: Interface Elements that Enhance Civility and Relevance in Online News Comments*. Vol. 12935 LNCS. Springer International Publishing, 433–450 pages. https://doi.org/10.1007/978-3-030-85610-6_25
- [6] Alex Bradley, Claire Lawrence, and Eamonn Ferguson. 2018. Does observability affect prosociality? *Proc. R. Soc. B Biol. Sci.* 285, 1875 (mar 2018). <https://doi.org/10.1098/RSPB.2018.0116>
- [7] Margaret M. Bradley and Peter J. Lang. 1994. Measuring emotion: The self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* 25, 1 (mar 1994), 49–59. [https://doi.org/10.1016/0005-7916\(94\)90063-9](https://doi.org/10.1016/0005-7916(94)90063-9)
- [8] Virginia Braun and Victoria Clarke. 2012. Thematic analysis. *APA Handb. Res. methods Psychol. Vol 2 Res. Des. Quant. Qual. Neuropsychol. Biol.* (mar 2012), 57–71. <https://doi.org/10.1037/13620-004>
- [9] Simon Buckingham Shum. 2008. Cohere: Towards Web 2.0 Argumentation. In *Proc. COMMA'08 2nd Int. Conf. Comput. Model. Argument.* 97–108. <https://doi.org/10.5860/choice.51-2973>
- [10] Roser Cañigueral and Antonia F.de C. Hamilton. 2019. Being watched: Effects of an audience on eye gaze and prosocial behaviour. *Acta Psychol. (Amst)* 195 (apr 2019), 50–63. <https://doi.org/10.1016/j.actpsy.2019.02.002>
- [11] Gina Masullo Chen and Shuning Lu. 2017. Online Political Discourse: Exploring Differences in Effects of Civil and Uncivil Disagreement in News Website Comments. *J. Broadcast. Electron. Media* 61, 1 (2017), 108–125. <https://doi.org/10.1080/08838151.2016.1273922>
- [12] 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 (jun 2017), 181–188. <https://doi.org/10.1016/j.chbh.2017.02.010>
- [13] Kevin Coe, Kate Kenski, and Stephen A. Rains. 2014. Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *J. Commun.* 64,

- 4 (2014), 658–679. <https://doi.org/10.1111/jcom.12104> arXiv:1708.02002
- [14] Thomas Davidson, Dana Wamsley, Michael Macy, and Ingmar Weber. 2017. Automated Hate Speech Detection and the Problem of Offensive Language. In *Proc. 11th Int. Conf. Web Soc. Media, IJWSM 2017*. AAAI Press, 512–515. arXiv:1703.04009 <http://arxiv.org/abs/1703.04009>
- [15] Nicholas Diakopoulos and Mor Naaman. 2011. Towards Quality Discourse in Online News Comments Human Factors. In *Proc. ACM 2011 Conf. Comput. Support. Coop. Work.* 133–142. <https://doi.org/10.1.1.188.3516>
- [16] Tawanna Dillahunt, Geoff Becker, Jennifer Mankoff, and Robert Kraut. 2008. Motivating Environmentally Sustainable Behaviour Changes with a Virtual Polar Bear. In *Pervasive 2008 Work. Proc.* 58–62. <https://doi.org/10.1504/IJARGE.2002.000023>
- [17] Christine DiStefano, Min Zhu, and Diana Mindrilă. 2009. Understanding and using factor scores: Considerations for the applied researcher. *Pract. Assessment, Res. Eval.* 14, 20 (2009).
- [18] Eleanor Eytam. 2020. Effect of Visual Design on the Evaluation of Technology- vs. Design-Based Novel Interactive Products. *Interact. Comput.* 32, 3 (2020), 296–315. <https://doi.org/10.1093/iwc/iwaa021>
- [19] Rui Fan, Onur Varol, Ali Varamesh, Alexander Barron, Ingrid A. van de Leemput, Marten Scheffer, and Johan Bollen. 2019. The minute-scale dynamics of online emotions reveal the effects of affect labeling. *Nat. Hum. Behav.* 3, 1 (jan 2019), 92–100. <https://doi.org/10.1038/s41562-018-0490-5>
- [20] Nancy Fraser. 1990. Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy. *Soc. Text* 25/26 (1990), 56. <https://doi.org/10.2307/466240>
- [21] Jake Gammon. 2015. Social media blunders cause more damage to important relationships today than two years ago. <https://today.yougov.com/topics/lifestyle/articles-reports/2015/07/22/social-media-blunders-cause-more-damage-important->
- [22] Becky Gardiner, Mahana Mansfield, Ian Anderson, Josh Holder, Daan Louter, and Monica Ulmanu. 2016. The dark side of Guardian comments | Technology | The Guardian. <https://www.theguardian.com/technology/2016/apr/12/the-dark-side-of-guardian-comments/>
- [23] David Gefen, Detmar Straub, and Marie-Claude Boudreau. 2000. Structural Equation Modeling and Regression: Guidelines for Research Practice. *Commun. Assoc. Inf. Syst.* 4, 1 (oct 2000), 7. <https://doi.org/10.17705/1cais.00407>
- [24] Tarleton Gillespie. 2018. *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press, New Haven & London.
- [25] Grammarly. 2022. Tone Detector | Grammarly. <https://www.grammarly.com/tone>
- [26] James Grimmelmann. 2015. The Virtues of Moderation. *Yale J. Law Technol.* 17 (2015). <https://heinonline.org/HOL/Page?handle=hein.journals/yjolt17&iid=42&div=&collection=>
- [27] Kirsikka Grön and Matti Nelimarkka. 2020. Party Politics, Values and the Design of Social Media Services. In *Proc. ACM Human-Computer Interact.*, Vol. 4. <https://doi.org/10.1145/3415175>
- [28] James J. Gross. 2015. Emotion Regulation: Current Status and Future Prospects. *Psychol. Inq.* 26, 1 (jan 2015), 1–26. <https://doi.org/10.1080/1047840X.2014.940781>
- [29] Ståle Grut. 2017. With a quiz to comment, readers test their article comprehension. <https://nrkbeta.no/2017/08/10/with-a-quiz-to-comment-readers-test-their-article-comprehension/>
- [30] Amos Guioira and Elizabeth A. Park. 2017. Hate Speech on Social Media. *Philosophia (Mendoza)*. 45, 3 (sep 2017), 957–971. <https://doi.org/10.1007/s11406-017-9858-4>
- [31] Jürgen Habermas. 1991. *The Structural Transformation of the Public Sphere: An inquiry into a category of bourgeois society*. Vol. 68. The MIT Press, Cambridge. arXiv:arXiv:1202.3162v2
- [32] Marc Hassenzahl, Michael Burmester, and Franz Koller. 2003. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. Vieweg+Teubner Verlag, 187–196. https://doi.org/10.1007/978-3-322-80058-9_19
- [33] Alina Huldigtren, Pascal Wiggers, and Catholijn M. Jonker. 2014. Designing for self-reflection on values for improved life decision. *Interact. Comput.* 26, 1 (2014), 27–45. <https://doi.org/10.1093/iwc/iwt025>
- [34] Jigsaw. 2017. Perspective API. <https://www.perspectiveapi.com/#/home>
- [35] Leon Karlsen Johannessen. 2017. The Young Designer’s Guide to Speculative and Critical Design.
- [36] Evangelos Karapanos, John Zimmerman, Jodi Forlizzi, and Jean Bernard Martens. 2009. User experience over time: An initial framework. *Conf. Hum. Factors Comput. Syst. - Proc.* (2009), 729–738. <https://doi.org/10.1145/1518701.1518814>
- [37] Joel Kiskola, Thomas Olsson, Heli Väättäjä, Aleksii H. Syrjämäki, Anna Rantasila, Poika Isokoski, Mirja Ilves, and Veikko Surakka. 2021. Applying critical voice in design of user interfaces for supporting self-reflection and emotion regulation in online news commenting. In *CHI ’21 Proc. 2021 CHI Conf. Hum. Factors Comput. Syst.* Association for Computing Machinery. <https://doi.org/10.1145/3411764.3445783>
- [38] Thomas B. Ksiazek. 2018. Commenting on the News. *Journal. Stud.* 19, 5 (apr 2018), 650–673. <https://doi.org/10.1080/1461670X.2016.1209977>
- [39] Thomas B Ksiazek, Limor Peer, and Kevin Lessard. 2016. User engagement with online news: Conceptualizing interactivity and exploring the relationship between online news videos and user comments. *New Media Soc.* 18, 3 (mar 2016), 502–520. <https://doi.org/10.1177/1461444814545073>
- [40] James J Lin, Lena Mamykina, Silvia Lindtner, Gregory Delajoux, and Henry B Strub. 2006. Fish n’Steps Encouraging Physical Activity with an interactive computer game. In *Int. Conf. ubiquitous Comput.* Springer Berlin Heidelberg, 261–278. https://www.ics.uci.edu/~lindtner/documents/Lin_FishnSteps2006.pdf
- [41] Rousiley C. M. Maia and Thaiane A. S. Rezende. 2016. Respect and Disrespect in Deliberation across the Networked Media Environment: Examining Multiple Paths of Political Talk. *J. Comput. Commun.* 21, 2 (mar 2016), 121–139. <https://doi.org/10.1111/JCC4.12155>
- [42] Iris B. Mauss and Michael D. Robinson. 2009. Measures of emotion: A review. <https://doi.org/10.1080/02699930802204677>
- [43] Hans K. Meyer and Michael Clay Carey. 2014. In Moderation. *Journal. Pract.* 8, 2 (mar 2014), 213–228. <https://doi.org/10.1080/17512786.2013.859838>
- [44] Douglas C. Montgomery, Elizabeth A. Peck, and G. Geoffrey Vining. 2021. *Introduction to linear regression analysis*. John Wiley & Sons.
- [45] Courtney Naples, Joel Tetreault, Aashish Pappu, Enrica Rosato, and Brian Provenza. 2017. Finding Good Conversations Online: The Yahoo News Annotated Comments Corpus. (2017), 13–23. <https://doi.org/10.18653/v1/w17-0802>
- [46] Adewale Obadimu, Esther Mead, Muhammad Nihal Hussain, and Nitin Agarwal. 2019. Identifying toxicity within youtube video comment. In *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, Vol. 11549 LNCS. Springer Verlag, 214–223. https://doi.org/10.1007/978-3-030-21741-9_22
- [47] Stefan Palan and Christian Schitter. 2018. Prolific.ac—A subject pool for online experiments. *J. Behav. Exp. Financ.* 17 (mar 2018), 22–27. <https://doi.org/10.1016/j.jbef.2017.12.004>
- [48] Zhenhui Peng, Taewook Kim, and Xiaojuan Ma. 2019. GremoBot: Exploring emotion regulation in group chat. In *Proc. ACM Conf. Comput. Support. Coop. Work. CSCW*. Association for Computing Machinery, New York, New York, USA, 335–340. <https://doi.org/10.1145/3311957.3359472>
- [49] Fabian Prochazka, Patrick Weber, and Wolfgang Schweiger. 2018. Effects of Civility and Reasoning in User Comments on Perceived Journalistic Quality. *Journal. Stud.* 19, 1 (jan 2018), 62–78. <https://doi.org/10.1080/1461670X.2016.1161497>
- [50] Ian Rowe. 2015. Civility 2.0: a comparative analysis of incivility in online political discussion. *Inf. Commun. Soc.* 18, 2 (feb 2015), 121–138. <https://doi.org/10.1080/1369118X.2014.940365>
- [51] Minna Ruckenstein and Linda Lisa Maria Turunen. 2020. Re-humanizing the platform: Content moderators and the logic of care. *New Media Soc.* 22, 6 (jun 2020), 1026–1042. <https://doi.org/10.1177/1461444819875990>
- [52] William Samuelson and Richard Zeckhauser. 1988. Status quo bias in decision making. *J. Risk Uncertain.* 1988 11 1, 1 (mar 1988), 7–59. <https://doi.org/10.1007/BF00055564>
- [53] 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. <https://doi.org/10.1145/3290605.3300836>
- [54] Guy Simon. 2020. OpenWeb tests the impact of “nudges” in online discussions. <https://www.openweb.com/blog/openweb-improves-community-health-with-real-time-feedback-powered-by-jigsaws-perspective-api/>
- [55] Seohae Sohn, Ho Chung Chung, and Namkee Park. 2019. Private Self-Awareness and Aggression in Computer-Mediated Communication: Abusive User Comments on Online News Articles. *Int. J. Hum. Comput. Interact.* 35, 13 (2019), 1160–1169. <https://doi.org/10.1080/10447318.2018.1514822>
- [56] Nina Springer, Ines Engelmann, and Christian Pfaffinger. 2015. User comments: motives and inhibitors to write and read. *Information. Commun. Soc.* 18, 7 (jul 2015), 798–815. <https://doi.org/10.1080/1369118X.2014.997268>
- [57] Natalie Jomini Stroud, Emily Van Duyn, and Cynthia Peacock. 2016. News commenters and news comment readers.
- [58] John Suler. 2004. The Online Disinhibition Effect. *CYBERPSYCHOLOGY, Behav.* 7, 3 (jul 2004), 321–326. <https://doi.org/10.1089/109493104129129195>
- [59] Aleksii H Syrjämäki, Mirja Ilves, Poika Isokoski, Joel Kiskola, Anna Rantasila, Thomas Olsson, Gary Bente, and Veikko Surakka. 2022. Emotionally Toned Online Discussions Evoke Subjectively Experienced Emotional Responses. (2022).
- [60] Richard H. Thaler and Cass R. Sunstein. 2009. *Nudge: Improving decisions about health, wealth, and happiness*. Penguin, New York.
- [61] Bruce M. Tharp and Stephanie Tharp. 2019. *Discursive design : critical, speculative, and alternative things*. MIT Press, Cambridge, Massachusetts & London, England, 617 pages.
- [62] Kamil Topal, Mehmet Koyuturk, and Gultekin Ozsoyoglu. 2016. Emotion -and area-driven topic shift analysis in social media discussions. In *Proc. 2016 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Mining,ASONAM 2016*. Institute of Electrical and Electronics Engineers Inc., 510–518. <https://doi.org/10.1109/ASONAM.2016.7752283>

- [63] Jared B. Torre and Matthew D. Lieberman. 2018. Putting Feelings Into Words: Affect Labeling as Implicit Emotion Regulation. *Emot. Rev.* 10, 2 (apr 2018), 116–124. <https://doi.org/10.1177/1754073917742706>
- [64] Joseph B. Walther. 1993. Impression development in computer-mediated interaction. *West. J. Commun.* 57, 4 (dec 1993), 381–398. <https://doi.org/10.1080/10570319309374463>
- [65] Joseph B. Walther. 1996. Computer-Mediated Communication. *Communic. Res.* 23, 1 (feb 1996), 3–43. <https://doi.org/10.1177/009365096023001001>
- [66] Yang Wang, Saranga Komanduri, Pedro Giovanni Leon, Gregory Norcie, Alessandro Acquisti, and Lorrie Faith Cranor. 2011. "I regretted the minute I pressed share": A Qualitative Study of Regrets on Facebook. In *Proc. Seventh Symp. Usable Priv. Secur. - SOUPS '11*. ACM Press, New York, New York, USA. <https://doi.org/10.1145/2078827>
- [67] Yang Wang, Pedro Giovanni Leon, Kevin Scott, Xiaoxuan Chen, Alessandro Acquisti, and Lorrie Faith Cranor. 2013. Privacy nudges for social media. In *Proc. 22nd Int. Conf. world wide web*, 22, 763–770. <https://doi.org/10.1145/2487788.2488038>
- [68] Florian Winterlin, Tim Schatto-Eckrodt, Lena Frischlich, Svenja Boberg, and Thorsten Quandt. 2020. How to Cope with Dark Participation: Moderation Practices in German Newsrooms. *Digit. Journal.* (jul 2020), 1–21. <https://doi.org/10.1080/21670811.2020.1797519>
- [69] J David Wolfgang. 2018. Taming the 'trolls': How journalists negotiate the boundaries of journalism and online comments. *Journalism* (mar 2018), 146488491876236. <https://doi.org/10.1177/1464884918762362>
- [70] 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. In *Conf. Hum. Factors Comput. Syst. - Proc.*, Vol. 2018-April. 1–12. <https://doi.org/10.1145/3173574.3174150>
- [71] Tai-Yee Wu and David Atkin. 2017. Online News Discussions: Exploring the Role of User Personality and Motivations for Posting Comments on News. *Journal. Mass Commun. Q.* 94, 1 (mar 2017), 61–80. <https://doi.org/10.1177/1077699016655754>
- [72] Xenophon, Carleton L. Brownson, E. C. Marchant, O. J. Todd, and Walter Miller. 1979. *Xenophon in seven volumes*. Harvard University Press, Cambridge, MA.
- [73] JungKyoan Yoon, Shuran Li, Yu Hao, and Chajoon Kim. 2019. Towards Emotional Well-Being by Design. In *Pervasive Heal. 19 13th EAI Int. Conf. Pervasive Comput. Technol. Healthc.* 351–355. <https://doi.org/10.1145/3329189.3329227>
- [74] Marc Ziegele, Mathias Weber, Oliver Quiring, and Timo Breiner. 2018. The dynamics of online news discussions: effects of news articles and reader comments on users' involvement, willingness to participate, and the civility of their contributions. *Information, Commun. Soc.* 21, 10 (oct 2018), 1419–1435. <https://doi.org/10.1080/1369118X.2017.1324505>

A APPENDICES

A.1 Full Surveys

The following lists the survey questions following the survey structure.

–Pre-survey–

Title: Survey on commenting online news

Thank you for your interest in this research! This is a pre-survey that is used for selecting the participants for an actual research survey.

The purpose of the study. We are interested in how often you read and/or comment news articles on online news sites and social media. It does not matter which devices you are using (desktop computer, laptop, mobile device, etc.). Also, any professionally produced news content counts (by commercial media corporations, public broadcasting organizations, national news sites, etc.).

Your participation in the study is fully voluntary. You may stop the survey at any moment by closing the page, in which case the survey tool will not send any of your answers.

Confidentiality, data processing and retention. All the data will be anonymized and used only for research purposes as required by the European Union's General Data Protection Regulation (GDPR).

After the study has ended, the data will be stored and managed carefully according to national recommendations on research integrity. This survey should only take a minute or so to fill.

Responsible researchers: anonymized.

By checking this box, I confirm I have read the study description and consent to participate in this study: Y or N

Please enter your Prolific ID if it has not been entered automatically

Please consider your use of online news sites

I read news articles on online news sites Several times a day, Daily, Weekly, Monthly, Yearly, Less than once a year, Never

I read at least some of the comments to news articles on online news sites Several times a day, Daily, Weekly, Monthly, Yearly, Less than once a year, Never

I comment on news articles on online news sites – Several times a day, Daily, Weekly, Monthly, Yearly, Less than once a year, Never

Please list your 1 to 3 most frequently visited news sites where you typically also read comments or add your own comments

Please consider your use of social media services, such as Facebook or Twitter

I post and comment news articles on social media services, such as Facebook or Twitter Several times a day, Daily, Weekly, Monthly, Yearly, Less than once a year, Never

I comment on others' posts about news articles on social media Several times a day, Daily, Weekly, Monthly, Yearly, Less than once a year, Never

I am interested in participating in a follow-up study: Y or N

DESIGN SURVEY

Title: Survey on improving discussion around online news articles

Thank you for your interest in this research!

The purpose of the study. This survey will ask about your behavior and attitudes related to commenting news on online news sites. You will be shown two speculative prototypes that might help improve discussion around online news articles or help to keep it good and we will ask what you think about them. Please answer honestly and truthfully to all the questions, rather than in a way that you think we would like to hear.

Your participation in the study is fully voluntary. You may stop the survey at any moment by closing the page, in which case the survey tool may send some or all the answers you have given until that moment but as a general rule, you will not be compensated. However, under certain circumstances we may still choose to pay partial compensation.

Confidentiality, data processing and retention. All the data will be anonymized and used only for research purposes as required by the European Union's General Data Protection Regulation (GDPR). Your answers to open questions may be reproduced in whole or in part for use in presentations or written results of this study. However, your level of education, age, or any other identifier will never be revealed outside of the research team. After the study has ended, the data will be stored and managed carefully according to national recommendations on research integrity.

This survey asks for your consent to participate as well as some background information. This survey should take about 20 minutes to fill.

Responsible researchers: anonymized

By checking this box I confirm I have read the study description and consent to participate in this study: Y or N

Please enter your Prolific ID if it has not been entered automatically

Basic background questions

Age: dropdown menu 18 to 99

Gender: male, female, other, prefer not to say

Level of education: Secondary education (e.g. GED or GCSE) (1), High School diploma A levels (2), Technical or community college (3), Undergraduate degree (BA or BSc or other) (4), Graduate degree (MA or MSc or MPhil or other) (5), Doctorate degree (PhD or other) (6), Don't know / not applicable (0)

Current country of residence

On commenting history

Q: Considering my history of commenting on various news sites, I believe that I have written altogether: More than 10,000 comments, More than 1000 comments, More than 100 comments, More than 10 comments, Less than 10 comments

Please note that you cannot return to the previous page of the survey. Returning may prevent you from finishing the survey or you may even lose your answers. This means that you cannot change your answers after you have clicked "Next." If you accidentally press back in your browser, the browser may ask you to re-submit data or page. If this happens, follow the browser's instructions.

About your views on commenting

In the following questions, please consider your experiences of the discussion on the news site where you are most actively reading and posting comments (using any device). Spend a moment to choose the one that you are most active on.

Now, spend a moment thinking about a typical comment section on a news article and how it feels to read the comments and take part in the discussion. For example, think back recent articles or comments that you can remember particularly well.

Please name the news site that you are thinking about. Note that it must be identifiable, so please, for example, provide a link if the name can be misunderstood

Please select your level of agreement or disagreement with the following statements:

(Strongly disagree, Disagree, Somewhat disagree, Neither agree nor disagree, Somewhat agree, Agree, Strongly agree)

[What the variables below could measure: View on the situation]

The comments on news articles are generally of high quality

The comments on news articles are respectful

The comments on news articles include inappropriate language
Trolling and other intentional misbehavior is common in the commenting section

The people commenting on the news are mindful of others when expressing their opinions

[What the variables below could measure: Views the news site provides a stable commenting environment]

The news site does not encourage respectful commenting

Overall, the news site feels like a place where disrespectful commenting simply does not belong The news site has moderation practices that ensure the quality of commenting

[What the variables below could measure: Toleration of incivility]

• If I see disrespectful comments on the news site, I will get anxious

• If I see inappropriate comments on the news site, it will bother me

• If I see hateful speech in the comments, I will not be bothered
[What the variables below could measure: Wish for more content moderation]

• Publishing inappropriate comments is a problem that should be taken more seriously on this news site

• The news site should moderate the discussion more than currently

• Inappropriate comments get quickly removed or are not published at all

Which of the following options for commenting would be the best on this news site? (Radio buttons)

• All news articles have a comment section

• Selected articles on specific topics have a comment section

• None of the news articles have a comment section

Now, please consider your commenting behavior in general
Consider your interests to write comments on various news sites. Please select your level of agreement or disagreement with the following statements:

(Strongly disagree, Disagree, Somewhat disagree, Neither agree nor disagree, Somewhat agree, Agree, Strongly agree)

[What the variables below could measure: Tends to be drawn in to comment by controversy]

• I tend to participate in the discussion only when the discussion is heated

• I tend to comment on news articles on topics that are controversial

• I typically comment on articles regardless of what the earlier discussion is like

[What the variables below could measure: Is an influencer of sorts]

• I am typically one of the first to comment on a new article

• My comments typically receive many likes or upvotes

• I tend to reply to others' comments

[What the variables below could measure: Acts on emotion in commenting]

• When reading others' inappropriate comments, I tend to write inappropriate responses

• When commenting, I tend to act based on my intuition and avoid overthinking my response

• I carefully think how others might interpret and feel about my comment

Motivations to read and write comments

Please consider what motivates you to read comments. Mark how often the different motivations are present when you read comments to news articles.

(Never, Rarely, Sometimes, Often, Always)

[Adapted from Springer et al.]

["Cognitive motive" items]

• I read comments to broaden my knowledge base

• I read comments to better understand others

["Social-integrative motive" items]

• I read comments to be part of the community

• I read comments to meet other users

["Entertainment motive" items]

- I read comments because it is entertaining to see others fight
- I read comments for a pastime

Please consider what motivates you to write comments. Mark how often the different motivations are present when you comment on a news article. Again, focus on the news site that you are most familiar with.

["Cognitive motive" items]

- I comment to understand events that are happening
- I comment to better understand others

["User-journalistic interactivity" items]

• I comment to show disagreement with the article, parts of it, or the journalist's opinion

- I comment to bring in my opinion

["User-user interactivity" items]

- I comment to discuss with others
- I comment because I enjoy to see that others think the same way I do

["Personal identity" items]

- I comment to establish my personal identity
- I comment to promote or publicize my expertise

Introduction to the Designs

Online news publishers have long sought means to improve the quality of comments on their news commenting sections. It has been argued that the discussion around news articles on news sites is too often disrespectful, uncivil, or otherwise impolite. Various solutions could be considered to solve these problems.

Next, we will show you two examples of different ways to possibly influence the commenting and reading behavior of the news site visitors. We want to understand how you experience them and what kind of opportunities or risks you see in them. Please, note that these are merely speculative prototypes created out of academic interests, rather than products that any news site would soon take into use.

(–The following block of questions were asked also for the second design shown to participant–)

The First Design / The Second Design and Questions About It

Please view this series of pictures of the design, and answer the questions below.

This design will later be referred to as: [short name, e.g., Highlight Emotions in Comments (see the list of short names at the end of this document)]

[Pictures of the design here]

• (Open question) How would you describe your immediate reaction to this solution? How do you feel about it?

On emotional impact

[Emotion dimension scales adapted from Bradley and Lang]

Please consider how the solution makes you feel.

The left end of the scale (–4) means that you feel completely unhappy, annoyed, unsatisfied, melancholic, despaired, or bored. The right end (+4) refers to the completely opposite feeling, feeling completely happy, pleased, satisfied, contented, or hopeful.

(Unpleasant –4, –3, –2, –1, 0, +1, +2, +3, +4 Pleasant)

The left end of the scale (–4) means that you feel completely relaxed, calm, sluggish, dull, sleepy, or unaroused. The right end (+4) refers to the completely opposite feeling, feeling completely stimulated, excited, frenzied, jittery, wide-awake, or aroused.

(Calm –4, –3, –2, –1, 0, +1, +2, +3, +4 Aroused)

The left end of the scale (–4) means that you feel completely controlled, influenced, cared-for, awed, submissive, or guided. The right end (+4) refers to the completely opposite feeling, feeling completely in control, influential, important, dominant, autonomous, or controlling.

(Controlled –4, –3, –2, –1, 0, +1, +2, +3, +4 In-control)

Based on your first impression, please select your level of agreement or disagreement with the following statements:

(Strongly disagree, Disagree, Somewhat disagree, Neither agree nor disagree, Somewhat agree, Agree, Strongly agree)

Discursive dissonance items (clarity, feasibility, familiarity, truthfulness, desirability). Inspired by Tharp and Tharp, 2019. Discursive Design

[What the variables below could measure: Clarity]

- I feel that it is clear what the solution aims at

- I feel that it is unclear how the solution would actually work

[What the variables below could measure: Feasibility]

• I feel that it is feasible for this to become a real, functioning solution

• I feel that this does not solve the problem of disrespectful commenting

[What the variables below could measure: Familiarity]

- The solution feels strange to me

- I have never seen such a solution before

[What the variables below could measure: Truthfulness]

• I feel the designer who made this is trying to deceive or ridicule me

- I feel that the solution is sarcastic or a spoof

[What the variables below could measure: Desirability]

- Overall, I find the solution desirable

- The solution matches what kind of solutions I wish for

On design qualities

[Adapted from AttrakDiff]

Please compare the solution to your experiences of using news sites and their commenting features.

To me, the solution feels...

- (conventional 1, 2, 3, 4, 5 inventive)

- (unimaginative 1, 2, 3, 4, 5 creative)

- (cautious 1, 2, 3, 4, 5 bold)

On behavioral effects

[These questions were presented only with the following 'reading-type' designs: Symbols, Highlight, Philosophy, Warning]

Please consider how the presented solution might affect your own behavior in terms of reading comments on online news sites:

(Strongly disagree, Disagree, Somewhat disagree, Neither agree nor disagree, Somewhat agree, Agree, Strongly agree)

• If an earlier comment annoyed me, this solution would help me avoid writing an angry reply

• The solution would help me take an objective and neutral perspective to reading the comments

• The solution would help me to decide whether I want to read the comments

[The following questions were presented only with the following 'writing-type' designs: Audience, Creature, Evaluate, Regret]

Please consider how the presented solution might affect your own behavior in commenting news on online news sites:

- The solution would help me to write more respectful comments
- The solution would affect how I phrase my comments
- The solution would not influence my writing style

[The following questions presented with all designs]

Please consider how the presented solution might affect your behavioral tendencies and whether it would work in practice or not.

[What the variables below could measure: Effect on emotion regulation in general]

- The solution would help me manage my emotional reactions
- If I was angry, the solution would make me even angrier
- The solution would have a calming effect on me

[What the variables below could measure: Effect on commenting activity]

- If this solution was implemented, I would take part in news commenting more actively
- I would likely comment less often on news if this solution was implemented
- This solution would likely engage me in more active discussion on news articles

[What the variables below could measure: Feasibility]

- The risks that the solution introduces are higher than its benefits

- The solution would not work in the long-term
- The solution would be accepted on the news sites I typically use

[What the variables below could measure: Freedom of speech]

- The solution would violate my freedom of speech too much to be acceptable
- The solution would help me express my opinions more freely

Misuse. If you expect that the solution would likely be misused, please tell how (free choice, text area)

(–End of the block of questions that were repeated for the second design–)

Questions on the Designs

Now, consider the two different solutions that you saw: XXX XXX (names in the end of this document). Which of them you found as the better solution for improving the commenting culture on online news?

[] XXX [] XXX

Why? If you are not sure why, please write “unsure.”

What were the strengths of the better design (e.g., effective in solving the problem, useful for self-reflection, easy to understand and use)?

What were the weaknesses of the worse design (e.g., unacceptable, too weird, hard to imagine them being used on the news sites that you know)?

*Explanation of the design naming scheme for the survey:

Reading/writing type (R/W), number, name in this paper, descriptive name shown to participants in the survey

R, 1, Highlight, Highlight Emotions in Comments

W, 2, Creature, Animated Creature

R, 3, Symbols, Feedback through Symbols

W, 4, Evaluate, Share feelings to comment

R, 5, Philosophy, Problematic Comments Get an Icon

W, 6, Regret, Option to Regret

R, 7, Warning, Warning About the Comments

W, 8, Audience, Virtual Audience of Experts

A.2 Designs as they were shown in the survey

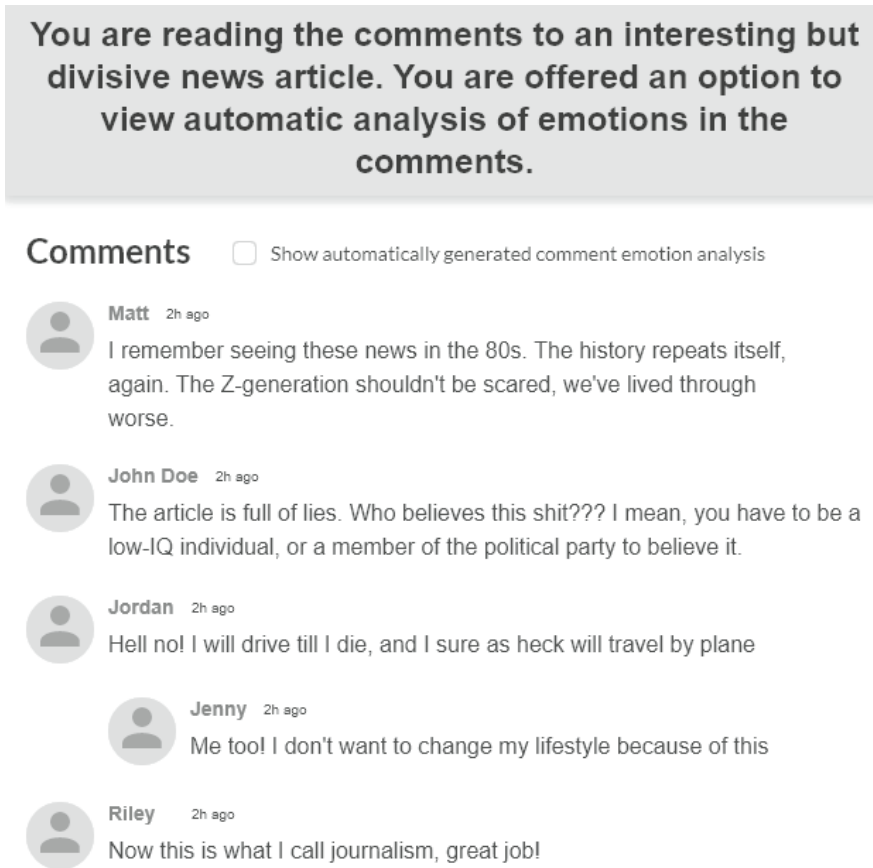


Figure 4: Highlight part 1/2. The following freely available resources were used in making the designs: Semantic UI kit. Icons: Font Awesome, Ionic and Feather.

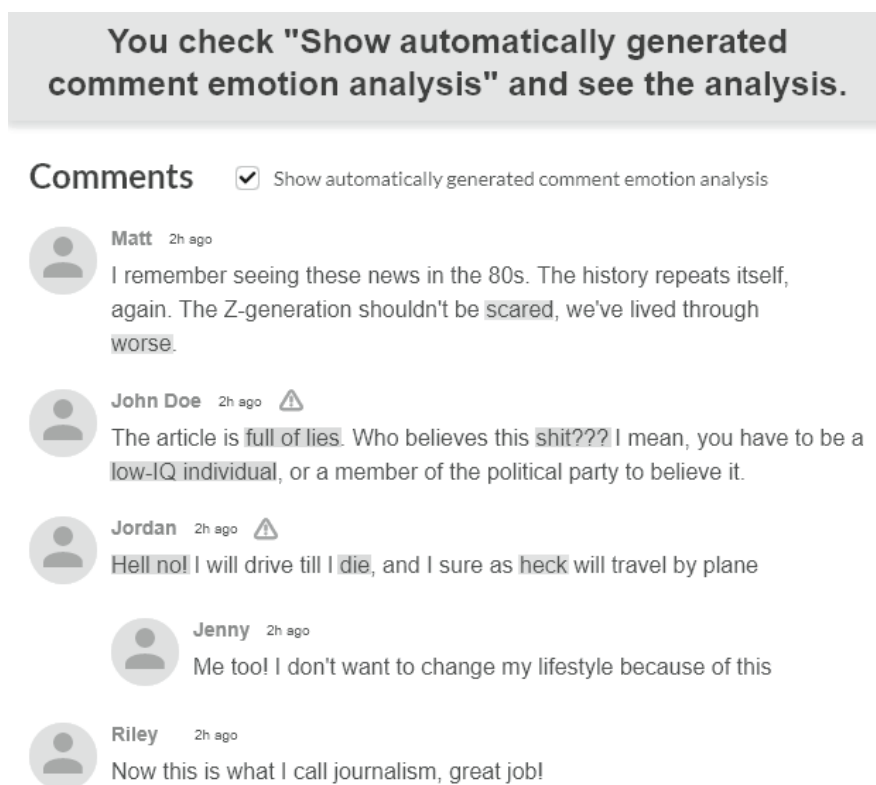


Figure 5: Highlight part 2/2. The following freely available resources were used in making the designs: Semantic UI kit. Icons: Font Awesome, Ionic and Feather.

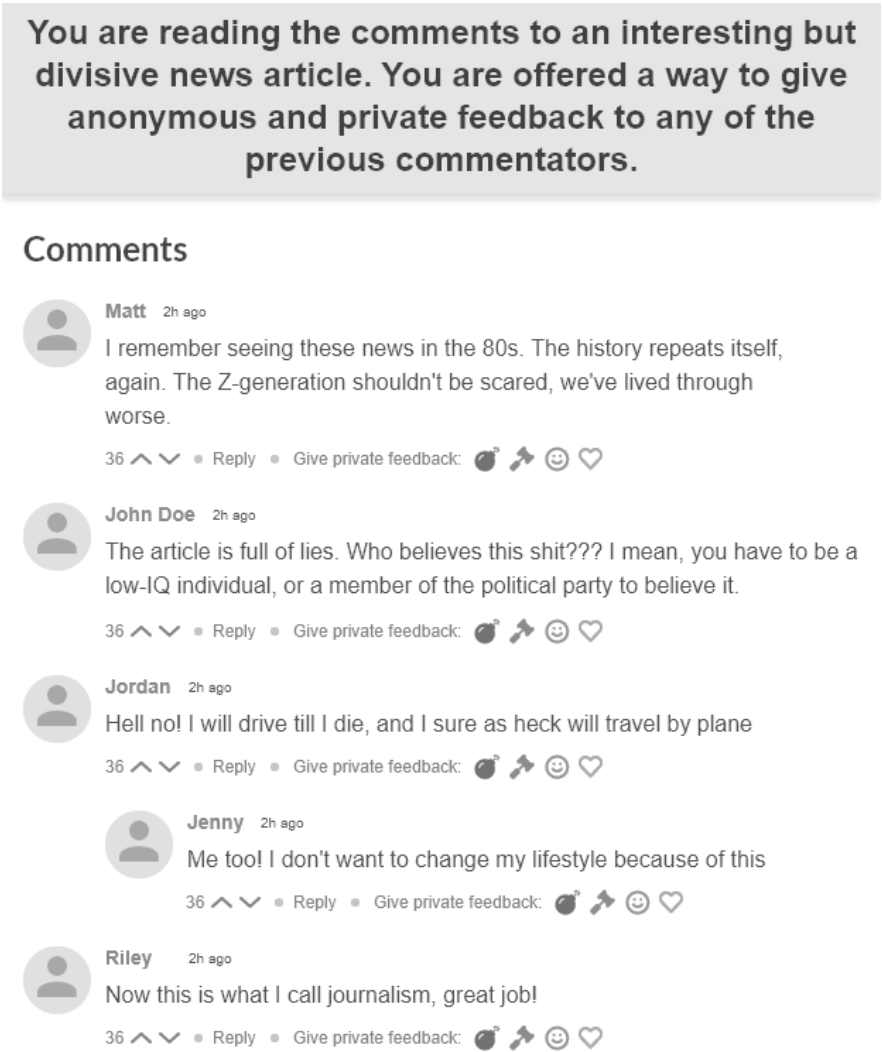


Figure 6: Symbols part 1/3. The following freely available resources were used in making the designs: Semantic UI kit. Icons: Font Awesome, Ionic and Feather.

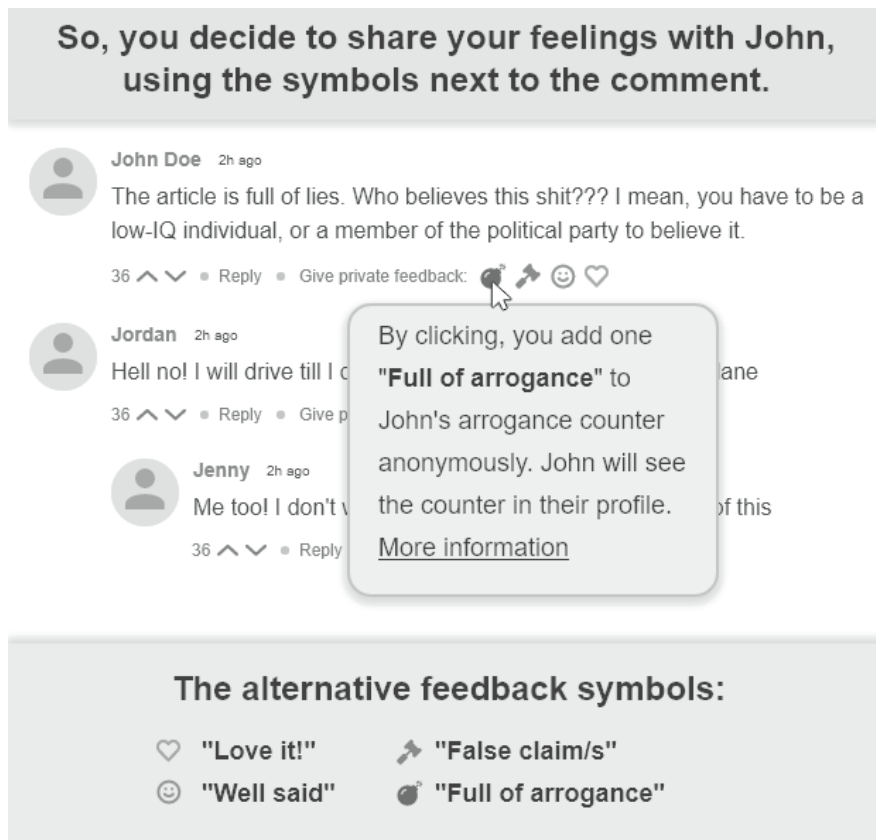


Figure 7: Symbols part 2/3. The following freely available resources were used in making the designs: Semantic UI kit. Icons: Font Awesome, Ionic and Feather.

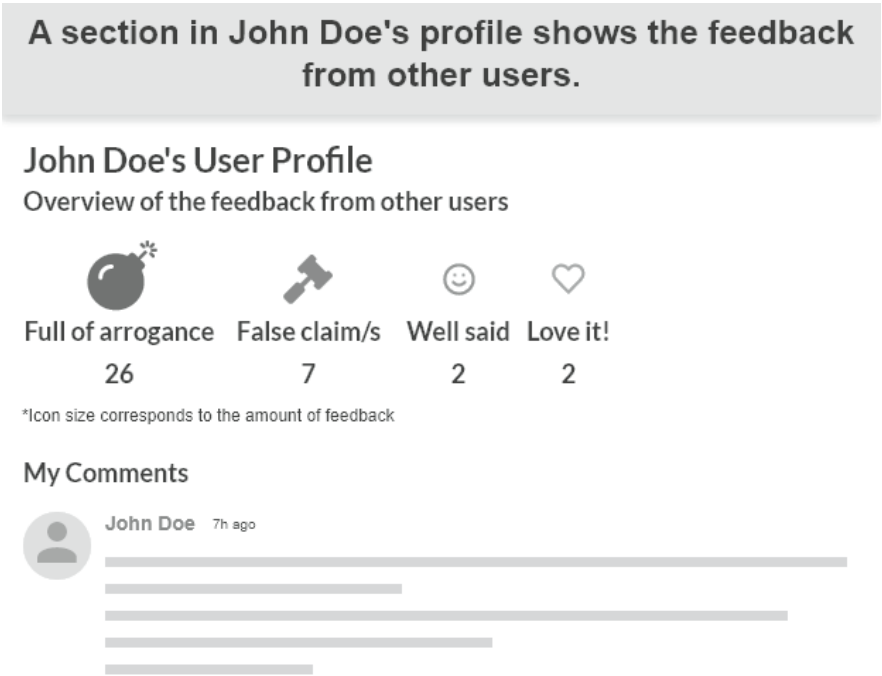




Figure 9: Creature part 1/2. The following freely available resources were used in making the designs: Semantic UI kit. Icons: Font Awesome, Ionic and Feather.

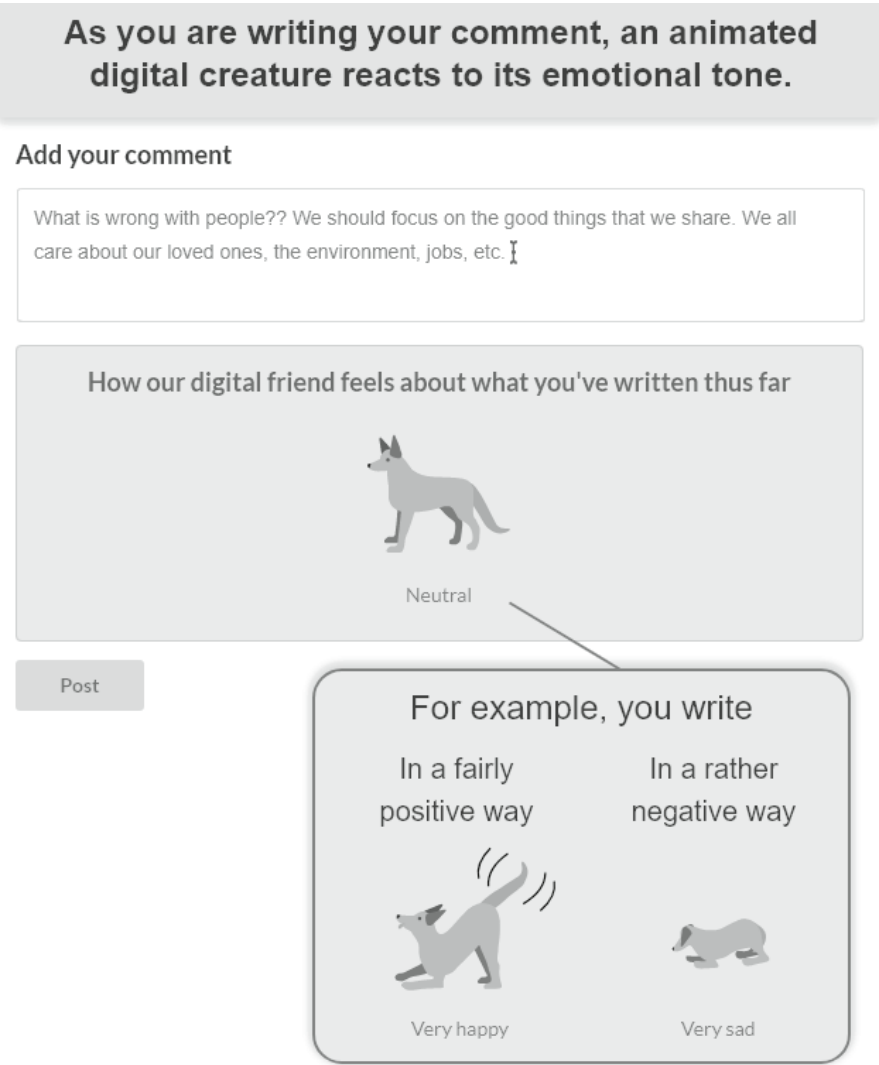


Figure 10: Creature part 2/2. The following freely available resources were used in making the designs: Semantic UI kit. Icons: Font Awesome, Ionic and Feather.

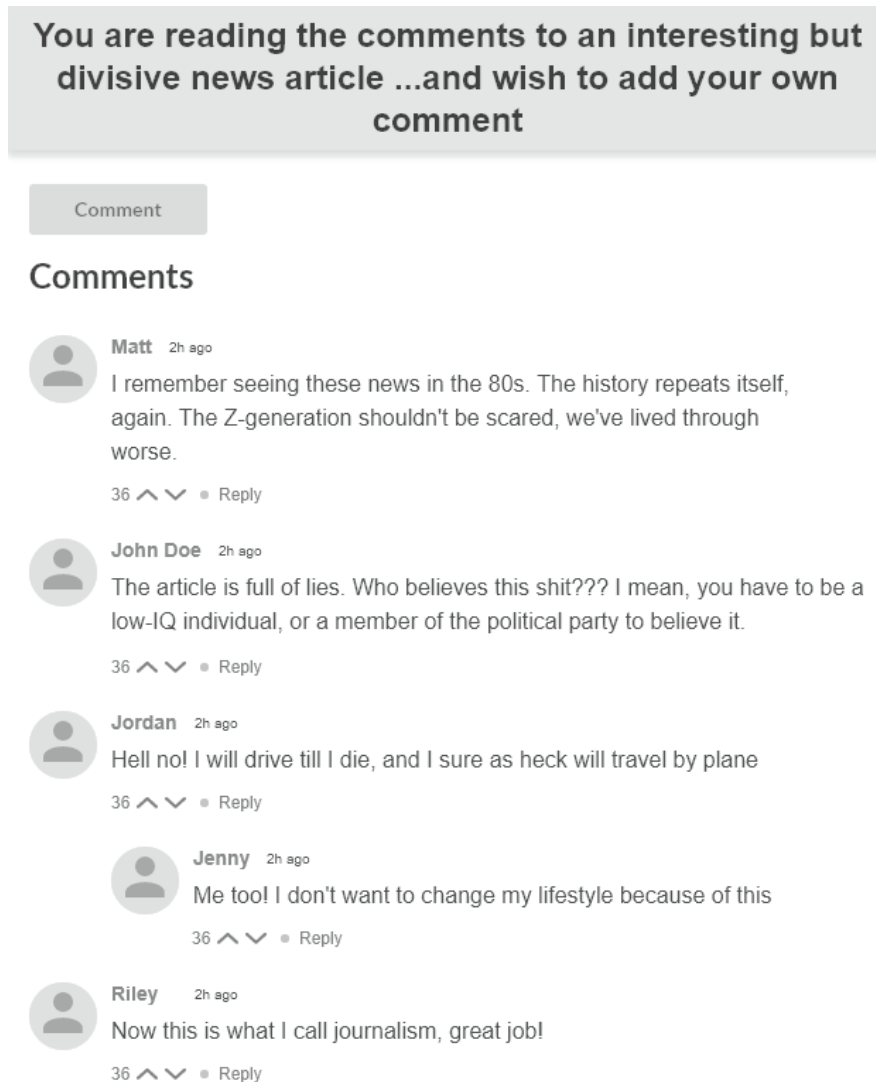





Figure 11: Evaluate part 1/2. The following freely available resources were used in making the designs: Semantic UI kit. Icons: Font Awesome, Ionic and Feather.


When you click to "Comment", you first need to tell how you feel before adding your comment.


1. Please tell which of the following best describes how you feel after reading the news:


Very negative


Negative


"meh"



Positive


Very positive


2. Add your comment

Post






Figure 12: Evaluate part 2/2. The following freely available resources were used in making the designs: Semantic UI kit. Icons: Font Awesome, Ionic and Feather.

You are reading the comments to an interesting but divisive news article. The problematic comments and comment threads are marked with an  icon.






Comments

- 
Matt 2h ago





I remember seeing these news in the 80s. The history repeats itself, again. The Z-generation shouldn't be scared, we've lived through worse.

12    Reply
- 
John Doe 2h ago 


The article is full of lies. Who believes this shit??? I mean, you have to be a low-IQ individual, or a member of the political party to believe it.

36    Reply
- 
Jordan 2h ago 





Hell no! I will drive till I die, and I sure as heck will travel by plane

23    Reply
- 

50%
Negative
thread


Jenny 2h ago

Me too! I don't want to change my lifestyle because of this

24    Reply
- 
Riley 2h ago

Now this is what I call journalism, great job!




36    Reply

Figure 13: Philosophy part 1/2. The following freely available resources were used in making the designs: Semantic UI kit. Icons: Font Awesome, Ionic and Feather.

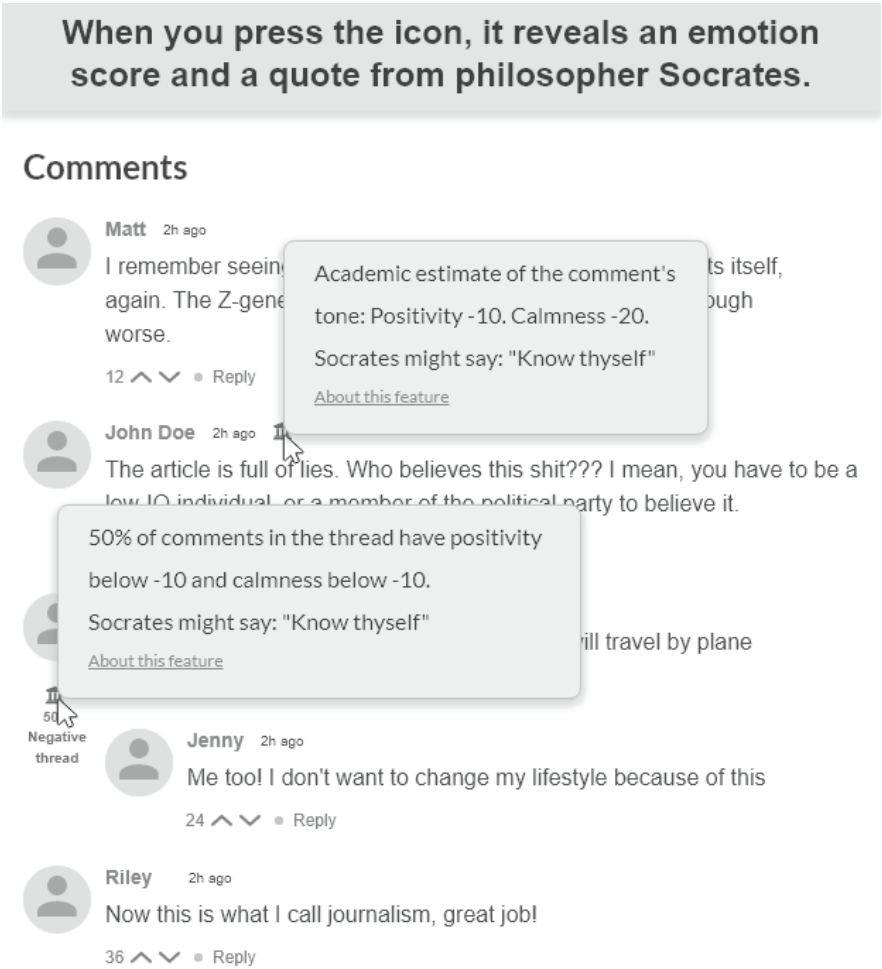


Figure 14: Philosophy part 2/2. The following freely available resources were used in making the designs: Semantic UI kit. Icons: Font Awesome, Ionic and Feather.

You, Anon123 are reading a news article online

Breaking News: Political Division all Time High

Political polarization – the vast and growing gap between liberals and conservatives, Democrats and Republicans – is a defining feature of American politics in 2029. 46% of U.S. citizens, almost all of them Republican, say the president did something wrong regarding the Gulf of Mexico oil spill and that it was enough to justify her removal from office. Another 28% of U.S. citizens say the president did something wrong but that it was not enough to warrant her removal, while 25% say she did nothing wrong.

..and then you post an angry comment on it.

The article is obviously full of lies. Only a low-IQ individual, or a party voter can believe this crap.



Figure 15: Regret part 1/5. The following freely available resources were used in making the designs: Semantic UI kit. Icons: Font Awesome, Ionic and Feather.

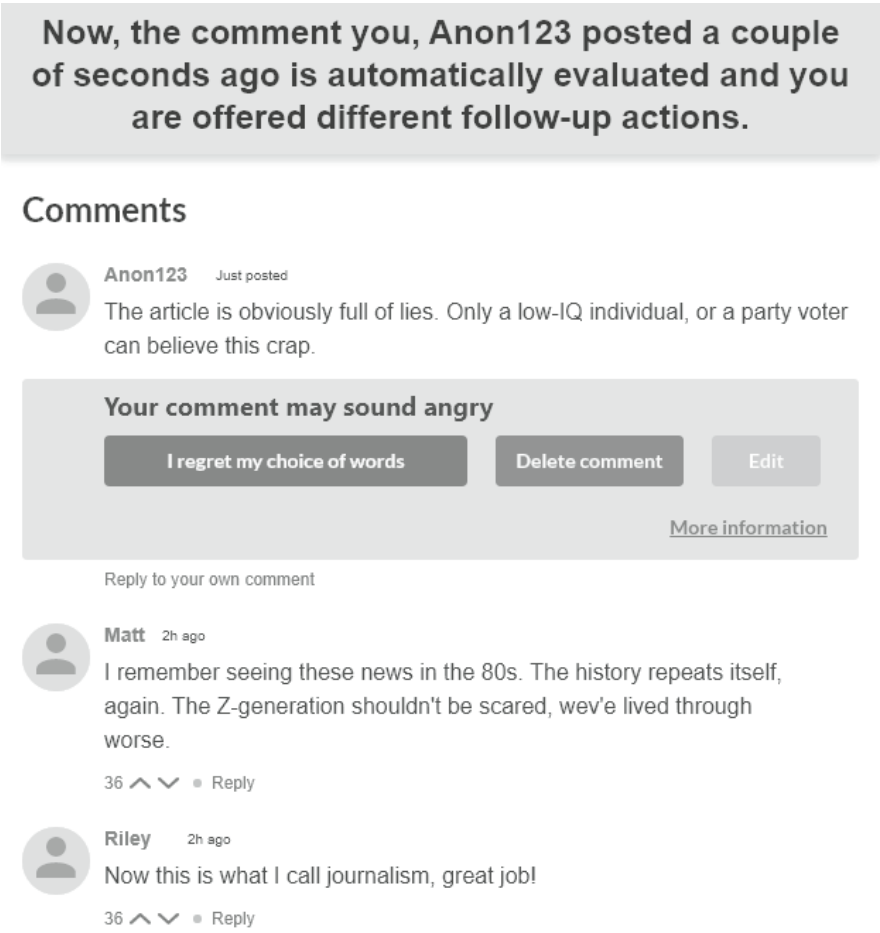


Figure 16: Regret part 2/5. The following freely available resources were used in making the designs: Semantic UI kit. Icons: Font Awesome, Ionic and Feather.

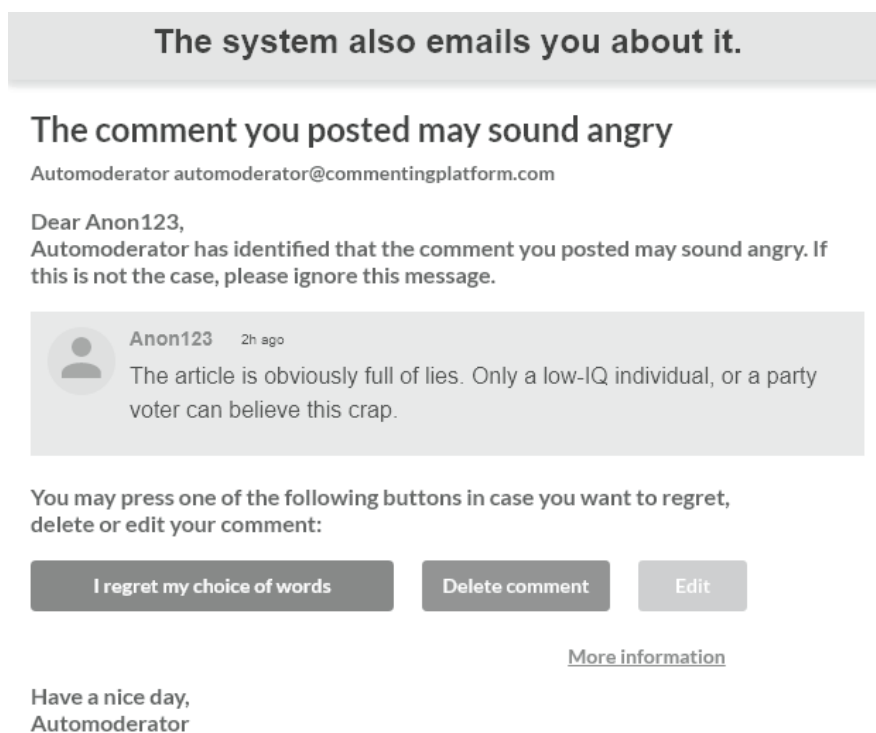


Figure 17: Regret part 3/5. The following freely available resources were used in making the designs: Semantic UI kit. Icons: Font Awesome, Ionic and Feather.

After thinking about it, you decide to use the Regret feature. A special label is added to your comment.

Comments



Anon123 2h ago

The article is obviously full of lies. Only a low-IQ individual, or a party voter can believe this crap.

Anon123 regretted their angry words

About this feature

Reply to your own comment



Matt 2h ago

I remember seeing these news in the 80s. The history repeats itself, again. The Z-generation shouldn't be scared, wev'e lived through worse.

36 ^ v • Reply



Riley 2h ago

Now this is what I call journalism, great job!


36 ^ v • Reply

Figure 18: Regret part 4/5. The following freely available resources were used in making the designs: Semantic UI kit. Icons: Font Awesome, Ionic and Feather.

Also, in case another user writes a reply to your comment, they are reminded that you regretted your words.

Reply to Anon123

Anon123 regretted their angry words

It's 

Post

Figure 19: Regret part 5/5. The following freely available resources were used in making the designs: Semantic UI kit. Icons: Font Awesome, Ionic and Feather.



Figure 20: Audience part 1/2. The following freely available resources were used in making the designs: Semantic UI kit. Icons: Font Awesome, Ionic and Feather.

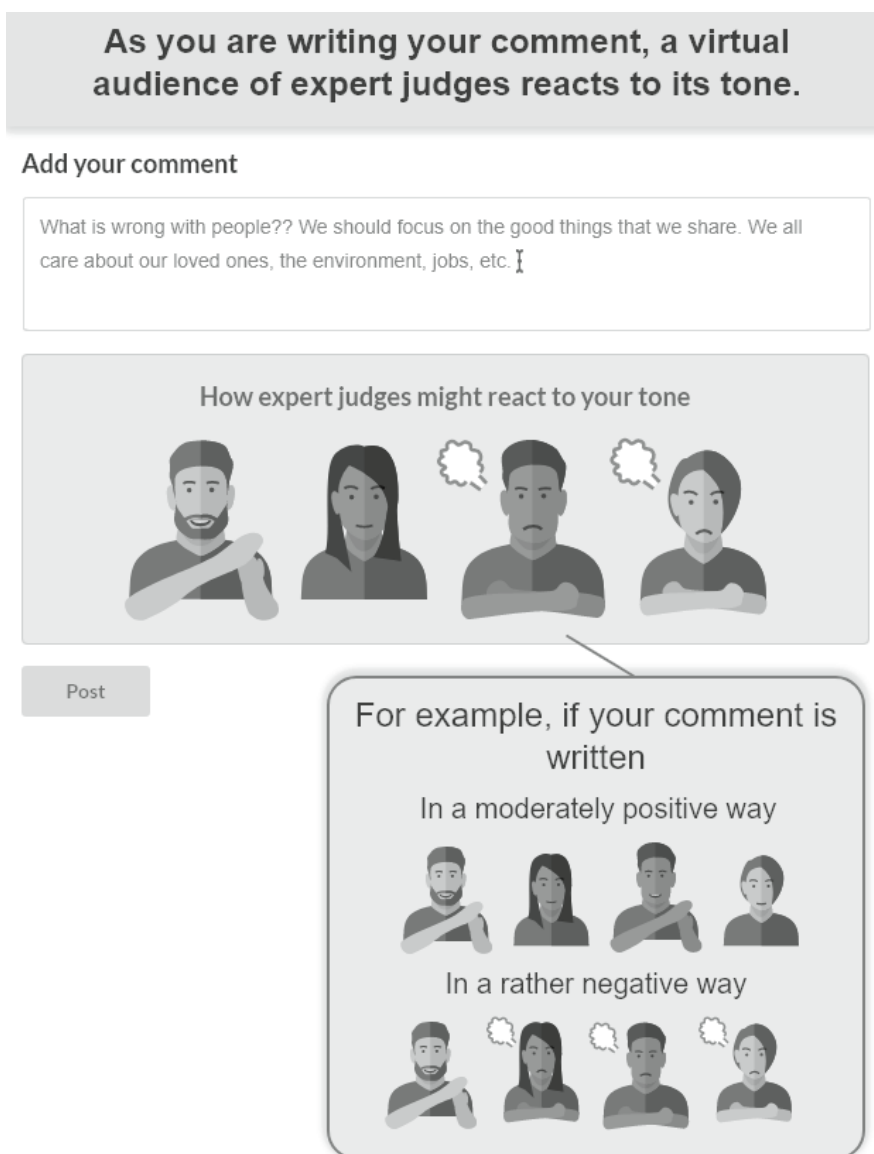


Figure 21: Audience part 2/2. The following freely available resources were used in making the designs: Semantic UI kit. Icons: Font Awesome, Ionic and Feather.

You are about to read the comments to an interesting but divisive news article. You are shown a notification about the argumentation that the comments include.

Breaking News: Political Division all Time High

Political polarization – the vast and growing gap between liberals and conservatives, Republicans and Democrats – is a defining feature of American politics in 2029. 46% of U.S. citizens, almost all of them Republican, say the president did something wrong regarding the Gulf of Mexico oil spill and that it was enough to justify her removal from office. Another 28% of U.S. citizens say the president did something wrong but that it was not enough to warrant her removal, while 25% say she did nothing wrong.



The discussion around this article contains



10% Hatefulness 5% Provocation 5% Encouragement 5% Agreement

Comments



Matt 2h ago

I remember seeing these news in the 80s. The history repeats itself, again. The Z-generation shouldn't be scared, we've lived through worse.

Figure 22: Warning. The following freely available resources were used in making the designs: Semantic UI kit. Icons: Font Awesome, Ionic and Feather.

PUBLICATION

III

User-centred quality of UI interventions aiming to influence online news commenting behaviour

Kiskola, J., Olsson, T., Rantasila, A., Syrjämäki, A. H., Ilves, M., Isokoski, P., & Surakka, V.

Behaviour & Information Technology, 1-33. 2022
10.1080/0144929X.2022.2108723

**Publication is licensed under a Creative Commons Attribution 4.0
International License CC-BY-NC-ND**



User-centred quality of UI interventions aiming to influence online news commenting behaviour

Joel Kiskola, Thomas Olsson, Anna Rantasila, Aleks H. Syrjämäki, Mirja Ilves, Poika Isokoski & Veikko Surakka

To cite this article: Joel Kiskola, Thomas Olsson, Anna Rantasila, Aleks H. Syrjämäki, Mirja Ilves, Poika Isokoski & Veikko Surakka (2022): User-centred quality of UI interventions aiming to influence online news commenting behaviour, Behaviour & Information Technology, DOI: [10.1080/0144929X.2022.2108723](https://doi.org/10.1080/0144929X.2022.2108723)

To link to this article: <https://doi.org/10.1080/0144929X.2022.2108723>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 10 Aug 2022.



Submit your article to this journal [↗](#)



Article views: 1308



View related articles [↗](#)










View Crossmark data [↗](#)

RESEARCH ARTICLE



User-centred quality of UI interventions aiming to influence online news commenting behaviour

Joel Kiskola , Thomas Olsson , Anna Rantasila , Aleksi H. Syrjämäki , Mirja Ilves , Poika Isokoski  and Veikko Surakka 

Faculty of Information Technology and Communication Sciences, Tampere University, Tampere, Finland

ABSTRACT

While HCI literature offers general frameworks for understanding user-centred quality, specific application areas may call for more detailed contextualisation of it. This paper focuses on socio-technical context of online news commenting by investigating speculative UI interventions intended to influence users' emotions and social behaviour. To understand the aspects of quality that matter to users in such UI interventions, we conducted an international online survey ($N=439$) and qualitatively analysed respondents' first impressions of eight different design proposals. The findings describe contextually relevant socio-technical viewpoints and offer actionable considerations for design. For example, the findings imply that designers should be mindful of possible unintentional misuse that may result from the UI reinforcing specific emotional states or affording stigmatisation of individual users. The study advances understanding of which aspects of quality should be considered when designing and deploying UI interventions for digital media services and evaluating them with potential end-users.

ARTICLE HISTORY

Received 19 December 2021
Accepted 28 July 2022

KEYWORDS

Design research; speculative design; design probe; Research through Design; UI intervention design

1. Introduction

Understanding perceptions of quality of user interfaces (UI) can be regarded as one of the core agendas in HCI. The breadth of aspects that are seen to influence perceived quality has expanded over time because of new theories, empirical knowledge, and the application of information technology in new areas. For example, the conceptual expansion from usability to user experience in the 2000s (e.g. [Diefenbach, Kolb, and Hassenzahl 2014]) introduced factors like pleasure and playfulness to be considered in the design and evaluation of IT systems. Following this trend, the increasing agency of IT systems and, for example, the recent discussion on the ethical aspects of IT (Shilton 2018) call for continuous revisiting the essence of perceived quality.

This paper analyses potential users' perceptions and articulations of quality of UI intervention designs that aim to influence users' emotions and behaviour in digital media discussions. We focus on speculative, low-fidelity designs and the specific activity of commenting on online news. While online news commenting has been studied with ethnographic and descriptive approaches (e.g. Diakopoulos and Naaman 2011), applying *research through design* in this area is rarer

with only few recent examples (e.g. Grön and Neli-markka 2020; Kiskola et al. 2021). The proposed intervention designs draw from prior research that suggest the use of 'nudging' mechanisms to influence user behaviour (Fogg 2009; Seering et al. 2019; Thaler and Sunstein 2009; Wang et al. 2014) and in mitigating online incivility (Taylor et al. 2019; Topal, Koyuturk, and Ozsoyoglu 2016). In particular, prior research suggests that increasing online news commenters' reflexivity and emotion regulation could be helpful (Kiskola et al. 2021; Topal, Koyuturk, and Ozsoyoglu 2016). The question of perceived quality becomes apparent as the designs propose to influence a delicate form of social activity where they could simultaneously be seen as both desirable and ethically questionable, depending on the perspective and criteria. Qualitative understanding is important due to the nuanced viewpoints that this application area introduces to nudging UIs.

Knowledge about what makes nudging UIs good in general (e.g. Bovens 2009; Desmet and Hekkert 2007; Fogg 2009; Tidwell, Brewer, and Valencia 2020; Galitz 2007) may not sufficiently inform the design of systems in the specific socio-technical application area of commenting behaviour. Online news commenting features

CONTACT Joel Kiskola  joel.kiskola@tuni.fi

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

complex social interactions, mediated by a relatively simple digital channel, which may manifest as undesirable phenomena, such as hate speech, intentional trolling, and inconsiderate commenting, which may develop into hateful discussion threads (Chen and Margaret Ng 2017; Cheng et al. 2017; Eberwein 2019). Users, journalists, and other stakeholders have varying views on if and how comment moderation should be implemented in this context (Kiskola et al. 2021; Stroud, Van Duyn, and Peacock 2016). Mindful of this complexity, we suggest that presenting potential users various speculative UI intervention designs and qualitatively analysing their opinions could increase understanding of potential expectations and requirements for such UI intervention designs. The present study focuses on the aspects of quality that potential end-users pay attention to in this context.

We conducted an international online survey ($N = 439$) in which each respondent evaluated two out of eight speculative UI intervention designs. While the survey featured multiple quantitative questions and items, in this article we focus on qualitative data from two viewpoints. First, we inductively analysed the commenters' first impressions of the designs. We think this offers insight into the aspects users may pay attention to and therefore need to be addressed when designing and deploying such interventions. Second, we inductively analysed respondents' explanations as to why they preferred one of two designs they had viewed. We applied a socio-cognitive lens in the qualitative analysis: we note the respondents choose to select some aspects of the reality and make them more prominent, so that certain problem definitions, causal interpretations, moral evaluations and/or outcomes are favoured and promoted (Entman 1993; Lin and Silva 2005; Orlikowski and Gash 1994). For example, we examined which problem the respondents think the intervention aims to solve. The approach was utilised to form a nuanced understanding of what good quality means to the users.

The contributions of this work are (1) Descriptions of relevant user requirements for UI intervention designs for enhancing discussions in digital media. This includes contextually relevant viewpoints on quality and critical perspectives to deployment of such technology; (2) Preliminary design guidelines for UI interventions in this context.

2. Theoretical background

The following offers a theoretical background by covering topics like speculative design, user expectations, nudging, and media studies. Situating the work in relation to literature on speculative design and user

expectations, Section 2.1 elaborates on the type of knowledge we sought with the analysis. Section 2.2 positions the work in the research on UI interventions and nudges, as well as outlines different ways to consider the quality of such systems. Section 2.3 focuses on the socio-technical context of online news commenting, shedding light on the communal level requirements for design and arguing for the need for social interaction design.

2.1. Using speculative designs to create knowledge on users' expectations

The current design proposals (Section 3.1) can be considered as speculative and discursive by nature (Tharp and Tharp 2019). Speculative design proposals can be useful research tools if they elicit informative reactions from study participants (Baumer, Blythe, and Tanenbaum 2020). This kind of knowledge creation follows the broad approach of research through design where design thinking, processes, and products are used as a method for inquiry (Bardzell, Bardzell, and Hansen 2015; Zimmerman and Forlizzi 2014). Epistemologically, speculative methods and designing speculative solutions can provide insight into social problems (Auger 2013; Baumer, Blythe, and Tanenbaum 2020). Provocative artefacts can be used to elicit users' values for the initial research phase of a project to design acceptable products (Johannessen, Keitsch, and Pettersen 2019). Hence, it can be useful to show people solution proposals that are not designed to be instantly adopted and that are framed as speculative. Further, discursive design aims to encourage critical thinking about design (e.g. about what values and behaviours design embodies), often with the intention of initiating subsequent debate (Tharp and Tharp 2019). Our designs feature this motive, in addition to the problem-solving motive.

The present work investigates the question of what quality means to the potential users in the context of speculative artefacts. Designers often see traditional measurement and evaluative techniques as inappropriate when developing new products that are not yet in existence (Suri 2002). Suri (2002) has argued that measurement, by its nature, forces designers to ignore all but a few selected variables. Hence, using well-established measures like AttrakDiff (Hassenzahl, Burmester, and Koller 2003), System usability Scale (Bangor, Kortum, and Miller 2008), or NASA Task Load Index (Hart 2016) would be misleading if the designers are not confident about which variables are relevant (Hart 2016). At the same time, we acknowledge that users may not accurately recognise their needs or wishes as regarding speculative products (Heikkinen, Olsson,

and Väänänen-Vainio-Mattila 2009; Yogasara et al. 2011; Orlikowski and Gash 1994).

In other words, this study seeks to understand the users' *assumptions* and *expectations* that affect the acceptance and adoption of technology (Orlikowski and Gash 1994). We inductively analyse respondents' reactions to the artefacts, and their argumentation regarding the quality of the artefacts. Hence, the work is related to studies of anticipated user experience (AUX), while, at the same time it is more speculative and explorative and less about measuring than typical studies of AUX (Olsson et al. 2013; Sánchez-Adame, Urquiza-Yllescas, and Mendoza 2020; Yogasara et al. 2011). We follow a similar approach as Bonino and Corno (Bonino and Corno 2011) who explored user expectations of smart homes of the future in Italy with a qualitative online survey. While the present study focuses on a different application area, the studies are analogous in that the participants know the environment where the technology would be implemented (home environment – online news commenting environment) and have expectations based on that knowledge. Like Bonino and Corno, we used an online survey to collect a broad sample of data and analysed answers to open-ended questions.

2.2. Conceptualising the motivational aspect of our design proposals

The literature features many ways to think about and name designs that aim to influence user behaviour: for example, *persuasive* design (Fogg 2009), *nudging* towards certain unconscious selections (Thaler and Sunstein 2009), and *friction* to hinder certain unwanted behaviours (Cox et al. 2016). In this broad conceptual landscape, there seem to be various interpretations about the terms – for example, what counts as a 'nudge' (Caraban et al. 2019; Zimmermann and Renaud 2021). To avoid terminological clashes and misunderstanding, the designs in the present work are simply termed *UI interventions*. To elaborate, we believe the term 'nudge' would take too strong a stance on the strength and pervasiveness of the intervention at this stage of design and in this design context. For example, an often-cited example of a 'nudge' is a traffic sign displaying a sad or a happy face depending on whether the driver obeys the speed limit (Weinmann, Schneider, and vom Brocke 2016; Zimmermann and Renaud 2021). In our design context, online news commenting, the user knows there are moderators (c.f. police) behind the intervention. As the user is not isolated from other people, the user might feel more than merely 'nudged'. Further, our designs can be perceived to intervene in

naturalistic behaviour in digital media, which is another reason for calling them interventions.

Good persuasive designs match the user's level of motivation and ability to act (Caraban et al. 2019; Fogg 2009) and are transparent (Bovens 2009). They support the user in acting in accordance with their overall preference structure (i.e. with their conception of the right thing to do in a given situation) (Bovens 2009; Sunstein 2018). However, should the user consider what interventions the *other* users might need, they might accept an intervention that is excessive compared to their personal needs. This further motivates us to study what potential users think about intervening in online commenting behaviour.

Little is known about users' perspectives on intervention designs utilised on online forums, news commenting platforms, and social media in general. While there is knowledge on how users perceive various content moderation strategies across various platforms (Cook, Patel, and Wohn 2021), it does not focus on the UI designs. A few recent studies investigate perceived benefits and drawbacks of guiding social media users or news commenters to stop and think before posting (de Carvalho, Olsson, and Kiskola 2021; Wang et al. 2014). Linhares de Carvalho et al. (2021) interviewed 18 university students about their perceptions of four proposed UI mechanisms for guiding users to emotional self-reflection when reading and commenting news articles online. The interviewees commented about the ease of use, usability, usefulness; feeling of control, censorship, intrusion; an unintended consequence of angering users; and level of trust towards the service. The study concluded that users do not want an intervention to interfere with fast-paced interaction in online news commenting. A study by Wang et al. (2014) featured two 'privacy nudges' to Facebook posting. Twenty-eight participants installed them as web-browser add-ons and used them for six weeks. The researchers discussed several perspectives to user-centred quality in the nudges: intrusiveness of the nudge; a sense of being watched or judged; control or customisation of the nudge by users; and usability and reliability of the nudge. The study concluded that 'privacy nudges' have great potential to assist users in avoiding unintended disclosures. Inspired by these two studies, we aim to explore user expectations and perceptions with larger samples of participants and with a larger number of design proposals.

2.3. Improving online news commenting as a socio-technical and systemic design problem

Online news commenting can be considered a socio-technical system (STS) where people communicate

with others through technology and their behaviour emerges rather than is dependent on technology alone (Whitworth 2009). For example, individual user's commenting behaviour depends on other users' earlier comments, the semantics and emotional associations related to the news article, their attitudes towards the topic, and the interaction affordances and conditions introduced by the discussion platform. This makes it difficult to predict how even a simple new design would be appropriated.

To further illustrate the complex nature of the problem, we apply the web of system performance model proposed by Whitworth & Zaic (Whitworth 2009; Whitworth and Zaic 2003), which has been used in information systems evaluation (Isaias and Tomayess 2015). Following the model, at the level of software, increasing the rule-based functioning of an intervention to commenting can decrease its ability to respond to environmental changes, and vice versa. At the *human* level, increasing the intervention's predictability can decrease its flexibility and vice versa. At the corresponding *communal* level, increasing the amount of order an intervention imposes on commenters can decrease their freedom. Other tensions the model proposes at the *communal* level are creation of benefit by social interaction (synergy) versus lack of social conflict (morale); respecting the right to be shielded (privacy) versus enabling everyone to easily see what is going on (transparency); and letting new people and ideas enter (openness) versus preventing ideological hijack (identity). To summarise, improving online news commenting can be difficult because it requires accounting for multiple charged perspectives to its quality and at multiple levels (e.g. communal, human, and software [Whitworth 2009]). A narrow focus on a single perspective or level can cause problems to pop up elsewhere (Alexander 1964; Whitworth 2009). For example, even a solution that seems to improve the quality of commenting without incurring any obvious costs might do so at the cost of human connectivity.

The socio-technical level and *communal* requirements are worth stressing as HCI and Design have long focused on the perspective of the individual (cf. *human* level [Whitworth 2009]). Designers should take responsibility as 'shapers' of society and not hide behind the needs and wishes of the consumer (Tromp, Hekkert, and Verbeek 2011). Further, existing conditions are often framed as problems and technological systems as solutions (Baumer and Silberman 2011), which is unhelpful when elimination of the problem is unlikely (Baumer and Silberman 2011) or the problem is socio-cultural. For example, as noted by (Sparrow, Gibbs, and Arnold 2021), 'the goal of completely

eradicating incivility is unfeasible and unreasonable'. Rather than imagining that a technology design offers solutions to extremely difficult problems, Baumer and Silberman (2011) suggest thinking of design as an intervention in a complex situation.

We found few articles with guidelines or principles for designing for online social behaviours, sociability, or social interaction. Of these, we want to mention Adrian Chan's 175-page explorative essay *Principles of Social Interaction Design* (2012). While the essay focuses mostly on social networking sites, it also offers general suggestions for social interaction design that seem applicable in this context. For example: anticipate the social practices that will emerge, consider who will be attracted to using the service, and who these users will attract in turn. Overall, according to Chan, good social interaction design accounts for the diversity of user experiences and for the development of a social tool over time.

3. Methodology

Following an explorative design process, we ran an international online survey to qualitatively analyse perceptions and opinions of people who at least occasionally comment on news on online news sites. The overall setup follows a common methodology where surveys are used to collect qualitative data with open-ended questions. Similar methodology has been applied, for example, in studies of user perceptions towards data disclosure for cognition-aware e-learning (Herbig, Schuck, and Krüger 2019), towards smart energy consumption metres (Jakobi et al. 2019), and towards augmented reality scenarios at early stages of technology development (Olsson et al. 2012). The use of an online survey allowed us both to invite viewpoints from a diverse sample of potential users and avoid the risks of real-world testing like failure to predict negative consequences (e.g. discouraging diverse discussion and supporting trolling) of intervention designs in the social context (Kiskola et al. 2021). The survey was implemented with LimeSurvey and the participants were recruited via Prolific, a platform for online research participant recruitment (Palan and Schitter 2018). To select a diverse sample of participants, we first conducted a short pre-survey regarding how often the candidate respondents read and commented on online news articles. The actual design survey asked participants about their behaviours and attitudes related to commenting on online news sites and invited them to evaluate two designs selected out of the eight design proposals. In this paper, we focus on the qualitative data from answers to two broad open-ended questions

as they were likely the best ones to reflect the respondents' ways of thinking.

3.1. Designs and scenarios in brief

The following summarises how the eight UI intervention designs were created and what the related scenarios of use are like, to elaborate what kind of artefacts the analysed perceptions of quality relate to. Only a summary is provided as *the designs are not intended as a novel contribution per se* in this paper. The designs are intended as propositions of possible future UIs, inviting the reader to assess their meaningfulness and speculate on the possible implications. The design work for this study builds upon our earlier research-through-design exploration (Kiskola et al. 2021), in which we envisioned unconventional solutions to the problem of uncivil commenting. In the study, we unpacked this same problem area and outlined critical perspectives on potential solutions by describing and analysing four designs that aimed to support emotion regulation by facilitating self-reflection. Next, we briefly recap the design process of the earlier study:

1. Existing design conventions were identified by analysing social media platforms and news websites. This was done to find a convention to be tweaked slightly, to avoid reinventing existing solutions, and to reflect on what kind of solutions might fit various news websites.
2. Approximately 60 concept ideas were sketched based on several idea generation sessions. Two general strategies mentioned in literature on critical design were used: (1) the designer picks a literary device (e.g. irony, sarcasm, parody, or ambiguity) and attempts to implement it in designs (Johannessen 2017) and (2) the designer picks a convention (cultural or UI) and tweaks it slightly, for example, by introducing a foreign concept, and then reflects on the result (Bardzell, Bardzell, and Stolterman 2014).
3. 19 of the sketched ideas were subjectively evaluated by the design team as more promising in terms of perceived criticality, novelty, feasibility, and effectiveness. Following this, the first author created UI mock-ups of the 19 ideas. Also, four of the 19 mock-ups were pictured and analysed in depth in the earlier study.

Eight of the ideas that represent a rich breadth of approaches to support self-reflection and emotion regulation in online discussion were chosen for the survey. The ideas were further developed and made more

presentable. The eight designs utilise several different 'emotion strategies' that Yoon et al. (2019) propose may be used in designs, such as suppression and avoidance. Also, the interventions are proposed to take place at different moments of use: before reading comments, while reading comments, while writing a comment, and/or after sending a comment. In addition, we subjectively assessed the designs as conceptually different from one another.

The names of the designs are EVALUATE, CREATURE, HIGHLIGHT, SYMBOLS, AUDIENCE, REGRET, PHILOSOPHY, and WARNING. For a full illustration of the designs and scenarios, see Appendix 1.

In the EVALUATE design (see Figure 1), the user must first indicate how they feel before they can add their comment. This is done by clicking a smiley face that represents their emotional state. The design aims to make comment writers more aware of their emotions. The design is inspired by and applies the theory of affect labelling (i.e., putting one's feelings into words) (Torre and Lieberman 2018).

To illustrate the design scenarios, the EVALUATE scenario was described to the respondents as follows: 'You are reading the comments to an interesting but divisive news article ... and wish to add your own comment'. (A couple of comments created by us are shown for illustration purposes). 'When you click "Comment", you first need to tell how you feel before adding your comment'.

In the CREATURE design (see Figure 1), an animated pet dog reacts to the emotional tone of a comment while the user is writing the comment. The benefits of using emotional attachment to pets to motivate behaviour change have been documented in previous research (e.g. Dillahun et al. 2008; Lin et al. 2006). The pet dog is displayed below the text area, and it is described as 'our digital friend'. If the user writes positively, the pet dog appears happy, as if ready to play. If the user is writing neutrally, the pet dog appears neutral (see Figure 1 top right). If the user is writing negatively, the dog communicates submission or fear. The design aims to motivate comment writers to consider their tone by giving feedback about it.

In the HIGHLIGHT design (see Figure 1), the user is offered an option to view an automatic analysis of the emotions in the comments. Negative emotional expressions would be highlighted in red, and comments containing strong negative expressions would be marked with an alert symbol. The design aims to make users more aware of the emotional expressions and to take a more analytical approach to reading comments. This design is also inspired by the theory of affect labelling (Torre and Lieberman 2018).

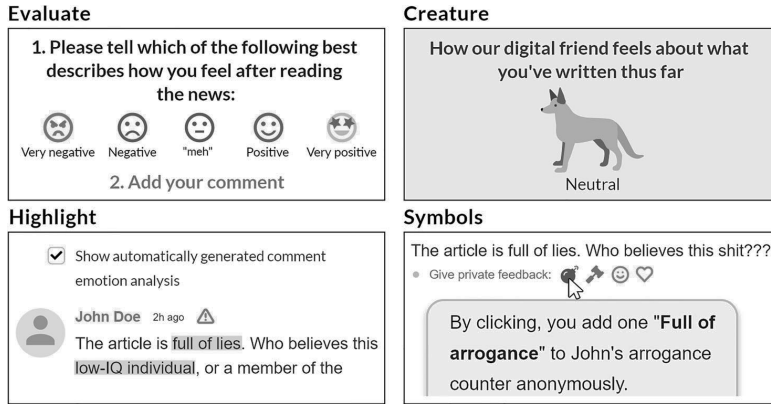


Figure 1. EVALUATE, CREATURE, HIGHLIGHT, and SYMBOLS designs in short.

In the SYMBOLS design (see Figure 1), the user is offered a way to provide anonymous, private feedback to any of the previous commenters. This is intended to decrease the likelihood of written personal attacks toward other commenters. It has been demonstrated that uncivil comments (including replies) promote further incivility (Chen and Lu 2017a; Ziegele et al. 2018), and that ad hominem attacks are a frequent type of incivility online (Coe, Kenski, and Rains 2014; Maia and Rezende 2016). In the design, there are buttons depicting a bomb, a gavel, a smiling face, and a heart next to every comment. The bomb symbolises 'Full of arrogance'; the gavel 'False claim/s'; the smiling face 'Well said'; and the heart 'Love it!'. Also, every user's profile contains a prominent section entitled 'Overview of the feedback from other users', which displays the same symbols and the number of times the user has received these feedback types. The design aims to

motivate comment writers to consider the quality of their writing and to guide the other users away from writing uncivil replies.

In the AUDIENCE design (see Figure 2), when a user is writing their comment, a virtual audience of expert judges reacts to its tone in real-time and their reaction is displayed below the text area. If the user writes in a moderately positive way, some members of the audience appear glad, and others have a neutral expression. If the user writes in a rather negative way, most members of the audience appear angry or frustrated. The design aims to motivate comment writers to consider their tone and who they are writing for. The audience's appearance in the proposal is also intended to communicate that the audience is ethnically diverse. Previous research has found that showing Facebook users profile pictures of people who will see (c.f. judge) their posts can help some of them avoid regrettable



Figure 2. AUDIENCE, REGRET, PHILOSOPHY, and WARNING designs in short.

disclosures (Wang et al. 2013). Also, the AUDIENCE design utilises the concept of being watched to induce self-awareness (e.g. Bradley, Lawrence, and Ferguson 2018; Cañigüeral and Hamilton 2019). Previous research implies that designs that induce self-awareness might reduce abusive comments to news (Sohn, Chung, and Park 2019).

In the REGRET design (see Figure 2), users' comments are automatically evaluated immediately after posting. If a comment sounds very angry, the user is notified and offered various follow-up actions below the published comment and by email. The user is offered options to regret the choice of words, to delete the comment, or to edit it. If the user chooses the regret option, a notification is attached to the comment, stating 'username regretted their angry words'. The design aims to motivate commenters to reconsider the emotional quality of their comments and provides a new affordance to show regret. Previous research has found that postings with profanity or obscenity can be a cause of regret for Facebook users (Wang et al. 2011).

In the PHILOSOPHY design (see Figure 2), problematic comments and comment threads are marked with a university icon providing subtle affordance to view analysis of the comment. If the user presses the icon, a box with the emotion score for the comment or comment thread and a quote from Socrates, 'Know thyself!' (Xenophon et al. 1979) is revealed. The emotion score has two dimensions, positivity and calmness. The design aims to motivate comment writers to consider the emotional quality of their comments and to enable other users to skip reading comments or alternatively to analyse the comments' emotional qualities.

In the WARNING design (see Figure 2), a notification is shown above the comment section, indicating a description of the argumentation within the comment section (e.g. '10% Hatefulness'). The design aims to make users aware of emotions in comments, to use a more analytical reading approach, and to allow a choice whether they want to read the comments. The design is mainly inspired by the theory of affect labelling (Torre and Lieberman 2018).

3.2. Participants and recruitment

The pre-survey deployed in Prolific involved 2000 voluntary participants who met the specified eligibility criteria: fluency in English, normal or corrected-to-normal vision, and a minimum approval rate of 70% in Prolific (percentage of total submitted studies minus returned).

The key criteria for inviting the pre-survey participants to take the design survey included having given

complete answers and commenting at least occasionally on online news sites. Furthermore, because we wanted to focus on news sites that have commenting sections, respondents who had mentioned some of the following sites as their main news sites were not invited to take part in the design survey: Facebook, Twitter, Reddit, Quora, YouTube, various blogs, and news aggregators where we could not find comment sections. That is, only responders who mentioned news publishers' sites, sports news sites, gaming news sites, or alternative news sites were invited. Based on these inclusion and exclusion criteria, altogether 480 Prolific users were invited to take the design survey.

Next, we briefly describe how the respondents were introduced to the design survey and the main parts of the survey. The survey study in Prolific was entitled 'Survey on improving discussion around online news articles'. The study description stated that it asks about the behaviours and attitudes related to commenting news on online news sites. Furthermore, respondents were told that two UI proposals will be shown as speculations of how discussion around online news articles could possibly be improved or kept at a good level.

Of the 480 survey responses, 41 were discarded as incomplete (i.e. missing answers), or duplicates (i.e. the same person completing the survey twice), or click-throughs (i.e. response times two standard deviations below the mean, or nonsensical answers to open questions). Of the 439 respondents with valid responses, 45.3% reported being females and 54.7% males. The respondents' age range was 18–75 years (average 33.5 years, $SD = 11.98$). 43.3% of them were from the United Kingdom (UK), 12.1% from Poland, 10% from the United States (US), and the rest 34.6% from altogether 36 other countries. All respondents reported to comment on online news sites at least occasionally.

3.3. Survey procedure and questions

Out of the altogether eight speculative UI intervention proposals, each respondent was shown two pseudo-randomly selected designs. Pseudo-randomisation was used instead of true randomisation to ensure that all eight designs were presented an approximately equal number of times in the sample. The two designs were then presented to the respondent in a random order. The designs and the associated scenarios of use are described in Section 3.1. Immediately after presenting a design, the respondents filled in a mandatory open-ended question (analysed in this paper) and several other, mostly closed-ended questions

(not discussed in this paper). The open-ended question was, 'How would you describe your immediate reaction to this solution? How do you feel about it?' Furthermore, after they had evaluated both designs, another mandatory open-ended question was presented (analysed in this paper): 'Now, consider the two different solutions that you saw: X & Y. Which of them you found as the better solution for improving the commenting culture on online news? Why?' We focus on these two open-ended questions as the answers likely reflect the respondents' own way of thinking about the designs, which is what we are interested in this study.

3.4. Data analysis

We qualitatively analysed the responses to the two open-ended questions: first reactions to designs and explanations for the choice of the better design. The average number of characters in the responses were 175 (standard deviation 148) and 100 (st. dev. 91), respectively.

We followed a data-driven explorative analysis informed by the socio-cognitive analytical lens of technological frames (users' assumptions, expectations, and knowledge) (Lin and Silva 2005; Orlikowski and Gash 1994). It was kept in mind that people generally choose to emphasise some aspects of reality, so that certain problem definitions, causal interpretations, moral evaluations and/or outcomes are favoured and promoted (Entman 1993; Lin and Silva 2005; Orlikowski and Gash 1994). Open and axial coding was conducted to highlight themes from the data and to build a hierarchy of categories. This and comparable coding methods have previously been extensively utilised to understand user expectations of new technologies or applications (Jakobi et al. 2019; Nicholas et al. 2017; Olsson et al. 2012). The responses were read and coded one at a time (i.e. given short words or phrases that describe the meaning of the responses [Saldaña 2013]). When reading the comments and coding them, the coders paid particular attention to the following aspects of the responses: (a) how the responses described the designs, (b) how the respondents described their reactions to the designs, and (c) what kind of vocabulary was used in the responses (e.g. style, tone, length of the response).

The codes were then further abstracted into categories presented below in Section 4. The categories were generated by abstracting out existing codes and by developing new concepts that encompass several of them. When reasonable, lower-level categories were generated to describe respondents' assumptions and

expectations in more detail (e.g. a category of helpfulness could be elaborated by considering who the helpfulness is directed to and for what reason).

In the end, the number of answers matching each category was counted. The quantifications are meant to be inferred merely as indicative; we argue that the contribution of the results lies in the diversity and qualitative descriptions of the identified themes and categories rather than in the quantity of the responses per category. New viewpoints and nuances to quality, and critical perspectives to the deployment of technology are valuable as far as they are meaningful, regardless of how many respondents provide them.

As the questions were open and the answers varied, the first author, who was primarily responsible for creating the designs, collaborated with two researchers to classify and quantify the data. He coded the data using Microsoft Excel and created preliminary classifications, then met with the other researchers to refine the codes and categories. Additionally, the third author rated 50 randomly selected responses twice and the ratings were compared to those given by the first author. While we had some disagreements about the ratings, they primarily resulted from ambiguity of the answers. Overall, we engaged in a highly iterative process where the individual codes and their interrelations were gradually clarified. However, analysis of inter-rater reliability was not seen useful as *the findings do not hinge on frequency counts*. The participant quotes presented in the paper are verbatim except for some corrected typos.

4. Findings

The following reports selected findings on the respondents' impressions of the design proposals. While the survey data features diverse perspectives, we focus on categories we found qualitatively most interesting and specific to the context of online news commenting, hence offering nuanced perspectives to user-centred quality in this area. In other words, the presented categories are not necessarily the most frequently identified in the data. In total, 274 first impressions and 285 arguments for the choice of the better design fell into at least one of the code categories presented in the following sections. The other categories of the first impressions (omitted from this report) contain, for example, short emotional reactions (e.g. 'angered', 'I think it's great!') and general comments about good or bad style of design. Those for the choice of the better design contain, for example, expressions of uncertainty, considerations of the ease of use or familiarity, and vague or unclear answers. Each subsection heading represents a relevant

high-level category that is represented by several categories identified in the analysis (*marked with cursive typeface*).

4.1. Respecting freedom to comment

Many respondents appeared to feel that some of the designs would restrict users' freedom. In their first impressions of the designs, 32 respondents appeared to *refer vaguely to restriction of their freedom*. In the choice task, 46 respondents found the chosen design better because it was vaguely perceived to *restrict their freedom less* than the other design. For example, in their first impression, a respondent wrote that CREATURE is vaguely restrictive:

It seems quite patronizing. I don't see why comments should be moderated to be positive. (From UK, comments monthly on news sites)

Some respondents appeared to think that a design can represent censorship. In their first impressions of the designs, 14 *clearly referred to censorship* (i.e. intentional suppression of speech). In the choice task, 28 specifically stated the chosen design *leaves more space for free expression*. For example, a respondent had the following first impression:

On the AUDIENCE: 'People should be allowed to express their opinion regardless of what it is. Failing to do so is asserting some sort of control on people's opinion and will be pathetic for democracy'. (US, comments monthly)

However, to illustrate how difficult it is to judge what is and is not restrictive at this stage of design, we quote two opposing arguments from the choice task. A respondent argued the PHILOSOPHY to be better than the AUDIENCE because they perceived that PHILOSOPHY is neutral:

Less offensive. The second solution [AUDIENCE] appears like you are being judged and tried by 4 other people. The icon solution is a neutral symbol. (UK, comments monthly)

In contrast, another respondent argued the opposite: that the AUDIENCE is better than the PHILOSOPHY because the latter represents censorship:

I think it is better to get people to think before they comment than it is to censor it after it has appeared. (US, rarely comments)

In other words, the answers in this category imply that high quality is marked by a capability of the intervention design to manage a balance between restriction of speech and promotion of civil discussion. Furthermore, we argue that emphasising free speech might be contrary to the wishes of those users who want more moderation.

4.2. Objectivity in assessing comments

In their first impressions, five respondents appeared to ponder the question of who decides whether a comment is problematic. The respondents appeared to indicate that they are doubtful that the proposed evaluators (other users or an algorithm) would evaluate the comments objectively. For example, a respondent asked the following in their first impression of the REGRET:

Who on the newspaper is the arbiter of what constitutes anger, and when it is justified? (UK, comments weekly)

In the choice task, 13 respondents said the chosen design is better because of the trustworthiness of the actor/s who evaluate the text. For example, a respondent argued that the REGRET is better than SYMBOLS because they perceive machines as incapable of giving biased feedback:

It analyses wording of the comment not the meaning of it. In the 'Feedback' solution it's the people who decide what feedback should they give, and they can give you a bad one just because they disagree with you on the topic - not because of your wording. (Poland, comments daily)

In sum, this implies that high quality is indicated by having actors that the users can trust as those who judge and moderate the comments. While the responses offered little guidance on what would increase trust, they implied a need for evidence that the intervention appears objective in assessing the tone of the comments.

4.3. Helping various users behave better

Most of the respondents who appeared to perceive uncivil commenting as a problem, appeared to think that the interventions are meant to help prevent users from accidentally or unintentionally behaving in a way that can come across as uncivil. More broadly, they thought that news commenting can lead to emotionally stressful situations. In their first impressions, 94 respondents appeared to say that the design would help to improve the quality of commenting of nonspecific users (e.g. respondent referring to 'the writer'). Also, 51 indicated that they personally have challenges and need help. A few respondents indicated that they are *not* the ones needing help. 12 respondents argued that a specific user group would need help (e.g. users who are easily 'triggered' or what they referred to as 'troublemakers'). In contrast, 35 respondents argued that the design would not stop an irritated user, and some argued it would not stop someone who irritates other users on purpose (9).

To illustrate the point of helping the user to avoid accidentally or unintentionally behaving in a way that

can come across as uncivil, a respondent had the following first impression of the REGRET:

The solution is actually great. It gives me the opportunity to think about the consequences of my choice of words and be able to make the necessary corrections. (UK, comments daily)

Also, to illustrate the closely related point of helping the user to avoid getting into emotionally stressful situations, a respondent had the following first impression of the SYMBOLS:

I think it's a good solution to show that I do not agree with this and not to enter into unnecessary discussions with the author. (Poland, comments daily)

The respondents' arguments in the choice task coincide with the first impressions. In the choice task, 169 respondents appeared to argue that the chosen design is better because it is *more effective or helpful*. As an example of this helpfulness argument, a respondent argued that the CREATURE is better than the HIGHLIGHT:

Animated creature might be a big help for people to not be misunderstood by using wrong choice of words. It can also make comment section more civilized where people instead of swears might use more cultural way to express their opinion or critics. (Poland, comments monthly)

In contrast, 21 respondents appeared to argue that the chosen design is *more effective as it is more forceful or restrictive*. For example, a respondent speculated that the AUDIENCE would be more effective than the WARNING:

Easier to appreciate, the visual effect is more shocking and therefore will be more effective. The arbitrary percentages of the latter just seemed too random. (UK, comments daily)

Summarizing these categories, high quality seems to be indicated by the design helping the users to avoid getting themselves into emotionally stressful situations. Still, for some users, good quality means that an intervention design also must be able to deal with those who intend to be offensive. Further, the fact that 94 respondents referred to nonspecific users could be an instance of the 'third-person effect' (Phillips Davison 1983): many believe that they personally would not be influenced, while other people would.

4.4. Use of apt metaphors

In their first reaction, 15 respondents appeared to associate the metaphors and manifestations of mimicry (i.e. copying properties of familiar objects, organisms, or environments) in the proposed designs to their

usefulness. The answers to the choice task also suggest that users may prefer metaphors and mimicry that matches their personal taste, values, and contextual expectations. 15 respondents argued that the chosen design is better because it fits the serious use context better, and 9 argued that it is better because of being more playful. However, as these two criteria could be considered contradictory, they reflect the variety of possible tastes that people can have. For example, the following two responses take contrary views on the CREATURE:

It is a very creative and worthy solution, almost everyone feels empathy with dogs so it might be effective. I feel empathy towards the dog, so I'd change my comment if it were sad. (Portugal, comments weekly)

This is not relevant to the posting of comments. I feel it downplays the issue of what impact your comments have and is almost more suited to children rather than adults. (UK, comments monthly)

The first response above seems to interpret the sadness of the pet dog as a metaphor for human suffering, while the second seems to interpret it more literally as a pet dog.

Further, to illustrate how mimicry in a design was connected to usefulness by some of the respondents, a respondent commented on the AUDIENCE:

It makes judgement more human; it seems people are closer to me; I can understand their feelings better. (Italy, comments monthly)

To this end, high quality and usefulness appear to be indicated by the applied metaphors and/or concepts matching their personal values. Creative use of metaphors might play an important role in supporting some users' reflective and empathetic thinking.

4.5. Avoiding risks of intentional misuse

In their first reactions to a design, 28 respondents spontaneously considered how the design could be purposefully misused to hurt other users. In the choice task, 16 respondents argued the chosen design is better entirely or partly because it is less open to misuse. This means that some respondents not only noted that the design could be intentionally misused, but also argued that it is important that intentional misuse is actively discouraged or prevented. This general category of expected intentional misuse comprised three more specific perspectives explained below:

4.5.1. Some users would seek to receive negative scores

This perspective illustrates a downside to giving negative feedback to uncivil commenters: it may encourage

further incivility, for example, due to a sense of being provoked or a will to explore the boundaries of the scoring system. This was expected of every design where the system was proposed to explicitly evaluate or grade the user's comment. For example:

On the AUDIENCE: 'in general I like the concept of it; however, it is open to interpretation depending upon the article - the article may generate a negative opinion which means people reply with a slightly negative attitude and it may only serve to encourage some people to carry on their comment further if they see it is generating a response that will gain replies by being overly negative'. (UK, comments daily)

4.5.2. The users would show that certain views are not welcome

Some respondents were concerned that an option to quickly give anonymous negative feedback to a commenter can be used to send the message that certain views are not welcome, for example:

On the SYMBOLS: 'I think it's a good idea, but people could give [the commenter a label of] full of arrogance only because they don't share their opinion'. (Ecuador, comments weekly)

On the SYMBOLS: 'Sometimes I comment on articles from other newspapers with very different views (e.g. Daily Mail), even though I know my comments will get downvoted, just to show them that some people think differently. But I would be a bit upset because I know that on that news site, my profile would get a bad rating, purely because my views differ from most of the readers'. (UK, comments weekly)

4.5.3. Bullies would target the users who stand out

This concern applied to the designs where individual comments are marked as different from others. For example:

On the REGRET: 'I feel like comment readers might start bullying those people who have a label of regret and create even harder conflict'. (Lithuania, rarely comments)

These categories imply that high quality is indicated by preventive actions (or assurance thereof) that minimise intentional misuse of the intervention. The design proposals featured indirect suggestions, and it became evident that the users might react to the suggestions in unintended ways. Many expected behaviours like 'gaming the system', which is extensively discussed in the literature (e.g. Petre, Duffy, and Hund 2019).

4.6. Avoiding risks of the intervention leading to unintended detrimental behaviour

This category involves unintended, unintentional uses of the designs. In their first reactions to a design, 14 of

the 439 respondents spontaneously considered how the design could be used in unproductive or harmful ways without an intent to do so. In the choice task, 5 respondents argued that the chosen design is better entirely or partly because it has less of a risk of unintended use. This category comprised five more specific perspectives, which we explain in what follows:

4.6.1. The user could be misdirected to aim for a positive analysis score for their text

This was expected of the designs that evaluate the comment while writing it. This also illustrates a downside to giving positive feedback to civil commenters: it may turn the receiving of positive feedback into a goal, which can distract the original activity of commenting on news. For example:

On the AUDIENCE: 'I'd be concerned that it would encourage me to write comments that make the virtual experts happy rather than helping me concentrate on what I'm thinking about the news issue'. (UK, comments daily)

4.6.2. Directing the user's focus on negativity

This expectation reveals a belief that online news commenting easily gravitates toward negativity. The expectation came up with the designs that propose to show to the readers whose comments might be problematic:

On the PHILOSOPHY: 'It highlights negative comments and hides the more positive ones. I found it unpleasant'. (UK, comments weekly)

On the HIGHLIGHT: 'I think this solution would be helpful but wouldn't fix the problem completely. It highlights uncivil comments what leads to us paying attention to them even more and as people tend to react to such strong feelings, it would probably cause even bigger fights because people would focus only on the negativity'. (Poland, comments weekly)

4.6.3. Individual users could be stigmatised over time

This concern applied to a scenario where the users give honest and accurate negative feedback to another user who is commenting in an uncivil way, and where the feedback stays on their profile for a long time. This may lead the other users to be overly judgmental toward the one with negative feedback in the future. For example:

On the SYMBOLS: 'I don't really like that. You might say something arrogant in one article and 500 people click your 'full of arrogance' and then there is no coming back from that, it will be like a stigma. If you comment next on another article, someone will see your profile

and judge you based on one number that may have come from one unpopular comment on another article that had nothing to do with the current article'. (Greece, comments monthly)

4.6.4. Directing the users to comment about the discussion platform rather than the news article

Particularly the designs with a provocative communication style were feared to cause this, for example:

On the PHILOSOPHY: 'I think this solution is not good. It seems self-indulgent to use Socrates. I don't think the wider public will understand the relevance of this and it won't have the desired effect. It is likely to generate negative comments about the system itself'. (UK, rarely comments)

4.6.5. Reinforcing the commenter's emotion

All the comments in this category were about the EVALUATE, where the user must click how they feel before writing a comment. The respondents were concerned that the increased awareness of the emotional state might make one more focused on it, hence reinforcing its negative aspects. For example:

I don't think it will work - may encourage people to feel more negative/angry by identifying the feeling. (UK, rarely comments)

In other words, the subcategories above imply that high quality would be indicated by explicit features and/or assurance that unproductive and unintended use of the intervention would be prevented.

5 Discussion

In the following, we discuss the meaning of the identified categories of quality at different levels. We propose preliminary design considerations, many of which introduce the needs for balancing acts between different extremes. The considerations are meant to help creating high-quality UI solutions and appropriately communicating them to users. Finally, we reflect on the validity of the reported study.

5.1 Design considerations per category

5.1.1 Respecting freedom to comment

Considering Whitworth's (2009) STS theory and its *communal* level, which concerns the exchange of norms, ideas and beliefs, people appear to cherish freedom and active audience participation in journalistic context. At the *human* level, which concerns personal level exchanges of meaning, the users seem to want the design to remain unnoticed, yet act when needed,

in order to allow for appropriate communication between news readers. This aligns with the UI design principles of supporting immersion and compatibility with the user's perspective (Galitz 2007). This requirement is also supported by related work of Wang et al. (2014) who found that a 'privacy nudge' that delays posting on Facebook can both prevent unwanted disclosures and feel intrusive.

Design consideration 1: Seek for a balance between restriction of speech and promotion of civil discussion.

The design could be made feel less restrictive, by letting the user have some degree of control over the intervention design, making the system at least a little bit flexible. For example, we speculate that more users could be satisfied if there were easily accessible settings to influence how often the user is likely to see the intervention. That said, the impact of this kind of customisability on the effectiveness of the intervention ought to be studied case by case.

5.1.2. Objectivity of intervention

Considering the *communal* level (Whitworth 2009), the users seemingly require the design to be in line with the protection of commenting as a place where different opinions are allowed. At the *human* level (Whitworth 2009), the users seemingly have a broad requirement of untampered communication. Also, previous research stresses the requirement for objective moderation (e.g. Wang 2021).

We argue objectivity to be important when considering contexts where people of differing opinions take part in commenting. Objectivity is also important in contexts where the users could perceive the discussion platform provider to have an interest in promoting certain types of opinions. In such contexts, the users probably need to know that the system was intended to avoid any bias.

Design consideration 2: Offer reasons for the users to trust that the comments are evaluated by objective actors.

For users who perceive that the intervention is somehow biased or wrongful towards their commenting, it is central to offer ways for them to defend themselves. For example, a new UI proposal could feature a possibility to directly chat with administrators or moderators in problematic situations.

5.1.3. Helping various users behave better

Some respondents seemed to want the designs to target users who are clearly trolling. However, most wanted that the average user is helped by an intervention. The call for help seems to illustrate, at the *communal* level (Whitworth 2009), that many users think the social interaction (synergy) in commenting should result in more benefits, such as production of information,

enjoyment, and understanding. At the *human* level (Whitworth 2009), the users' need for help suggests that many users think the current commenting systems do not afford enough capability to control one's tone or to empathise with other users when communicating. This is also supported by literature: the current, largely text-based interfaces may limit the ability to control one's emotions or to empathise with other people (Walther 1993). Also, previous research has found that some social media users would like to get help in controlling their tone of writing (Wang et al. 2014).

Unfortunately, our data does not indicate *how much* help the system should give, in what contexts, and to whom exactly. On one hand, helping when it is not needed could feel patronising. On the other hand, the more the design feels like an intelligent assistant, the higher the risk of 'infantilisation': individuals may come to rely on the guiding interventions and become unable to make decisions on their own (Acquisti et al. 2017; Bovens 2009).

Design consideration 3: Seek for a balance between helping the users too much and helping the users too little.

Design consideration 4: Help the user to avoid getting involved into emotionally upsetting situations.

The designs could be explicitly communicated as attempts to improve social interaction as this could increase the likelihood that the user accepts the design. In the light of the designs presented in this paper, it might be wiser to imply that the users lack the ability to control their tone of writing rather than a motivation to control it (Fogg 2009).

5.1.4 Use of apt metaphors

Considering the *communal* level (Whitworth 2009), the findings suggest that the style of addressing the commenters should match the commenters' values and contextual expectations. For example, if commenting is considered a serious matter, playful metaphors may be a bad idea. At the same time, at the *human* level (Whitworth 2009), the findings suggest the design should match user's personal requirements. This seems to call for personalising or customising the design. However, we do not have strong reasons to believe that the users would creatively customise a UI intervention design's appearance. Also, we speculate that a high degree of personalisation of a UI intervention (e.g. highly personalised metaphors) would scare off a large portion of users.

Design consideration 5: Utilise metaphors with caution.

We emphasise the need to try different metaphors (e.g. dog vs. cat vs. abstract creature) as well as basing them on knowledge of the cultural meanings in the

target culture. In a great product metaphor, the metaphor's source has high salience (i.e. significance in a person's representation of a 'category') (Cila, Hekkert, and Visch 2014; Ortony et al. 1985). For example, reflecting on our design choices in the *CREATURE*, a pet dog appearing fearful is not a typical exemplar of the concept of suffering, therefore its salience might not be high. In addition, in a great product metaphor, the 'source' (e.g. a tornado) should have obvious similarity with the 'target' (e.g. a vacuum cleaner) (Cila, Hekkert, and Visch 2014). As the connection between a fearful dog and a negative comment is arguably not that obvious in the *CREATURE*, it could be seen as a decent metaphor, but not a great one.

5.1.5 Avoiding risks of intentional misuse and unintended detrimental behaviour

Considering the *communal* level (Whitworth 2009), intentional misuse of commenting UI can be seen to create strong conflicts and exhaust users' morale. The same is true for the other detrimental behaviours that the respondents mentioned, though their effect might be less drastic. At the *human* level (Whitworth 2009), the unproductive behaviours can harm the perceived ease of use of commenting or one's capability to comment. Previous research indicates that many people avoid commenting because of conflict in comments (Stroud, Van Duyn, and Peacock 2016). Further, we note an earlier work has found that some journalists expect that some users would use automatic notifications about uncivil writing as a guide to write uncivil comments (Kiskola et al. 2021). The expectation of intentional misuse did not, however, come up in an earlier work where 18 university students were interviewed (de Carvalho, Olsson, and Kiskola 2021).

Design consideration 6: To discourage creative misuse, make the design harder to use for unintended purposes.

Design consideration 7: Analyse which UI affordances might encourage detrimental behaviour and try to avoid including them in the design.

When designing future UI interventions for social contexts, it could be a beneficial exercise to anticipate and model intended use processes, and then identify unintended forms of use. For example, typical and atypical deviations, and completely aberrant behaviours could be identified, and considered from the perspectives of natural, accidental, and intentional evil (Klein 2007; Merton 1936; Nelson and Stolterman 2012; Van Der Vegte et al. 2004). Also, some crude user personas (e.g. a worrier, a hedonist, a controversialist, or an inconsiderate person) could support the analysis. Moreover, besides this design work, it could be wise to vaguely communicate readiness to address unintended

behaviours to potential users. This could help potential users accept the technology despite seeing flaws in it.

5.2. Considering quality at multiple levels in design and evaluation

All in all, the analysis implies that the behavioural issues related to uncivil commenting are largely socio-technical by nature. Rather than being caused by either technology or behavioural conventions alone, the issues emerge from the application of technological solutions in complex and socially constructed circumstances (Whitworth 2009). For example, good quality is not only unambiguously linked to the artefact's qualities but also to a belief that everything people might do with the artefact has been considered. This implies that quality also refers to addressing various particularities of the intended socio-technical-cultural context. This idea is strongly in line with Chan's (Chan 2012) normative notion that good social interaction design accounts for the development of a social tool over time.

Next, we reflect the identified user perspectives on quality against common notions of user-centred quality in HCI. The respondents appeared to often evaluate the designs from the perspective of the community or society (e.g. 'It can also make comment section more civilized ...', '... pathetic for democracy ...'). Hence, in this context, the concept of user-centred quality also covers *communal* requirements, such as freedom, order, morale, and synergy, as highlighted by Whitworth (2009).

Many qualities commonly focused on in UI design (e.g. ease of use, clarity, desirability) can support *communal* requirements in this context by, for example, making it easier to comment, understand other users, trust other users, and follow the predefined community rules. However, particularly the *adaptability* (cf. reliability, [Whitworth 2009]) of the design seems relevant: a person may consider the trouble it would take for a user or a news site to use a design for unintended or unadvertised purposes (adaptability). Hence, perception of adaptability is related to both fears that users will misuse the design and fears that a news site will use the design to censor and manipulate users. While a low cost of adaption does not guarantee use for unfruitful and malicious purposes, a high cost of adaption makes such use impractical.

Reflecting on the prevalence of the expectation of misuse, we found it surprising that as many as approximately ten percent of the respondents raised the possibility of intentional misuse and other behaviours that can cause harm. Perhaps this is connected to a wider social context of online incivility and the public debate about it (Diakopoulos and Naaman 2011; Gillespie 2018). The topic of

online incivility has been debated for about a decade (Gillespie 2018; Grön and Nelimarkka 2020), and especially the most actively commenting respondents likely have first-hand experience on it.

From the perspective of design evaluation, the findings can be seen to support the premise that traditional, unavoidably reductionist measurement instruments like specific user experience questionnaires might indeed disregard relevant qualities of UI interventions in this area. As argued by Suri (2002), traditional, reductionist measurement and evaluation techniques are often not helpful to understand how novel products would be perceived and experienced. As they require knowledge about what would be relevant to measure, many aspects of perceived quality will likely be missed.

Further, while high quality may be described using short quality attributes, for quality attributes to offer actionable guidance to design, the design context must be well known, and the attributes contextualised accordingly. For example, recognising that a good motivational intervention to online discussion is effective would leave much contextual nuance unspecified. Accordingly, in this study, rather than reducing the qualities into a list of adjectives, we offered longer qualitative descriptions.

5.3. Reflection on the research process and methodology

Considering the methodological approach, the use of Prolific in recruiting participants for the survey resulted in over-representation of participants from the UK and other Western countries. Thus, the findings on how good quality is perceived represent mostly Western viewpoints. The socio-technical nature of the context area would benefit from data from, for example, more collectivistic cultures, and cultures that typically have different views of authority (see e.g. Baggini 2018). Further, we note that the monetary compensation for acceptable survey participation in Prolific might have caused the respondents to give longer answers to make sure their response gets accepted.

Regarding the extensiveness of the findings, they are based on online news commenters' opinions and arguments on eight speculative intervention designs focusing on the tone of commenting and emotional reflection and are therefore limited in both number and type. Opinions on intervention designs focusing on, for example, good argumentation in commenting or socialisation could be different. It would also be interesting to receive additional viewpoints from people who never comment on online news sites.

Despite these shortcomings, we argue that the methodological choices were justifiable vis-à-vis the set goals

because: First, the online survey enabled us to reach a large number and relatively broad spectrum of people who actively comment on online news sites. Second, presenting the designs as speculative resulted in meaningful answers. The answers offered an extensive overall picture of the potential end-users' assumptions and expectations. They offered meaningful new viewpoints and nuance to quality, and critical perspectives to the deployment of technology. Also, the speculative interfaces brought forth new insights that would remain latent when using more conventional interfaces: for example, concerning the use of metaphors and the consideration of cultural sensitivity. We note that all the identified requirements for good quality are important to some users and therefore need to be addressed when designing and publishing these kinds of systems and evaluating their quality. Also, as the requirements were spontaneously raised by the respondents, the findings could inform which user-centred qualities are relevant to measure in future studies.

6. Conclusion

This paper reported a case study on user-centred quality of UI intervention designs intended to influence online discussion in the context of news commenting. We analysed news commenters' first reactions to speculative intervention designs and the arguments they used to justify choosing between two designs. This resulted in several user requirements that relate to the communal, socio-technical perspective to news commenting as a form of social interaction and that are relatively rarely highlighted in the literature. For example, many users think a good intervention design should feature technological and/or human capability to prevent its intentional misuse. They expect the UI interventions to be objective and to utilise metaphors that are personally relevant and, hence, appropriate, and effective.

All in all, the study advances our understanding of how potential users perceive quality in UI interventions to online discussion. All the identified requirements are important to at least some users and therefore need to be addressed when designing and deploying these kinds of systems and evaluating their quality. To this end, we provide seven design considerations about different facets of user-centred quality, which can help designers make more well-informed decisions.

Acknowledgements

We thank all the participants in the survey study.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Academy of Finland [grant number 320766,320767].

ORCID

Joel Kiskola  <http://orcid.org/0000-0003-0884-8092>
 Thomas Olsson  <http://orcid.org/0000-0002-1106-2544>
 Anna Rantasila  <http://orcid.org/0000-0002-7703-401X>
 Aleks H. Syrjämäki  <http://orcid.org/0000-0003-0909-5678>
 Mirja Ilves  <http://orcid.org/0000-0002-7763-3741>
 Poika Isokoski  <http://orcid.org/0000-0002-9769-3506>
 Veikko Surakka  <http://orcid.org/0000-0003-3986-0713>

References

- Acquisti, Alessandro, Idris Adjerid, Rebecca Balebako, Laura Brandimarte, Lorrie Faith Cranor, Saranga Komanduri, Pedro Giovanni Leon, et al. 2017. "Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online." *ACM Computing Surveys* 50 (3): 1–41. doi:10.1145/3054926.
- Alexander, Christopher. 1964. *Notes on the Synthesis of Form*. Cambridge, Massachusetts: Harvard University Press.
- Auger, James. 2013. "Speculative Design: Crafting the Speculation." *Digital Creativity* 24 (1): 11–35. doi:10.1080/14626268.2013.767276.
- Baggini, Julian. 2018. *How the World Thinks: A Global History of Philosophy*. London: Granta Books.
- Bangor, Aaron, Philip T. Kortum, and James T. Miller. 2008. "An Empirical Evaluation of the System Usability Scale." *Intl. Journal of Human-Computer Interaction* 24 (6): 574–594. doi:10.1080/10447310802205776.
- Bardzell, Jeffrey, Shaowen Bardzell, and Lone Koefoed Hansen. 2015. "Immodest Proposals: Research Through Design and Knowledge." *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, 2093–2102. doi:10.1002/j.2326-1951.1971.tb00858.x.
- Bardzell, Jeffrey, 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:10.1145/2556288.2557137.
- Baumer, Eric P.S., Mark Blythe, and Theresa Jean Tanenbaum. 2020. "Evaluating Design Fiction: The Right Tool for the Job." *DIS 2020 - Proceedings of the 2020 ACM Designing Interactive Systems Conference*, 1901–1913. doi:10.1145/3357236.3395464.
- Baumer, Eric P. S., and M. S. Silberman. 2011. "When the Implication Is Not to Design (Technology).". In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems - CHI '11* 2271–2274. doi:10.1145/1978942.1979275.
- Bonino, Dario, and Fulvio Corno. 2011. "What Would you ask to Your Home if it Were Intelligent? Exploring User Expectations About Next-Generation Homes." *Journal of*

- Ambient Intelligence and Smart Environments* 3 (2): 111–126. doi:10.3233/AIS-2011-0099.
- Bovens, Luc. 2009. “The Ethics of Nudge.” In *Preference Change: Approaches from Philosophy, Economics and Psychology*, edited by Till Grüne-Yanoff and S.O. Hansson, 207–219. Berlin and New York: Springer. doi:10.1007/978-90-481-2593-7_10.
- Bradley, Alex, Claire Lawrence, and Eamonn Ferguson. 2018. “Does Observability Affect Prosociality?” *Proceedings of the Royal Society B: Biological Sciences* 285: 1875. doi:10.1098/RSPB.2018.0116.
- Cañigueral, Roser, and Antonia F.de C. Hamilton. 2019. “Being Watched: Effects of an Audience on Eye Gaze and Prosocial Behaviour.” *Acta Psychologica* 195: 50–63. doi:10.1016/J.ACTPSY.2019.02.002.
- Caraban, Ana, 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*, 1–15. doi:10.1145/3290605.3300733.
- Chan, Adrian. 2012. *Principles of Social Interaction Design*. <https://www.gravity7.com>.
- Chen, Gina Masullo, and Shuning Lu. 2017a. “Online Political Discourse: Exploring Differences in Effects of Civil and Uncivil Disagreement in News Website Comments.” *Journal of Broadcasting & Electronic Media*, 61 (1): 108–125. doi:10.1080/08838151.2016.1273922.
- Chen, Gina Masullo, and Yee Man Margaret Ng. 2017b. “Nasty Online Comments Anger you More Than me, but Nice Ones Make me as Happy as you.” *Computers in Human Behavior* 71: 181–188. doi:10.1016/j.chb.2017.02.010.
- Cheng, Justin, Michael Bernstein, Cristian Danescu-Niculescu-mizil, and Jure Leskovec. 2017. “Anyone can Become a Troll: Causes of Trolling Behavior in Online Discussions.” In *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, 1217–1230. doi:10.1145/2998181.2998213.
- Cila, Nazli, Paul Hekkert, and Valentijn Visch. 2014. “Source Selection in Product Metaphor Generation: The Effects of Salience and Relatedness.” *International Journal of Design* 8 (1): 15–28.
- Coe, Kevin, Kate Kenski, and Stephen A. Rains. 2014. “Online and Uncivil? Patterns and Determinants of Incivility in Newspaper Website Comments.” *Journal of Communication* 64 (4): 658–679. doi:10.1111/jcom.12104.
- Cook, Christine L., Aashka Patel, and Donghee Yvette Wohn. 2021. “Commercial Versus Volunteer: Comparing User Perceptions of Toxicity and Transparency in Content Moderation Across Social Media Platforms.” *Frontiers in Human Dynamics* 3 (February): 1–8. doi:10.3389/fhumd.2021.626409.
- Cox, Anna L., Sandy Gould, Marta E. Cecchinato, Ioanna Iacovides, and Ian Renfree. 2016. “Design Frictions for Mindful Interactions: The Case for Microboundaries.” In *Conference on Human Factors in Computing Systems - Proceedings*, 1389–1397. doi:10.1145/2851581.2892410.
- de Carvalho, Mariana Linhares, Thomas Olsson, and Joel Kiskola. 2021. “Exploration of User Interface Mechanisms with Affect Labeling to Enhance On-line Discussion.” In *AcademicMindtrek '21: Proceedings of the 24rd International Conference on Academic Mindtrek*.
- Desmet, Pieter, and Paul Hekkert. 2007. “Framework of Product Experience.” *International Journal of Design* 1 (1): 57–66.
- Diakopoulos, Nicholas, 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* 1133–142. doi:10.1.1.188.3516.
- Diefenbach, Sarah, Nina Kolb, and Marc Hassenzahl. 2014. “The “Hedonic” in Human-Computer Interaction: History, Contributions, and Future Research Directions.” In *Proceedings of the 2014 Conference on Designing Interactive Systems*. Accessed June 17, 2021. doi:10.1145/2598510.2598549.
- Dillahunt, Tawanna, Geoff Becker, Jennifer Mankoff, and Robert Kraut. 2008. “Motivating Environmentally Sustainable Behaviour Changes with a Virtual Polar Bear.” In *Pervasive 2008 Workshop Proceedings*, 58–62. doi:10.1504/IJARGE.2002.000023.
- Eberwein, Tobias. 2019. ““Trolls” or “Warriors of Faith”? Differentiating Dysfunctional Forms of Media Criticism in Online Comments.” *Journal of Information, Communication and Ethics in Society* 18 (1): 131–143. doi:10.1108/JICES-08-2019-0090.
- Entman, Robert M. 1993. “Framing: Toward Clarification of a Fractured Paradigm.” *Journal of Communication* 43 (4): 51–58. doi:10.1111/j.1460-2466.1993.tb01304.x.
- Fogg, Bj. 2009. “A Behavior Model for Persuasive Design.” In *Proceedings of the 4th International Conference on Persuasive Technology* 2009 Apr 26, 1–7. doi:10.1145/1541948.1541999.
- Galitz, Wilbert O. 2007. *The Essential Guide to User Interface Design: An Introduction to GUI Design Principles and Techniques*. 2. New York, Chichester, Weinheim, Brisbane, Singapore, Toronto: John Wiley & Sons.
- Gillespie, Tarleton. 2018. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven & London: Yale University Press.
- Grön, Kirsikka, and Matti Nelimarkka. 2020. “Party Politics, Values and the Design of Social Media Services.” In *Proceedings of the ACM on Human-Computer Interaction*. doi:10.1145/3415175.
- Hart, Sandra G. 2016. “Nasa-Task Load Index (NASA-TLX); 20 Years Later.” <http://dx.doi.org.libproxy.tuni.fi/10.1177/154193120605000909>: 904–908. doi:10.1177/154193120605000909.
- Hassenzahl, Marc, Michael Burmester, and Franz Koller. 2003. “AttrakDiff: Ein Fragebogen zur Messung Wahrgenommener Hedonischer und Pragmatischer Qualität.” *Vieweg + Teubner Verlag*, 187–196. doi:10.1007/978-3-322-80058-9_19.
- Heikkinen, Jani, Thomas Olsson, and Kaisa Väänänen-Vainio-Mattila. 2009. Expectations for User Experience in Haptic Communication with Mobile Devices. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '09*.
- Herbig, Nico, Patrick Schuck, and Antonio Krüger. 2019. “User Acceptance of Cognition-Aware E-Learning: An Online Survey CCS CONCEPTS.” *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia*. doi:10.1145/3365610.

- Isaias, Pedro, and Issa Tomayess. 2015. *High Level Models and Methodologies for Information Systems*, 91–120. New York, NY: Springer.
- Jakobi, Timo, Sameer Patil, Dave Randall, Gunnar Stevens, and Volker Wulf. 2019. "It is About What They Could do with the Data: A User Perspective on Privacy in Smart Metering." *ACM Transactions on Computer-Human Interaction* 26 (1): 1–44. doi:10.1145/3281444.
- Johannessen, Leon Karlsen. 2017. *The Young Designer's Guide to Speculative and Critical Design*. Norwegian University of Science and Technology.
- Johannessen, Leon Karlsen, M. M. Keitsch, and I. N. Pettersen. 2019. "Speculative and Critical Design-Features, Methods, and Practices." *Proceedings of the Design Society: International Conference on Engineering Design* 1 (1): 1623–1632. doi:10.1017/dsi.2019.168.
- Kiskola, Joel, Thomas Olsson, Heli Väättäjä, Aleks H. Syrjämäki, Anna Rantasila, Poika Isokoski, Mirja Ilves, and Veikko Surakka. 2021. "Applying Critical Voice in Design of User Interfaces for Supporting Self-reflection and Emotion Regulation in Online News Commenting." In *CHI '21: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3411764.3445783.
- Klein, Gary. 2007. "Performing a Project Pre Mortem." *Harvard Business Review* 85 (9): 18–19.
- Lin, James J, Lena Mamykina, Silvia Lindtner, Gregory Delajoux, and Henry B Strub. 2006. "Fish'nSteps Encouraging Physical Activity with an Interactive Computer Game." In *International Conference on Ubiquitous Computing*, 261–278. https://www.ics.uci.edu/~lindtner/documents/Lin_FishnSteps2006.pdf.
- Lin, Angela, and Leiser Silva. 2005. "The Social and Political Construction of Technological Frames." *European Journal of Information Systems* 14 (1): 49–59. doi:10.1057/palgrave.ejis.3000521.
- Maia, Rousiley C. M., and Thaianie A. S. Rezende. 2016. "Respect and Disrespect in Deliberation Across the Networked Media Environment: Examining Multiple Paths of Political Talk." *Journal of Computer-Mediated Communication* 21 (2): 121–139. doi:10.1111/JCC4.12155.
- Merton, Robert K. 1936. "The Unanticipated Consequences of Purposive Social Action." *American Sociological Review* 1 (6): 894–904.
- Nelson, Harold G, and Erik Stolterman. 2012. *The Design Way: Intentional Change in an Unpredictable World*. Cambridge, Massachusetts: The MIT Press.
- Nicholas, Jennifer, Andrea S. Fogarty, Katherine Boydell, and Helen Christensen. 2017. "The Reviews are in: A Qualitative Content Analysis of Consumer Perspectives on Apps for Bipolar Disorder." *Journal of Medical Internet Research* 19: 4. doi:10.2196/jmir.7273.
- Olsson, Thomas, Tuula Kärkkäinen, Else Lagerstam, and Leena Ventä-Olkkonen. 2012. "User Evaluation of Mobile Augmented Reality Scenarios." *Journal of Ambient Intelligence and Smart Environments* 4 (1): 29–47. doi:10.3233/AIS-2011-0127.
- Olsson, Thomas, Else Lagerstam, Tuula Kärkkäinen, and Kaisa Väänänen-Vainio-Mattila. 2013. "Expected User Experience of Mobile Augmented Reality Services: A User Study in the Context of Shopping Centres." *Personal and Ubiquitous Computing* 17 (2): 287–304. doi:10.1007/s00779-011-0494-x.
- Orlikowski, Wanda J., and Debra C. Gash. 1994. "Technological Frames: Making Sense of Information Technology in Organizations." *ACM Transactions on Information Systems (TOIS)* 12 (2): 174–207. doi:10.1145/196734.196745.
- Ortony, Andrew, Richard J. Vondruska, Mark A. Foss, and Lawrence E. Jones. 1985. "Salience, Similes, and the Asymmetry of Similarity." *Journal of Memory and Language* 24 (5): 569–594.
- Palan, Stefan, and Christian Schitter. 2018. "Prolific.ac—A Subject Pool for Online Experiments." *Journal of Behavioral and Experimental Finance* 17: 22–27. doi:10.1016/j.jbef.2017.12.004.
- Pettré, Caitlin, Brooke Erin Duffy, and Emily Hund. 2019. "Gaming the System.: Platform Paternalism and the Politics of Algorithmic Visibility." *Social Media + Society* 5: 4. doi:10.1177/2056305119879995.
- Phillips Davison, W. 1983. "The Third-Person Effect in Communication." *Public Opinion Quarterly* 47 (1): 1–15. doi:10.1086/268763.
- Saldaña, Johnny. 2013. *The Coding Manual for Qualitative Researchers*. 2. Los Angeles, London, New Delhi, Singapore, Washington DC: SAGE Publications Ltd.
- Sánchez-Adame, Luis Martín, José Fidel Urquiza-Yllescas, and Sonia Mendoza. 2020. "Measuring Anticipated and Episodic ux of Tasks in Social Networks." *Applied Sciences (Switzerland)* 10 (22): 1–17. doi:10.3390/app10228199.
- Seering, Joseph, 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." In *Conference on Human Factors in Computing Systems - Proceedings*, 1–14. doi:10.1145/3290605.3300836.
- Shilton, Katie. 2018. "Values and Ethics in Human-Computer Interaction." *Foundations and Trends in Human-Computer Interaction* 12 (2): 107–171. doi:10.1561/11000000073.
- Sohn, Seohee, Ho Chung Chung, and Namkee Park. 2019. "Private Self-Awareness and Aggression in Computer-Mediated Communication: Abusive User Comments on Online News Articles." *International Journal of Human-Computer Interaction* 35 (13): 1160–1169. doi:10.1080/10447318.2018.1514822.
- Sparrow, Lucy A., Martin Gibbs, and Michael Arnold. 2021. "The Ethics of Multiplayer Game Design and Community Management." *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–13. doi:10.1145/3411764.3445363.
- Stroud, Natalie Jomini, Emily Van Duyn, and Cynthia Peacock. 2016. "News Commenters and News Comment Readers."
- Sunstein, Cass R. 2018. "“Better off, as Judged by Themselves”: A Comment on Evaluating Nudges." *International Review of Economics* 65 (1): 1–8. doi:10.1007/s12232-017-0280-9.
- Suri, Jane Fulton. 2002. "Designing Experience: Whether to Measure Pleasure or Just Tune in." In *Pleasure with Products: Beyond Usability*, edited by William S. Green and Patrick W. Jordan, 161–174. London: CRC Press.
- Taylor, Samuel Hardman, Dominic Difranzo, Yoon Hyung Choi, Shruti Sannon, and Natalya N. Bazarova. 2019.

- "Accountability and Empathy by Design: Encouraging Bystander Intervention to Cyberbullying on Social Media." *Proceedings of the ACM on Human-Computer Interaction* 3. doi:10.1145/3359220.
- Thaler, Richard H., and Cass R. Sunstein. 2009. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New York: Penguin.
- Tharp, Bruce M., and Stephanie Tharp. 2019. *Discursive Design: Critical, Speculative, and Alternative Things*. Cambridge, Massachusetts & London, England: MIT Press.
- Tidwell, Jennifer, Charles Brewer, and Aynne Valencia. 2020. "Designing Interfaces: Patterns for Effective Interaction Design". 3. Beijing, Boston, Farnham, Sebastopol, Tokyo: O'Reilly Media.
- Topal, Kamil, Mehmet Koyuturk, and Gultekin Ozsoyoglu. 2016. "Emotion -and Area-Driven Topic Shift Analysis in Social Media Discussions." In *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2016*, 510–518. doi:10.1109/ASONAM.2016.7752283.
- Torre, Jared B., and Matthew D. Lieberman. 2018. "Putting Feelings Into Words: Affect Labeling as Implicit Emotion Regulation." *Emotion Review* 10 (2): 116–124. doi:10.1177/1754073917742706.
- Tromp, Nynke, Paul Hekkert, and Peter Paul Verbeek. 2011. "Design for Socially Responsible Behavior: A Classification of Influence Based on Intended User Experience." *Design Issues* 27 (3): 3–19. doi:10.1162/DESI_a_00087.
- Van Der Vegte, Wilfred, Yoshinobu Kitamura, Yusuke Koji, and Riichiro Mizoguchi. 2004. "Coping with Unintended Behavior of Users and Products: Ontological Modelling of Product Functionality and Use." In *Proceedings of the ASME Design Engineering Technical Conference*, 683–692. doi:10.1115/detc2004-57720.
- Walther, Joseph B. 1993. "Impression Development in Computer-Mediated Interaction." *Western Journal of Communication* 57 (4): 381–398. doi:10.1080/10570319309374463.
- Wang, Sai. 2021. "Moderating Uncivil User Comments by Humans or Machines? The Effects of Moderation Agent on Perceptions of Bias and Credibility in News Content." *Digital Journalism* 9 (1): 64–83. doi:10.1080/21670811.2020.1851279.
- Wang, Yang, Saranga Komanduri, Pedro Giovanni Leon, Gregory Norcie, Alessandro Acquisti, and Lorrie Faith Cranor. 2011. "'I Regretted the Minute I Pressed Share': A Qualitative Study of Regrets on Facebook." In *Proceedings of the Seventh Symposium on Usable Privacy and Security - SOUPS '11*. doi:10.1145/2078827.
- Wang, Yang, Pedro Giovanni Leon, Alessandro Acquisti, Lorrie Faith Cranor, Alain Forget, and Norman Sadeh. 2014. "A Field Trial of Privacy Nudges for Facebook." In *Conference on Human Factors in Computing Systems - Proceedings*, 2367–2376. doi:10.1145/2556288.2557413.
- Wang, Yang, Pedro Giovanni Leon, Kevin Scott, Xiaoxuan Chen, Alessandro Acquisti, and Lorrie Faith Cranor. 2013. "Privacy Nudges for Social Media." In *Proceedings of the 22nd international conference on world wide web*, 22: 763–770. doi:10.1145/2487788.2488038.
- Weinmann, Markus, Christoph Schneider, and Jan vom Brocke. 2016. "Digital Nudging." *Business and Information Systems Engineering* 58 (6): 433–436. doi:10.1007/s12599-016-0453-1.
- Whitworth, Brian. 2009. "The Social Requirements of Technical Systems." In *Handbook of Research on Socio-Technical Design and Social Networking Systems*, edited by Brian Whitworth and Aldo De Moor, 2–22. Hershey & New York: Information Science Reference.
- Whitworth, Brian, and Michael Zaic. 2003. "The WOSP Model: Balanced Information System Design and Evaluation." *Communications of the Association for Information Systems* 12. doi:10.17705/1cais.01217.
- Xenophon, Carleton L, E. C. Marchant Brownson, O. J. Todd, and Walter Miller. 1979. *Xenophon in Seven Volumes*. Cambridge, MA: Harvard University Press.
- Yogasara, Thedy, Vesna Popovic, Ben Kraal, and Mari- anella Chamorro-Koc. 2011. "General Characteristics of Anticipated User Experience (AUX) with Interactive Products." *Diversity and Unity: Proceedings of IASDR2011, the 4th World Conference on Design Research*, 1–11.
- Yoon, JungKyo, Shuran Li, Yu Hao, and Chajoong Kim. 2019. "Towards Emotional Well-Being by Design." In *Pervasive Health' 19: 13th EAI International Conference on Pervasive Computing Technologies for Healthcare*, 351–355. doi:10.1145/3329189.3329227.
- Ziegele, Marc, Mathias Weber, Oliver Quiring, and Timo Breiner. 2018. "The Dynamics of Online News Discussions: Effects of News Articles and Reader Comments on Users' Involvement, Willingness to Participate, and the Civility of Their Contributions." *Information, Communication & Society* 21 (10): 1419–1435. doi:10.1080/1369118X.2017.1324505.
- Zimmerman, John, and Jodi Forlizzi. 2014. "Research Through Design in HCI." In *Ways of Knowing in HCI*, edited by Judith S. Olson and Wendy A. Kellogg, 167–189. New York: Springer. doi:10.1007/978-1-4939-0378-8_8.
- Zimmermann, Verena, and Karen Renaud. 2021. "The Nudge Puzzle: Matching Nudge Interventions to Cybersecurity Decisions." *ACM Transactions on Computer-Human Interaction* 28 (1): 1–45. doi:10.1145/3429888.

Appendix 1. Designs as they were shown in the survey.

Participant was shown two of the designs. The following freely available resources were used in making the designs: Semantic UI kit. Icons: Font Awesome, Ionic and Feather.

Evaluate

You are reading the comments to an interesting but divisive news article ...and wish to add your own comment

Comment

Comments



Matt 2h ago

I remember seeing these news in the 80s. The history repeats itself, again. The Z-generation shouldn't be scared, we've lived through worse.

36 ^ v • Reply



John Doe 2h ago

The article is full of lies. Who believes this shit??? I mean, you have to be a low-IQ individual, or a member of the political party to believe it.

36 ^ v • Reply



Jordan 2h ago

Hell no! I will drive till I die, and I sure as heck will travel by plane

36 ^ v • Reply



Jenny 2h ago

Me too! I don't want to change my lifestyle because of this

36 ^ v • Reply



Riley 2h ago

Now this is what I call journalism, great job!

36 ^ v • Reply

When you click to "Comment", you first need to tell how you feel before adding your comment.

1. Please tell which of the following best describes how you feel after reading the news:



2. Add your comment

Post

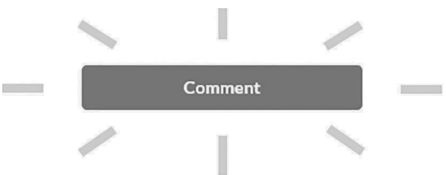
Creature

You are reading a news article online

Breaking News: Political Division all Time High

Political polarization – the vast and growing gap between liberals and conservatives, Democrats and Republicans – is a defining feature of American politics in 2029. 46% of U.S. citizens, almost all of them Republican, say the president did something wrong regarding the Gulf of Mexico oil spill and that it was enough to justify her removal from office. Another 28% of U.S. citizens say the president did something wrong but that it was not enough to warrant her removal, while 25% say she did nothing wrong.

..and then you press "Comment."



As you are writing your comment, an animated digital creature reacts to its emotional tone.

Add your comment

What is wrong with people?? We should focus on the good things that we share. We all care about our loved ones, the environment, jobs, etc. I

How our digital friend feels about what you've written thus far



Neutral

Post

For example, you write

In a fairly
positive way



Very happy

In a rather
negative way



Very sad

Highlight

You are reading the comments to an interesting but divisive news article. You are offered an option to view automatic analysis of emotions in the comments.

Comments

☐ Show automatically generated comment emotion analysis



Matt 2h ago

I remember seeing these news in the 80s. The history repeats itself, again. The Z-generation shouldn't be scared, we've lived through worse.



John Doe 2h ago

The article is full of lies. Who believes this shit??? I mean, you have to be a low-IQ individual, or a member of the political party to believe it.



Jordan 2h ago

Hell no! I will drive till I die, and I sure as heck will travel by plane



Jenny 2h ago

Me too! I don't want to change my lifestyle because of this



Riley 2h ago

Now this is what I call journalism, great job!

You check "Show automatically generated comment emotion analysis" and see the analysis.

Comments

☒ Show automatically generated comment emotion analysis



Matt 2h ago

I remember seeing these news in the 80s. The history repeats itself, again. The Z-generation shouldn't be scared, we've lived through worse.



John Doe 2h ago 

The article is full of lies. Who believes this shit??? I mean, you have to be a low-IQ individual, or a member of the political party to believe it.



Jordan 2h ago 

Hell no! I will drive till I die, and I sure as heck will travel by plane



Jenny 2h ago

Me too! I don't want to change my lifestyle because of this



Riley 2h ago

Now this is what I call journalism, great job!

You are reading the comments to an interesting but divisive news article. You are offered a way to give anonymous and private feedback to any of the previous commentators.

Comments



Matt 2h ago

I remember seeing these news in the 80s. The history repeats itself, again. The Z-generation shouldn't be scared, we've lived through worse.

36 ^ v • Reply • Give private feedback:    



John Doe 2h ago

The article is full of lies. Who believes this shit??? I mean, you have to be a low-IQ individual, or a member of the political party to believe it.

36 ^ v • Reply • Give private feedback:    



Jordan 2h ago





Hell no! I will drive till I die, and I sure as heck will travel by plane

36 ^ v • Reply • Give private feedback:    



Jenny 2h ago

Me too! I don't want to change my lifestyle because of this

36 ^ v • Reply • Give private feedback:    









Riley 2h ago


Now this is what I call journalism, great job!

36 ^ v • Reply • Give private feedback:    

So, you decide to share your feelings with John, using the symbols next to the comment.

**John Doe** 2h ago
The article is full of lies. Who believes this shit??? I mean, you have to be a low-IQ individual, or a member of the political party to believe it.
36 ^ v • Reply • Give private feedback:    

**Jordan** 2h ago
Hell no! I will drive till I c
36 ^ v • Reply • Give p

**Jenny** 2h ago
Me too! I don't v
36 ^ v • Reply

By clicking, you add one
"Full of arrogance" to
John's arrogance counter
anonymously. John will see
the counter in their profile.

[More information](#)

The alternative feedback symbols:

- | | |
|---|---|
|  "Love it!" |  "False claim/s" |
|  "Well said" |  "Full of arrogance" |

A section in John Doe's profile shows the feedback from other users.

John Doe's User Profile

Overview of the feedback from other users



*Icon size corresponds to the amount of feedback

My Comments

**John Doe** 7h ago

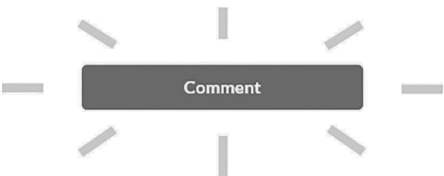
Audience

You are reading a news article online

Breaking News: Political Division all Time High


Political polarization – the vast and growing gap between liberals and conservatives, Democrats and Republicans – is a defining feature of American politics in 2029. 46% of U.S. citizens, almost all of them Republican, say the president did something wrong regarding the Gulf of Mexico oil spill and that it was enough to justify her removal from office. Another 28% of U.S. citizens say the president did something wrong but that it was not enough to warrant her removal, while 25% say she did nothing wrong.

..and then you press "Comment."



As you are writing your comment, a virtual audience of expert judges reacts to its tone.

Add your comment

What is wrong with people?? We should focus on the good things that we share. We all care about our loved ones, the environment, jobs, etc. 

How expert judges might react to your tone



Post

For example, if your comment is written

In a moderately positive way



In a rather negative way



Regret

You, Anon123 are reading a news article online

Breaking News: Political Division all Time High

Political polarization – the vast and growing gap between liberals and conservatives, Democrats and Republicans – is a defining feature of American politics in 2029. 46% of U.S. citizens, almost all of them Republican, say the president did something wrong regarding the Gulf of Mexico oil spill and that it was enough to justify her removal from office. Another 28% of U.S. citizens say the president did something wrong but that it was not enough to warrant her removal, while 25% say she did nothing wrong.

..and then you post an angry comment on it.

The article is obviously full of lies. Only a low-IQ individual, or a party voter can believe this crap.



Now, the comment you, Anon123 posted a couple of seconds ago is automatically evaluated and you are offered different follow-up actions.

Comments



Anon123 Just posted

The article is obviously full of lies. Only a low-IQ individual, or a party voter can believe this crap.

Your comment may sound angry

I regret my choice of words

Delete comment

Edit

[More information](#)

Reply to your own comment



Matt 2h ago

I remember seeing these news in the 80s. The history repeats itself, again. The Z-generation shouldn't be scared, we've lived through worse.

36 ^ v • Reply



Riley 2h ago

Now this is what I call journalism, great job!

36 ^ v • Reply

The system also emails you about it.

The comment you posted may sound angry

Automoderator automoderator@commentingplatform.com

Dear Anon123,

Automoderator has identified that the comment you posted may sound angry. If this is not the case, please ignore this message.



Anon123 2h ago

The article is obviously full of lies. Only a low-IQ individual, or a party voter can believe this crap.

You may press one of the following buttons in case you want to regret, delete or edit your comment:

I regret my choice of words

Delete comment

Edit

[More information](#)

Have a nice day,
Automoderator

After thinking about it, you decide to use the Regret feature. A special label is added to your comment.

Comments



Anon123 2h ago

The article is obviously full of lies. Only a low-IQ individual, or a party voter can believe this crap.

Anon123 regretted their angry words

[About this feature](#)

Reply to your own comment



Matt 2h ago

I remember seeing these news in the 80s. The history repeats itself, again. The Z-generation shouldn't be scared, we've lived through worse.

36   • Reply



Riley 2h ago

Now this is what I call journalism, great job!

36   • Reply

Also, in case another user writes a reply to your comment, they are reminded that you regretted your words.


Reply to Anon123

Anon123 regretted their angry words

It's 

Post

Philosophy

You are reading the comments to an interesting but divisive news article. The problematic comments and comment threads are marked with an  icon.

Comments



Matt 2h ago

I remember seeing these news in the 80s. The history repeats itself, again. The Z-generation shouldn't be scared, we've lived through worse.

12   • Reply



John Doe 2h ago 

The article is full of lies. Who believes this shit??? I mean, you have to be a low-IQ individual, or a member of the political party to believe it.

36   • Reply



Jordan 2h ago 

Hell no! I will drive till I die, and I sure as heck will travel by plane

23   • Reply


50%
Negative
thread



Jenny 2h ago

Me too! I don't want to change my lifestyle because of this

24   • Reply




Riley 2h ago

Now this is what I call journalism, great job!

36   • Reply

When you press the icon, it reveals an emotion score and a quote from philosopher Socrates.

Comments



Matt

2h ago

I remember seeing
again. The Z-gene
worse.

12


^

v

•

Reply

Academic estimate of the comment's
tone: Positivity -10. Calmness -20.
Socrates might say: "Know thyself"
[About this feature](#)



John Doe

2h ago

The article is full of lies. Who believes this shit??? I mean, you have to be a
low IQ individual or a member of the political party to believe it.

50


^

v


•

Reply

50% of comments in the thread have positivity
below -10 and calmness below -10.
Socrates might say: "Know thyself"
[About this feature](#)



will travel by plane



Jenny

2h ago

Me too! I don't want to change my lifestyle because of this


24

^

v

•

Reply



Riley

2h ago

Now this is what I call journalism, great job!

36

^

v

•

Reply

Warning

You are about to read the comments to an interesting but divisive news article. You are shown a notification about the argumentation that the comments include.

Breaking News: Political Division all Time High

Political polarization – the vast and growing gap between liberals and conservatives, Republicans and Democrats – is a defining feature of American politics in 2029. 46% of U.S. citizens, almost all of them Republican, say the president did something wrong regarding the Gulf of Mexico oil spill and that it was enough to justify her removal from office. Another 28% of U.S. citizens say the president did something wrong but that it was not enough to warrant her removal, while 25% say she did nothing wrong.



The discussion around this article contains



10% Hatefulness 5% Provocation 5% Encouragement 5% Agreement

Comments



Matt 2h ago

I remember seeing these news in the 80s. The history repeats itself, again. The Z-generation shouldn't be scared, we've lived through worse.

PUBLICATION IV

Evaluating Alerts to Impolite Online News Commenters: The Impact of Previous Commenter's Politeness and the Form and Amount of Guidance

Kiskola, J., Olsson, T., Syrjämäki, A. H., Rantasila, A., Ilves, M., Isokoski, P., & Surakka, V.

Under review at Behaviour & Information Technology

**Publication is licensed under a Creative Commons Attribution 4.0
International License CC-BY-NC-ND**

