We thank the reviewers for the feedback. We have revised the paper accordingly, and below we respond to each of the specific reviewer comments.

Reviewer 4 (1AC)

R1 and R2 found the work to be highly relevant and timely and well-motivated (R1) as well as well-written and easy to follow (R2).
We thank the reviewers for recognizing the value in our work.

However, the paper received mixed reviews: MAJOR, MAJOR/ MINOR, and MINOR. The paper was discussed at the first editorial board meeting and we decided for MINOR revisions, which will be reviewed again after submission by ACs as well as external reviewers. The following concerns need to be addressed before the final decision at the second editorial board meeting.
We appreciate the faith of the editorial board and hope that the revision we have prepared based on addressing the detailed feedback from the reviewers satisfies the reviewers, ACs, and the editorial board. If needed, we would be happy to make further edits to incorporate additional reviewer feedback on the revision.

2AC gave suggestions on how to shorten the paper to make space for the additional content.
Thank you. We have tried our best to reduce the paper length in spite of adding content to the Discussion and Implication sections as requested by the reviewers. Please note that figures and tables might still make the paper appear longer in terms of page count. If reviewers have further specific suggestions for what could be cut to reduce length, we would be happy to incorporate them.

NOVELTY & IMPLICATIONS Reviewers' main concern is on the novelty of the implications. We have added a Discussion section discussing the novelty and importance of our findings and included further detail regarding the suggestions in the implications and their ties to our findings.

2AC's main criticism is the length of the paper compared to its contributions, as they point out that the discussion and implications providing little novelty and insights. We have made a full pass through the paper for brevity, clarity, focus, flow, and coherence. We hope that the length of the paper is now appropriate. We would be happy to make further revisions to reduce length should reviewers deem it necessary.

Similarly, R1 points out that "the connection between findings and implications can be improved" and asks to expand on the implications, in particular the last two (similar as 2AC). I agree that the implications (in particular the first two) are not really novel. Limiting notifications does not really do anything for protecting users' privacy, as they might be overseen even when they are arriving more spread out or at intervals. Furthermore, users might still be wanting to receive notifications in real time. The actual novel implications are only briefly mentioned at the end of 5.3 and in 5.4, but only superficially (as pointed out by reviewers). We have refined the implications section to improve the connection with the findings. As we point out, limiting notifications is connected to privacy via the 'intrusion' facet of privacy [7]. Moreover, reducing notification frequency reduces the number of occasions for potential privacy violations, thus reducing the risk.

Note also that allowing users better control over notification frequency and volume does not preclude them from receiving some (or all) notifications in real time, should they so desire. Further, when coupled with our suggestions for keyword based filtering, users can ensure that notifications that

match certain criteria can be delivered in real time. We have added further detail to the last two implications.

Connected to that, R1 "would have loved to see a more in-depth discussion on existing privacy measures for notifications on mobile devices and their effectiveness". And I agree with this. Going back to the implications, the first two are already available in mobile OSs. An overview over the state of the art of how applications as well as devices give users control over notifications could have helped to situate the findings better. For example, many popular message apps (WhatsApp, Signal, Telegram) allow users to set chat specific notification settings, and Android as well as iOS have extensive settings that also allow for privacy protection. None of these have been covered by the questionnaire. We refined the text to point out notification control mechanisms in current operating systems or apps, where relevant. We did not include these in the questionnaire because several are available only in the most recent versions of the operating systems or apps, so not everyone who responded to the questionnaire may be familiar with them. Regardless, as we found, participant reported little use of any control features beyond the crude ones (e.g., silent mode). Further, these features do not address all of the privacy-related issues connected to notifications that we report in the paper.

An overview / discussion of existing privacy measures should be added. Where relevant, we have pointed to existing features of operations systems or apps that can potentially be used to manage notification privacy.

At the same time remove the focus on productivity/interruption from related work—or elaborate on the relevance. As we explain in more detail below, we removed the text pertaining to productivity/interruption.

RELATED WORK R1 suggests removing the focus on productivity in the related work, and instead focus on privacy concerns related to notification in 2.3.1. As mentioned above and detailed below, we removed the focus on productivity in the related work.

On the same note R2 asks for the connection between the quote from [18] and the presented work to be made clearer. We removed the quote.

R1 asks "why there is a change in notification privacy as compared to related work from 2012" as stated by the paper. As explained below, we have revised the content, coherence, and flow of the related work such that this comment is no longer applicable.

In several places, R1 gives very valuable advice and pointers to related work that could help strengthen the paper. We incorporated all pointers to related work provided by R1.

METHODOLOGY The study participation was limited to US residents. However, R1 questions why this choice has been made, yet ethnic backgrounds of participants were collected. Is there indeed a cultural or ethnic variation? As we explain below in detail, the participation was restricted to US residents to avoid introducing cultural variance (since privacy is known to vary across cultures). As we explain, the collection of ethnic backgrounds is a standard practice useful for describing the sample and unconnected to cultural variation. We do not have sufficient data to investigate ethnic variation within the US cultural context (nor was that the focus of our investigation).

Furthermore, R1 points out several limitations in the findings due to the wording of the question-

naire. This should be discussed. As we explained in the Method section, the questionnaire went through multiple rounds of pilot testing (with native English speakers from the United States) for clarity and comprehension. Moreover, we received no comments or complaints from participants regarding lack of clarity. Therefore, we are confident that the questionnaire as worded was understood by the participants as we intended.

R1 asks for more details on the analysis approach used. It seems to be an adaptation of thematic analysis, which could have introduced a heavy researcher bias. This should be discussed or the approach explained in more detail. We added a "Data Analysis" subsection in the Findings section that describes our procedure in detail, including the measures we took to avoid researcher bias.

PRESENTATION Furthermore, R2 requests careful proofreading, as many sentences are lacking a word or contain wrong word and reference 1 is wrong not correct. E.g. "...reported least one .... –¿ ...reported at least one ...." and "Majority of the participants (60%) reported *at* least one negative experience connected to notifications." We have made a full proofreading/editing pass through the entire paper to avoid such errors. Hopefully, we caught and fixed all of these issues!

I would further like to point out that there is a discrepancy of number of participants in abstract (235) and in line 256 (213). Thank you for catching this error. We corrected the abstract to 213.

MINOR COMMENT A non-standard paper format was used, resulting potentially in an increased page count. I suggested Letter or A4 format. Thank you for catching this error. We removed the acmsmall option from the LaTeX documentclass command. It did not result in much difference in page count.

Reviewer 3 (2AC)

This submission presents the results of a study with N=235 participants (study implemented via a questionnaire in Amazon Mechanical Turk) that were asked to describe their practices and preferences regarding smartphone notifications, such as their negative experiences related to notifications. The results of the study were used to propose four design recommendations for better ways to deliver smartphone notifications that are privacy-oriented.

The paper is long and I am sure that many sections can be shortened, such as the description of the method (Section 3) and condensing the presentation of the results in Section 4 (for instance, Tables 5 and 8 can be merged, perhaps Tables 4 and 7 as well, etc.), the appendix can be excluded from the paper and added as supplementary material.
We have made a shortening pass through the entire paper. We shortened the description of the Method and the presentation of the results. We merged Tables 5 and 8 and 4 and 7. We removed the Appendix from the paper and provided it as supplementary material.

The most important part of this work is the implications (Section 5) that, unfortunately, occupy just 1.5 pages out of the 19 pages of the manuscript (excluding the references and the appendix). Moreover, the implications seem standard and not particularly insightful: (1) limit the number of notifications, (2) disable notifications, (3) personalize what notifications to receive and how frequently, and (4) automatically turn off notifications for shared devices. The last implication is the most interesting one, but it is the least described and clearly something for future work in terms of implementation and validation, while the first three implications are just standard for mobile devices.

We have separated the discussion aspects of the findings into a new Discussion section and expanded the Implications sections to add more detail to these suggestions. While the first three implications might be standard, they are either not currently present (at least as far as notifications are concerned) or not rich or convenient enough to be usable such that privacy violations of the kinds surfaced in our study can be avoided. However, removed the second implications because the DND mode in the latest Android version is quite similar. As we explained above, we have clarified the connection of the first one (third one in the revision) to privacy and included additional detail. The third one in the original was found to be valuable by R1 so we have kept and expanded it (first one in the revision). For each of the three implications in the revision, we added further explanation that connects them to our findings and explains how these might be implemented to build upon existing (standard) mechanisms in mobile devices.

In conclusion, I think that this paper could be improved a lot by shortening Sections 1 to 4 and by considerably adding to the discussion from Section 5 and making implications more insightful to foster new work and development beyond what already exists.
As mentioned above, we shortened Sections 1 to 4, added to the discussion in Section 5, and developed the implications section with more insightful suggestions.

I am therefore recommending a "Major Revision" for this submission because it needs a second round of review.
Thank you for allowing us the opportunity to revise the paper. We believe that reviewer feedback has helped us greatly improve the paper, and we hope that reviewers are satisfied with the revision.

Other comment: - Figures 5, 6, and 8 are stretched horizontally, probably by accident
Thank you for noticing this glitch. We have improved readability, presentation, and layout of all figures.

Reviewer 1

The authors explore privacy concerns and potential intrusion by notifications on mobile devices, by means of an online survey (N=235). They found that the majority of their participants at least once encountered privacy intrusion by mobile notifications. Based on their findings, the authors suggest various designs to address this issue.

The authors' research focus is definitely highly relevant and timely. The authors motivate their work well throughout the Introduction and Related Work.

The study seems well thought-through except for minor limitations that the authors acknowledge in Section 6.
Thank you.

However, I have the following concerns with the current state of the paper:
We have responded to each concern below.

Related Work ———— The authors provide an extensive overview of related work in Section 2.

However, I have two suggestions:

I believe Section 2.2 could motivate more clearly why there is a change in notification privacy as compared to related work from 2012: e.g., smartphones are now even more ubiquitous, and there is an increasing amount of apps with potentially sensitive notification content (e.g., very recently, contact-tracing apps in relation to Covid-19).
We have revised this section such that the above issue does not apply anymore (because we have cited more recent work and because we point out that none of the work in this subsection explicitly considers privacy of notifications).

I was wondering why Section 2.3 has a strong focus on productivity/interruptions, as I did not see this at the focus of this work. I think the more interesting and relevant part is what is being presented in 2.3.1, regarding privacy concerns related to notifications. The authors could revise this Section to focus more on this rather than the productivity/interruptions related aspects.
Thank you for pointing this out. We agree. We have significantly trimmed the content related to interruptions/productivity (except for retaining parts that are connected to the content of our paper) and merged it with 2.3.1 for a focused discussion such that there is no need to have a separate subsubsection within this subsection.

Method ————

Regarding the method, I was wondering why the authors limited their study to US participants, as especially a comparison between cultures could be of interest in privacy-related topics.
As we have pointed out in the paper, it is already known that privacy varies by culture. Including participants from a variety of cultures would have required a substantially larger sample for adequate statistical power to make meaningful cultural comparisons. Moreover, Mechanical Turk does not have the same kind of adoption outside of the United States and India. Restricting the sample to a single cultural context (i.e., United States) permitted us to collect data from a sample matched to our resource constraints without inducing cultural variance that would have lowered statistical power. That said, we wholeheartedly agree that cultural comparisons would be of interest, and we have pointed this out as an avenue for future work when mentioning the limitations of our sample in the Limitations section.

Also, it was unclear to me why the authors asked for participants' ethnic backgrounds. I did not see this impacting the analysis and the collection of participants' personal data should generally be kept to a minimum.
We asked the standard demographic questions that are typically asked in surveys in the United States. While our analysis did not include direct consideration of ethnic backgrounds, the information is important to report the characteristics of the sample from which the findings are drawn. For instance, the information reported in the Demographics subsection in the Method section shows that we do capture the responses of a multitude of ethnicity as opposed to 100% of the responses derived from a single ethnic group. As we have specified in the Method section, none of the questions (including Demographics) were mandatory. Participants who did not wish for us to collect that data could have simply skipped answering (or selected the "Prefer not to answer" option available for many questions). Moreover, the questionnaire collects no personally identifiable information, so the responses are non-identifiable.

At the same time, this indicates that there is indeed cultural variation in the participants sample,

which the authors meant to exclude in the first place.
We meant to exclude the larger cultural variations attributable to nation states, not the smaller cultural variations *within* the United States (which are much smaller than those between two countries, say United States and India). Note further that ethnic background does not necessarily refer to the participant not being from the United States and thus being connected to the larger culture of the United States because the United States is a country that has citizens that can claim many different ancestries going back generations. As we explained above, the ethnicity question was asked to be able to describe the sample and ensure that we captured the views of United States citizens/residents of a variety of ethnic backgrounds.

The authors asked participants "How concerned do you think the person would be if the content of the notification was seen by someone other than you?" (p. 26, lines 1263-1264). This question seemed pretty abstract to me and hard to imagine for participants. Alternatively, the authors could have asked how the participants themselves would feel if their information was revealed through a notification on somebody else's phone. At least, the results of this question should be taken with care.
As we have mentioned in the Method section, the questionnaire was thoroughly pilot tested for language and comprehension. Further, this question was asked not in abstract terms but in connection to a *specific* notification (i.e., the latest notification received on their phone) and only if the participant indicated that the notification contained someone else's information. Asking how the participant's information is revealed on someone else's phone would not be applicable to the concrete context of the latest notification received on the participant's phone.

Moreover, we did not receive any comments or complaints related to this issues from the participants (in the open-ended question at the end that asked them to tell us anything they wished to convey). Based on experience from other studies, we know that AMT participants do use such questions to point out any confusion or issues with study comprehensive or mechanics. So we have reasonable confidence that participants understood the question as intended.

That said, we agree that it would be interesting future work to ask about how someone would feel if their information is revealed to unintended parties because of notifications on someone else's phone.

As the authors also state in their limitations, a real-world evaluation is a relevant subject to future work. For this, I suggest the authors to look into work by Weber, Voit, and Henze [1,2,3].
Thank you for suggesting these pointers. We have incorporated all three references in the Limitations as follows:
Future work should compare these results with real-world analytics of behavioral data and metadata that captures notification content and user interactions with notifications via tools such as Clear All [5] or Notification Log [4]. Further, tools like Annotif [6] can enable the collection of corresponding retrospective user reflection that adds context to the captured logs and facilitates more nuanced interpretation based on a combination of analytics and user responses.

Findings ———— Regarding the analysis and findings, I have the following suggestions:

For the analysis of open-ended questions, the authors should add a reference for their approach – I guess thematic analysis as described by Braun and Clarke would be a good fit [4].
Thank you for the pointer. We moved the description of analysis procedures for open-ended ques-

tions to a "Data Analysis" subsection with the Method section and added the suggested citation which is indeed the correct citation for the approach we followed.

It seems like the analysis was conducted by multiple researchers, and full agreement was reached, which suits the above-mentioned method. However, it seems like the themes were established by only one researcher (i.e., the themes were defined prior to the in-depth analysis by the three coders), which to me seems unusual. The authors should provide more details on how these themes were derived: p.6, line 285 states "by the first author after detailed examination of the responses" – hence, I would argue that this initial analysis step was prone to experimenter bias, which should be acknowledged as a limitation. Moreover, it was unclear to me how ultimately the specific codes for each theme were defined, and how the authors made sure full agreement was reached. It would also be helpful to provide the full coding tree (or otherwise clarify that there were no codes beyond the ones mentioned in Tables 2 and 6).
As mentioned above, we added a subsection called "Data Analysis" to the Method section. In this section, we include a more detailed and accurate description of the coding procedures we followed. The description should alleviate the above concerns.

We apologize for inadvertently using the term 'themes' instead of 'codes' in line 284 of the original submission ("Each coder first independently coded the open-ended responses based a list of themes generated by the first author after detailed examination of the responses.) The correct sentence should have been: " Each coder first independently coded the open-ended responses based a list of codes generated by the first author after detailed examination of the responses."

As we explained, the initial list of code was developed in consultation with the second author and was open to refinement and additions by the other coders during their independent coding pass. In fact, the discussion among coders to resolve discrepancies did yield a more refined set of codes that all coders agreed with. As we mention, full agreement was reached based on synchronous discussion among all three coders.

We did have codes in addition to the ones mentioned in Tables 2 and 6. However, these codes are irrelevant to our research questions, so they have not reported them the paper. For your reference, we are including the coding tree in this response document.

In 4.1.2, the participant's quote (p. 12, line 555) reports on the "notification ring", which does in itself not reveal the notification content. It was unclear to me how this fits the overall scope of the paper, as I assume the focus to be on privacy intrusion by notification content. This should be clarified (or, alternatively, the quote could be removed).
We have clarified in the revision that we approach privacy aspects of notification with a broad view of privacy as conceptualized in the privacy literature. Such a perspective goes beyond simply the content of the notification and includes all aspects of notification delivery and user experience. In this particular case, the disruption caused by the sound is connected to the intrusion facet of privacy. Further, the "notification ring in office meeting" is a breach of intimacy of the social setting. The privacy facet of 'Intimacy,' as taken from one of the four states of privacy defined by Westin [7], refers to "small group seclusion for members to achieve a close, relaxed, frank relationship" [1].

As an additional practical point (not mentioned in the paper as it was not reported by the participants) consider that audible dings or vibrations for notifications received in a public setting can make bystanders and onlookers curious, and it is not uncommon in such situations that people

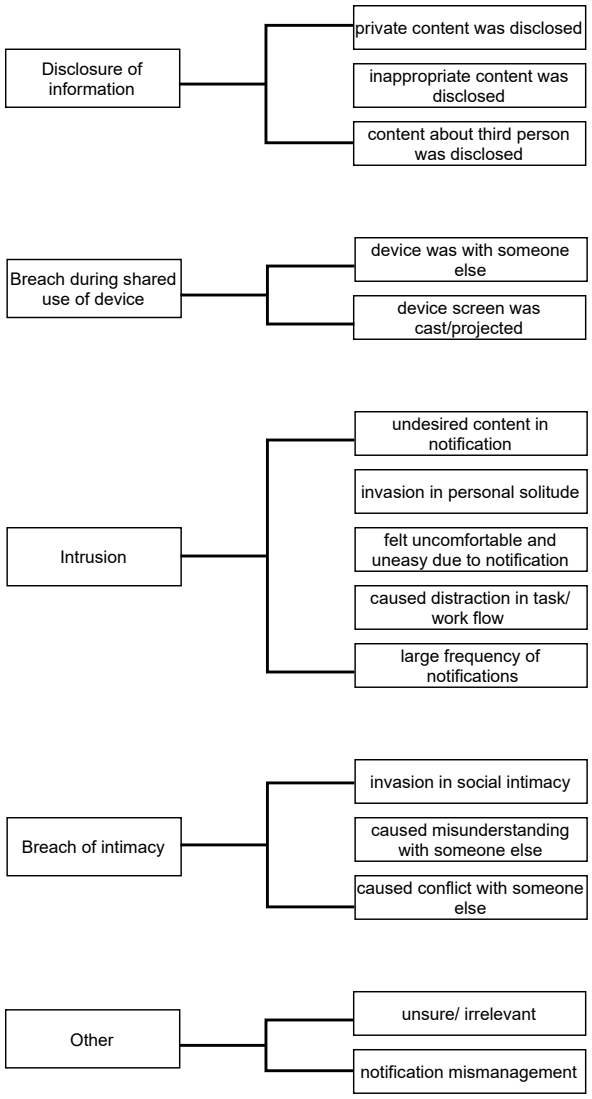

Figure 1: Coding Tree

inquire about the contents or circumstances connected to the notification which the notification receiver might not wish to divulge.

In 4.2.1, the authors report that a large number of negative experiences seems not related to participants' digital proficiency, however, their reason in lines 826-828 seems purely speculative.
Out intuition was that those with less technical proficiency may report large numbers of negative experiences/privacy violations due to notifications because of a lack of efficacy in controlling notification settings. However, this intution is not supported by the data.

However, the reviewer is correct that our explanation is speculative. We have tempered it and moved it as a point of discussion (in the Discussion section) to drive future work.

Furthermore, I suggest separating discussion aspects from the report of findings (e.g., I would see

lines 722-729 and lines 813-815 in the Discussion/Implications rather than in Findings).
We agree. We have separated these and other discussion aspects from the report of findings. We provide these in a newly added Discussion section after the Findings section.

Implications ———— Although I suggest generally separating the report of findings and discussion/implications, the connection between findings and implications can be improved. For instance, Section 5 starts with a "trade-off" (p. 17, line 832) between benefits and privacy, while it is unclear where this comes from.
We have expanded the Implications section to improve the connection between the findings and the implications as well as connection with the literature, where pertinent. We have removed the reference to the trade-off from the beginning of the Implications.

Also, I believe the authors could elaborate more on their valuable suggestions in 5.3 and 5.4, as I see these as one of the core contributions of this paper and an inspiration for future research. For both suggested mechanisms, the authors could discuss concrete design opportunities and how users would benefit from them.
We have expanded these suggestions as requested.

Lastly, I would have loved to see a more in-depth discussion on existing privacy measures for notifications on mobile devices and their effectiveness. Aspects that I suggest to consider, especially for the discussion around disabling notifications in 5.2: - iOS can hide notification content on the lockscreen by default. Only if the phone is unlocked, this content is then shown on the lockscreen [5].
Thank you the suggestion. We have included a discussion of this feature and its limitations for addressing various privacy concerns surfaced by our study.

- Android's "do not disturb" mode has very detailed settings regarding phone calls vs notifications, e.g. silencing everything except phone calls (cf. line 868 – 870) or except marked contacts [6].
Thank you the suggestion. We have included a discussion of this feature and its limitations for addressing various privacy concerns surfaced by our study.

Summary ———— To summarize, this paper presents highly relevant and interesting work on potential privacy intrusion by mobile notifications, and I would love to see this presented at an HCI conference in the near future.
Thank you.

Unfortunately, there is a number of aspects that make me unable to argue for acceptance in the current state of the paper, but need revision: 1) revise focus of related work (Section 2) 2) add details on the method and data analysis procedure (Section 3 and 4) 3) revise report of findings, separate discussion aspects (Section 4) 4) elaborate on concrete designs for suggested mechanisms (Section 5) 5) revisions with regards to language, citations, and presentation (see below)
We have revised all of the above aspects and described the respective revisions in our responses to the corresponding comments.

I hope my review to be helpful to the authors to improve their paper and am looking forward to a revised version.
We greatly appreciate your time and effort in providing us such a detailed review. We believe that addressing your comments has indeed improved the paper, and we hope you agree!

Miscellaneous ———————— - I suggest not using references as part of a sentence, e.g. "These four states are summarised aptly by [28]", p.2, lines 94-95 (rather use the authors' name)
Thank you for catching this inadvertent writing error. We fixed this particular error and ensure that citation numbers are not used like words.

- reference 45, p.3, line 138: the authors' full last name is "von Zezschwitz"
Thank you for spotting this inadvertent error. We fixed the text to include the full last name.

- citation style varies, e.g., p. 3, line 142 "Research by Hang et al.(2012) [21]" additionally provides the year – I suggest to be consistent in style
Thanks for pointing out the inconsistencies in citations. We removed the year in this particular case and have made a pass through the entire paper to make referencing consistent throughout.

- subsection 2.3 only has one subsubsection – I suggest either having (at least) two subsubsections (2.3.1 and 2.3.2) or otherwise remove the subsubsection (e.g., use a

instead)
Thank you for spotting this issue. Upon revising subsection 2.3 for brevity, clarity, and focus, it is not necessary to subsubsection or paragraph.

- line 187 refers to "Supplementary Material", this should be replaced with a (numbered) reference to the "Appendix"
As suggested, we removed the Appendix from the main paper and submitted it as Supplementary Material. Therefore, we have kept the reference to the Supplementary Material unchanged.

- p. 6, Demographics: for small numbers, I suggest to not use percentages (e.g., "0.47%(n = 1)", line 260). Also, most of the demographic data could be presented in a table for a better overview.
Thank you for pointing this out. We have revised the Demographics section for better readbility of the reporting which includes not using percentages for small numbers.

- Also for the report of findings, I suggest using words rather than percentages, e.g. "Half (n=107) of the participants . . . ", p. 12, line 571 We used words where it made sense to do so for increasing the readability as the reviewer suggests.

- Findings should be reported in past rather than present tense (e.g., p.7, line 340 "We found that . . . " instead of "We find that . . . ")
Thank you for pointing out the tense inconsistencies in the reporting of the Findings. We made a pass through the entire paper to ensure that findings are reported in the past tense. However, we use present tense when we mention a (potential) general truth based on our results.

- the layout of plots/figures should be improved (almost all labels in plots seem distorted)
As mentioned above, we have improved readability, presentation, and layout of all figures.

- Table 2 and 6: it should be clear that these are "sample" comments by participants. Also, the layout of both of these tables can be improved.
Thank you for pointing out this issue. We changed the description of this column to "Example Comments."

- participants should be cited with an ID or similar
We added participant IDs to each quote. We would make the edit.

- Table 7 and 8 are far away from the respective text
We have tried to improve the placement of the Figures and Table as much as possible. Ultimately, LaTeX sometimes runs into situations where not all placement constraints can be satisfied and concessions need to be made.

- reference [1] in the bibliography is broken
We fixed the issue with Reference [1].

- language / grammar issues, e.g.
p.3, line 118 "or" instead of the comma
The comma is more appropriate here as this is a list of two separate scales. In any case, we have made a pass through the full paper to identify and fix language/grammar issues. We hope we caught all of them!

* p. 3, line 140 remove "the": "(. . .) the mobile display by bystanders"
Done.

* carefully check the use of "users"' vs. "user's" (e.g., p. 4, lines 162-163)
We have made a thorough editing/proofreading pass through the entire paper to fix this and other writing issues.

* if using "-" as part of a sentence, "–" should be set in LaTeX
We have made a thorough editing/proofreading pass through the entire paper to fix this and other writing issues.

* avoid contractions such as "don't", "doesn't", etc. (use "do not", "does not" instead)
We have made a thorough editing/proofreading pass through the entire paper to fix this and other writing issues.

* I suggest a thorough proofread with regards to punctuation and sentence construction starting from section 2.3.
As mentioned above, we have made a thorough editing/proofreading pass through the entire paper.

References ——————— [1] Dominik Weber, Alexandra Voit, Gisela Kollotzek, and Niels Henze. 2019. Annotif: A System for Annotating Mobile Notifications in User Studies. In Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia (MUM '19). Association for Computing Machinery, New York, NY, USA, Article 24, 1–12. DOI:https://doi.org/10.1145/3365610.3365611

[2] Dominik Weber, Alexandra Voit, and Niels Henze. 2019. Clear All: A Large-Scale Observational Study on Mobile Notification Drawers. In Proceedings of Mensch und Computer 2019 (MuC'19). Association for Computing Machinery, New York, NY, USA, 361–372. DOI:https://doi.org/10.1145/3340764.3

[3] Dominik Weber, Alexandra Voit, and Niels Henze. 2018. Notification Log: An Open-Source Framework for Notification Research on Mobile Devices. In Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing

and Wearable Computers (UbiComp '18). Association for Computing Machinery, New York, NY, USA, 1271–1278. DOI:https://doi.org/10.1145/3267305.3274118

[4] Braun, V., & Clarke, V. (2012). Thematic analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), APA handbooks in psychology®. APA handbook of research methods in psychology, Vol. 2. Research designs: Quantitative, qualitative, neuropsychological, and biological (p. 57–71). American Psychological Association. https://doi.org/10.1037/13620-004

[5] `https://www.macworld.com/article/3261127/ios-11-how-to-hide-sensitive-info-in-notificati` `html`, last accessed March 3, 2021

[6] `https://support.google.com/android/answer/9069335?hl=en#zippy=\%2Cset-what-to-block\` `%2Cset-who-can-interrupt-you`, last accessed March 3, 2021

Thank you for these pointers. We have incorporated all of them in pertinent places within the paper.

Reviewer 2

In this paper, the author(s) explore privacy concerns around notifications in smartphone ecosystems. By conducting a user study (235 participants) they analysed users' preferences and concerns regarding privacy aspects of notifications. The results show that privacy breaches from notifications can happen in different ways. Based on this, the author(s) also provide certain design recommendations aiming at supporting users to better protect their privacy when it comes to device notifications.

Strengths: This is timely topic and well-aligned with MobileHCI. The paper is well-written and easy to follow. Methods and experiments are clearly presented and supported with appropriate literature materials.
Thank you.

Just a couple of minor concerns:
We have addressed each concern below.

- Title: what does the term 'violations' mean? Violations of what?
The word "privacy" in the title refers to concerns as well as violations. We feel that it would be awkward to have a title like: Exploring Privacy Concerns and Privacy Violations...

- Section 1 (Intro): On average smartphone users receive 63.5 notifications per day [39]. How is this reliable? The authors have cited a work published in 2014 and I am not convinced if this is really the case. The authors should try to support their arguments with more recent works (and not only one reference, rather a combination of previous work to better support their arguments).
A more recent study by Pielot et al. [3] reported the median number of smartphone notifications to be 56. We have added this more recent reference and tempered the claim by revising the sentence to read: On average, smartphone users receive more than 50 notifications per day [2, 3].

I don't really understand the relation between the quote (ref 18) and the author's work. Whilst this

quote is more about user experience design, the authors have stated that their work is primarily aimed at reducing the privacy risks associated with notifications. My questions is: how do these two go hand in hand? (the quote from ref. 18 and the authors' contributions). The authors should try to make this clearer.
The reviewer is correct that the quote is not directly connected to our work, so we have removed it (which also helped us shorten the paper as requested in other comments).

- Section 3: I would suggest to rename the title (method of what? Research method?)
The title refers to the Method we used to carry out the research. This is a standard section title used in HCI publications, so we have left it unchanged. However, in the signposting paragraph at the end of the Introduction section, we have clarified that the Method section describes the procedures we used to collect and analyze data.

Language: Recheck grammar: . . . reported least one . . . – . . . reported at least one . . .
Thank you for spotting this error. We fixed it.

Reference 1 is not displayed properly.
As mentioned above, we fixed the issue with Reference 1.

References

[1] S. T. Margulis. Three theories of privacy: An overview. *Privacy online*, pages 9–17, 2011.

[2] M. Pielot, K. Church, and R. de Oliveira. An in-situ study of mobile phone notifications. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services*, MobileHCI '14, page 233–242, New York, NY, USA, 2014. Association for Computing Machinery.

[3] M. Pielot, A. Vradi, and S. Park. Dismissed! a detailed exploration of how mobile phone users handle push notifications. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '18, New York, NY, USA, 2018. Association for Computing Machinery.

[4] D. Weber, A. Voit, and N. Henze. Notification log: An open-source framework for notification research on mobile devices. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, UbiComp '18, pages 1271–1278, New York, NY, USA, 2018. Association for Computing Machinery.

[5] D. Weber, A. Voit, and N. Henze. Clear All: A large-scale observational study on mobile notification drawers. In *Proceedings of Mensch Und Computer 2019*, MuC '19, pages 361–372, New York, NY, USA, 2019. Association for Computing Machinery.

[6] D. Weber, A. Voit, G. Kollotzek, and N. Henze. Annotif: A system for annotating mobile notifcations in user studies. In *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia*, MUM '19, New York, NY, USA, 2019. Association for Computing Machinery.

[7] A. F. Westin. Privacy and freedom. *Washington and Lee Law Review*, 25(1):166, 1968.