A Proposed Evaluation of *Just Not Sorry*, a Technology to Influence Language Use

Janet Davis[1]

[1] Department of Computer Science Whitman College, Walla Walla, WA 99362, USA
davisj@whitman.edu

Abstract. This poster will present plans for an empirical evaluation of *Just Not Sorry*, a Gmail plug-in designed for women in business and leadership. As the user composes email, *Just Not Sorry* highlights words that could undermine an image of confidence or authority. The proposed study seeks not only to evaluate the effectiveness of *Just Not Sorry* as a persuasive technology, but also to characterize reasons and values un- derlying decisions about adoption.

1 Background

Just Not Sorry [3] is a Gmail plug-in that runs while the user is composing an email. It highlights words of apology such as "sorry," hedge words such as "just," and intensifiers such as "very." Red underlines appear as if the words had been misspelled. When the user mouses over an underline, a motivational quote appears as a tooltip. For example: "Using 'sorry' frequently undermines your gravitas and makes you appear unfit for leadership - Sylvia Ann Hewlett." The intention behind *Just Not Sorry* is to support women in business who want to reduce their use of such words and appear more confident in their emails [6].

While it is clearly a behavior change support system, *Just Not Sorry* is not a product of academic persuasive system design. In 2015, tech industry CEO Tami Reiss wrote about her motivations for proposing the tool and guiding its development [6]. The tool gained broad media coverage—some supportive, some critical—and came to my attention as part of a survey of existing tools designed to influence language use [9]. At its peak, Reiss claimed hundreds of thousands of users [8], while the Chrome Web Store currently reports tens of thousands [3].

2 Goals and Research Questions

The goal of this poster presentation will be to obtain feedback on a proposed empirical evaluation of *Just Not Sorry*. Through this study, I seek to address two main research questions:

RQ1. What reasons and values inform decisions (not) to adopt *Just Not Sorry* ?
RQ2. Once adopted, does *Just Not Sorry* change behaviors and attitudes?

Why study adoption? At a surface level, *Just Not Sorry* is a boring technology [10], an email plug-in inspired by the "spellcheck" metaphor [6]. It is simple to understand and adopt, and fits effortlessly into everyday habits. At another level, the behavior change that *Just Not Sorry* seeks to promote is surprisingly controversial. A recent media analysis investigates published opinions about *Just Not Sorry* with respect to underlying values and perspectives on feminism [2]. An empirical study would investigate whether reasons for deciding against adoption are similar to reasons for opposition expressed in the media.

Why study effectiveness? First, Reiss claims that *Just Not Sorry* is effective at influencing both behaviors and attitudes [6], but presents only anecdotal evidence to support this claim. Second, while *Just Not Sorry* achieved a significant user base, it has shrunk over time by a factor of ten. This fact could be interpreted as evidence of either failure of long-term adoption, or of success at changing attitudes or behaviors making further use unnecessary. Finally, to the best of my knowledge, no work presented in the PERSUASIVE conference series has addressed the effectiveness of persuasive technology at changing habits of language use. Beyond intrinsic interest, the proposed study may inform future evaluation studies of tools now under development, e.g., to mitigate gender bias in academic recommendation letters [7].

3 Investigating reasons and values

Our recent media analysis [2] suggests some possible answers to RQ1, but there is no reason to think that published opinions about *Just Not Sorry* are either exhaustive or representative of a broader public. Therefore, a study addressing RQ1 should support both quantitative analysis to assess the representativeness of the published opinions, and qualitative analysis to uncover values and reasons that were left out. The choice to use surveys, interviews, or a combination of both will depend on decisions about the need to reach a large number of participants versus the ability to probe their answers.

The media analysis suggests possible reasons for adopting *Just Not Sorry* that we would seek to confirm:

– to enhance one's confidence, achievement, or social power;
– to become more mindful of one's habitual use of apologies, hedge words, and intensifiers, and hence to use such language more intentionally or sincerely;
– to promote the status of women in the workplace and thus gender equality.
The media analysis also discovered reasons for opposing the adoption of *Just Not Sorry* :

– concerns that Just *Not Sorry* will undermine rather than enhance confidence, achievement, or social power;
– concerns that *Just Not Sorry* is insensitive to linguistic or situational context;
– valuing apologies, hedge words, and intensifiers as women's language or as a form of politeness;

– a belief that *Just Not Sorry* is a distraction from more fundamental problems of gender inequality.

We can also imagine more personal reasons for choosing not to adopt *Just Not Sorry* that may not have appeared in the media as reasons to recommend against its use, such as

– satisfaction with one's current behavior;
– annoyance at receiving feedback while writing;
– reliance on an email system other than GMail or a web browser other than Chrome.

Just Not Sorry 's support for users' autonomy and privacy were discussed as positive attributes in the media opinions, and we would be surprised to see these cited as reasons to choose not to adopt the tool. While the operation of *Just Not Sorry* is generally transparent, currently there is no end-user documentation of the phrases it flags [4], which might be a deterrent for some.

To add nuance to discussion of gender equality and tools to support changing language use, an interview might invite participants to compare *Just Not Sorry* to a tool such as the #GenderBias Slack plug-in [1], which also employs a spellcheck metaphor but flags biased terms used to describe other people.

4   Investigating effectiveness

Hypotheses regarding RQ2 are informed by the "Change" dimension of the O/C matrix [5], focusing on interactions between compliance, behavior change, and attitude change. Specifically:

H1. The number of suggestions per email per user will decline over time, as users change their behavior to avoid proscribed words and phrases.

H2. Compliance will never reach 100%, as users will make deliberate choices not to comply with suggestions they deem inappropriate.

H3. Compliance will decline over time. As a user begins to avoid proscribed words and phrases, an increasing portion of the suggestions will be deemed inappropriate.

H4. Some users will stop using the extension because their behavior has changed and suggestions are no longer deemed valuable.

H5. Users will develop a more confident attitude with respect to the tone of their emails.

H6. Users will report that they avoid using target words and phrases in other speech and writing, beyond email.

To address hypotheses 1-3, we will develop an instrumented version of *Just Not Sorry* that tracks, for each user, the suggestions offered by the system, which the user complies with, and which they do not. Participants who indicate a willingness to use *Just Not Sorry* in the initial interview about adoption will

be invited to install this instrumented version and use it for an extended period of time, likely 4–6 weeks.

The collected data will also reveal when users stop using *Just Not Sorry,* which addresses part of hypothesis 4. Reasons for discontinuing use, as well as attitude change, will be addressed through a post-intervention survey or interview.

5 Acknowledgements

References

1. Catalyst: #BiasCorrect Slack Plugin (2020), https://www.catalyst.org/biascorrect/
2. Davis, J., Nyatichi, B.: Values and Politics of a Behavior Change Support System. In: 18th International Conference on the Ethical and Social Impacts of ICT. La Rioja, Spain (Jun 2020), meeting Name: ETHICOMP 2020 Reporter: 18th Inter- national Conference on the Ethical and Social Impacts of ICT
3. Def Method: Just Not Sorry – the Gmail Plug-in (v1.6.0) (Mar 2019), https://chrome.google.com/webstore/detail/just-not-sorry-the- gmail/fmegmibednnlgojep-midhlhpjbppmlci?hl=en-US, reporter: Chrome Web Store
4. JimLynchCodes: Document Phrases The Algo Looks For In README? · Issue #57 defmethodinc/just-not-sorry (Jan 2020), https://github.com/defmethodinc/just- not-sorry/issues/57
5. Oinas-Kukkonen, H.: A foundation for the study of behavior change support sys- tems. Personal and Ubiquitous Computing 17 (Aug 2013), reporter: Personal and Ubiquitous Computing
6. Reiss, T.: Just Not Sorry! (the backstory) (Dec 2015), https://medium.com/@tamireiss/just-not-sorry-the-backstory-33f54b30fe48, reporter: The Medium
7. Saderi, D.: This open-source tool shines a light on gender bias (May 2018), https://medium.com/read-write-participate/this-open-source-tool-shines- a-light-on-gender-bias-4a51bb5ff4ac
8. @tamireiss: "@alicegoldfuss I made an app that 350k ppl use #justnotsorry #WITBrag-Day https://t.co/a3pyq9ujp3" (Aug 2017), https://twitter.com/tamireiss/sta-tus/896384586833432576, reporter: Twitter
9. Twersky, E., Davis, J.: "Don't Say That!". In: Persuasive Technology. pp. 215–226. Lecture Notes in Computer Science, Springer International Publishing (2017), reporter: Persuasive Technology
10. Wai, C., Mortensen, P.: Persuasive Technologies Should Be Boring. In: Persua- sive Technology. pp. 96–99. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg (2007). https://doi.org/10.1007/978-3-540-77006-012