Blame the Scapegoat: How Creating an Easy Target Can Help Non-Designers Critique Designs

Bradley Reyes
Stanford University
Stanford, CA 94305, USA
breyes28@stanford.edu

Tamara Prstic
Stanford University
Stanford, CA 94305, USA
tprstic@stanford.edu

Ramon Flores
Stanford University
Stanford, CA 94305, USA
rrflores@stanford.edu

CHI'17 Extended Abstracts, May 06-11, 2017, Denver, CO, USA
ACM 978-1-4503-4656-6/17/05.
http://dx.doi.org/10.1145/3027063.3053184

Abstract
We present a study testing whether adding a scapegoat design (one that is clearly stylistically worse than the other designs and thus easier to critique) will make users feel more comfortable with giving more feedback on the other designs that the experimenters want feedback on. We found that participants who are shown a scapegoat design in parallel with other designs did not only provide more critiques, but also more substantial critiques - resulting in both higher quality and quantity of user feedback. Our results imply that the psychology of scapegoating can be successfully adapted to help the HCI community improve the effectiveness of user testing.

Author Keywords
Design; Usability Testing; UX Research; Design Methods; User Centered Design; Scapegoat; Parallel Prototyping

ACM Classification Keywords
H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous

Introduction
The concept of scapegoating in psychological terms has been well documented, and it is often used as a defense mechanism to eliminate negative feelings about oneself and motivate someone to raise his or her own status [4]. Un-

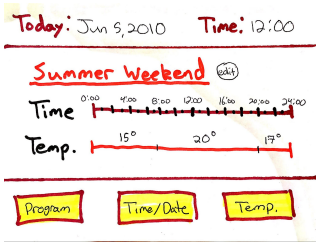

Figure 1: The Linear Paper Prototype

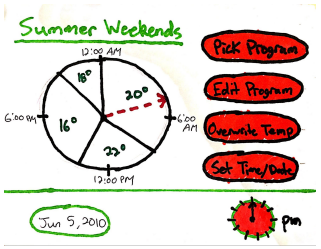

Figure 2: The Circular Paper Prototype

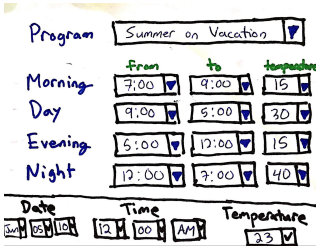

Figure 3: The Tabular Paper Prototype

derstood as the process of taking the blame off oneself by attributing it to somebody else, scapegoating can evoke a range of emotions in individuals who engage in it: by allowing them to eliminate negative feelings from themselves by attacking others, scapegoating often gives the perpetrator a sense of gratification and empowerment [6]. Furthermore, scapegoating often helps insecure individuals gain confidence by raising their own status [1].

In the context of usability testing, we believe the concept of scapegoating can be especially beneficial to non-designers when asked to provide feedback on multiple designs. Due to a lack of experience or technical background, non-designers often provide little or unsubstantial feedback. These psychological insights lead us to hypothesize that introducing an easy critique target – a scapegoat design – could help non-designers overcome their insecurities; heavily critiquing the scapegoat first would elevate their status, leading them to provide more and better feedback on other designs. This increase in feedback would ultimately help the designer improve their product or make a final design decision. Developing a scapegoat might vary across branches of design. In this work, we focused on digital design and thus strongly relied on Jakob Nielsen's usability heuristics for interface design as an indicator of quality [5].

When developing our study, we adopted a number of experimental techniques described in Bill Buxton's seminal work [7]. Buxton's experiment demonstrates that quality design feedback is best elicited by presenting three designs in parallel rather than presenting each design individually. We take Buxton's findings a step further by delving into the efficacy of replacing a parallel design with a scapegoat design with respect to gathering quality feedback. To maintain consistency, we designed similar prototypes and developed a similar procedure to Buxton. Furthermore, inspired by an intervention technique proposed in Elder and Zhou [2] that sought to test how interventions during the design process might help produce better results, we set out to test whether scapegoating can be manipulated to help the HCI community make the most of user tests. Novice critiquers are often reluctant to offer quality feedback, so we hope to determine whether or not the scapegoat design elevates their confidence and gives them an easy target to begin critiquing. The testing of these ideas is part of a larger agenda of gaining a better understanding of user psychology and how we might adapt it to improve the design process.

Research Questions

Our approach was to design and conduct an experiment where participants perform a design critique on three stylistically comparable designs (control) or performed the same procedure on one scapegoat and two of the three comparable designs (experiment). We hypothesized that being shown a scapegoat design would impact user feedback in the following ways:

H1: Participants will offer more overall feedback when seeing the scapegoat design first.

H2: Participants will offer more actionable feedback when seeing the scapegoat design first.

H3: Participants will offer more substantial suggestions when seeing the scapegoat design first.

Methodology

The study was a between-subjects design. Different groups of participants were presented with either three "good" prototypes (Figures 1-3) or a combination of two "good" designs and one scapegoat (Figure 4). Here, our goal was to assess the impact of these different conditions on the number of positive and negative user comments, as well as the number of substantial and superficial suggestions. We studied a total of 30 participants, all Stanford students,

and ensured that they had little to no experience critiquing designs.

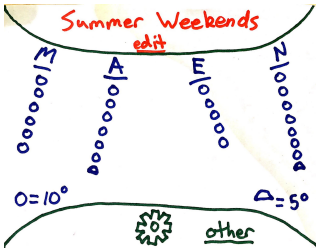

Figure 4: The Scapegoat Paper Prototype

Paper Prototypes

We tested four prototype designs total, all of which were of house climate control systems. The first prototype was the scapegoat. The other three were ”main” designs, which we previously labeled as ”good” designs. These were adopted from the Buxton paper. After pilot testing, we had to revise the circular display to improve its quality and add a line to the script reminding the user that we are testing the UI, not the user, as they often got discouraged after completing tasks on the scapegoat. Functionally, all four prototypes were equivalent. We then pilot tested 10 participants to see whether the main three prototypes were comparable and that the scapegoat was significantly worse than those three. We asked them to rate all four prototypes on their appeal and functionality on scale of 1-9. The three main designs got an average rating of 6.7, 6.8, and 6.9, making all of them comparable in quality. We tested out three different scapegoats and our final scapegoat received an average rating of 2.9 – significantly worse than the three main designs.

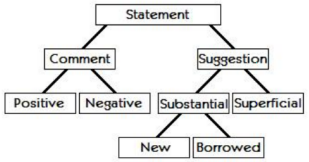

Figure 5: Feedback Statement Classification Tree

Creating the Scapegoat Design

When intentionally developing a stylistically low-quality design, we cared about both aesthetic and functionality. Aesthetically, we made the layout unappealing and unclear. The wording and icons were made inconsistent. Functionality wise, we ensured interaction was unintuitive while ensuring all tasks could be fully executed (and thus did not opt for the lowest scoring scapegoat which did not fulfill this condition). We ensured the scapegoat was not blatantly different from main designs. A useful reference was Nielsen’s usability heuristics [5] which we intentionally violated. Testing multiple options, maintaining visual consistency across all designs and ensuring functionality at all costs finally resulted in an effective scapegoat design.

Procedure

We followed a script in which we explained what a house climate control system was and asked participants to interact with the prototypes thinking out loud. Interaction included performing four tasks that included manipulating temperature, date, and time settings. One team member served as a ”Computer” who simulated interaction. We also emphasized that we were testing the UI, not the users, to remove discomfort. In the experiment condition, we made sure that the scapegoat was presented first. It was also important to note that the designs were not our own so as to remove the risk of users giving falsely positive critiques.

Post-Experiment Interview

After completing the tasks, participants were asked to list out what they liked or disliked about each of the designs in parallel. We then picked one of the three designs and asked about what the user would change if we were to move forward with that design as our final one. Lastly, we asked for any additional comments or feedback.

Experimental Conditions

Participants were randomly assigned either to the control condition (without scapegoat) or experiment condition (with scapegoat). In the control condition, participants were presented with the three main prototypes (Figures 1-3), with no mention of other existing designs. In the experiment condition, each participant was presented with two of the three main prototypes and one scapegoat prototype (Figure 4). In this condition, the scapegoat was always presented to the participant first. In both conditions, we randomized which and in what order the main designs were presented in. We wanted to ensure that the specific designs or order of pre-

sentation did not significantly affect our results. Thirty participants took part in all: fifteen in the control condition and fifteen in the experiment condition. During the experiment, we recorded audio and took notes on the participants' comments.

Analysis of Comments and Suggestions
We adopted Buxton's methodology for classifying feedback statements [7]: "Thus, we identified two broad classes of statements: participants either made 'comments' (facts or personal opinions), or provided 'suggestions' for change to improve the current design. In turn, we found the comments could be classified as either 'positive' (e.g. 'I like the way it guides you through each step of the process'), or as 'negative' (e.g. 'this is too cluttered'). Suggestions, on the other hand, were found to be either 'substantial' (e.g. 'it would be nice to allow for more than 4 intervals per day') or 'superficial' (e.g. 'the colors are dull'). In terms of the substantial suggestions, we further classified them as ideas for improvement which were original or 'new' (e.g. 'it would be nice to have both Celsius and Fahrenheit') or as 'borrowed' from ideas they had seen in other interfaces (e.g. 'show the current temperature like in the Tabular interface').

Results

To assess the impact of showing two main designs with a scapegoat as opposed to three main designs, we observed the effects on amount of feedback (H1), amount of actionable feedback (H2), and the number and type of suggestions for design change (H3). An unpaired t-test was used in all of the hypotheses to test for significance.

We also performed a two-way ANOVA with factors of group and permutation of prototypes to do a sanity check on our pilot test's conclusion that the three prototypes were of comparable quality. Our results were as expected; the change in amount of feedback was attributable to the presence of the scapegoat and not the designs of the prototypes themselves. Our pilot test was fairly accurate in saying the three prototypes were of similar quality.

	Total Feedback Results		
	Control	*Experiment*	*Avg*
Linear	4.4	11.4	7.2
Circular	5.9	10.5	7.8
Tabular	3.9	10.5	6.5
Scapegoat	N/A	11.4	11.4
Avg	4.7	10.8	7.2

Table 1: Average number of feedback statements by type of prototype (linear, circular, tabular, or scapegoat) or type of group (experiment or control). Avg. means the average.

Impact of Showing Scapegoat Designs on Amount of Participant's Feedback(H1)
For H1, we measured the total amount of feedback for each design in both groups. This includes all statements, including both comments and suggestions. We then averaged out the number of total statements, because the control group has three meaningful prototypes and the experiment group only has two. A t-test was carried out on feedback statements, and $p < 0.0001$. Table 1 shows the average number of statements for each prototype for each experiment group. The experiment group received more statements than the control group for every single one of the prototypes. Calculations omitted the scapegoat design. Overall, the results show a strong support for hypothesis H1.

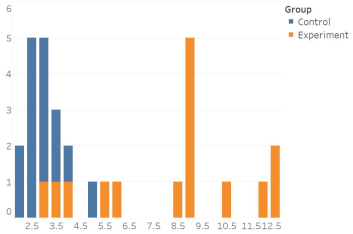

Figure 7: Histogram of average number of actionable statements given per prototype. Blue bars represent the control group, orange bars represent the experiment group.

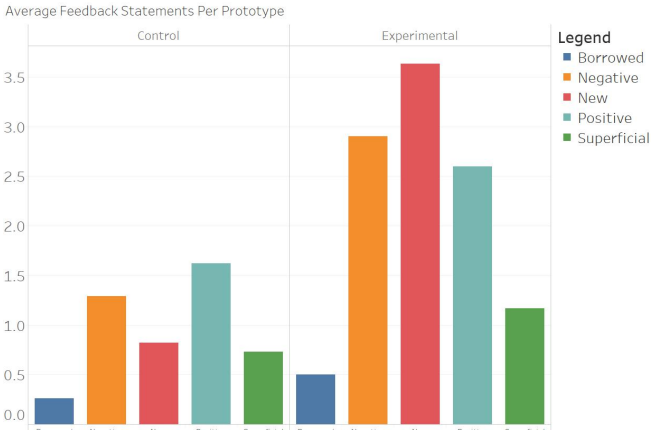

Figure 6: Average comments per prototype for each type of feedback statement (see Figure 5 for types of feedback statements)

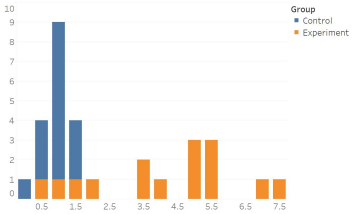

Figure 8: Histogram of average number of substantial suggestions given per prototype. Blue bars represent the control group, orange bars represent the experiment group.

Impact of showing scapegoat designs on amount of participant's actionable feedback (H2)

For H2, we measured the total amount of actionable feedback per main prototype in both groups. Based off of Figure 5, actionable feedback accounts for everything that is not a positive comment. This includes negative comments and all suggestions. Our rationale for this is that both negative comments and suggestions provide actionable insight to the designer. We believe that this actionable feedback is relatively more helpful than positive comments because they provide a specific pain point upon which to act on. A t-test was carried out on the number of actionable feedback given and it has a $p < 0.0001$. As seen in Figure 7, there is a clear ceiling for the amount of feedback received by the control group. Additionally, despite a slight overlap between the two groups, the experiment group clearly has a much larger number of feedback statements than the control group. Overall, the results show a strong support for hypothesis H2.

Impact of showing scapegoat designs on amount of participant's substantial suggestions (H3)

For H3, we measured the total amount of substantial suggestions for each design in both groups. Based off of Figure 5, substantial feedback pertains to both new and borrowed substantial suggestions. We then averaged out the number of main prototypes critiqued. We believe that substantial suggestions are the most useful of all the statements our participants could make. They offer an insight into pain points and provide specific suggestions to explore. A t-test was carried out on the number of substantial suggestions and it has a p value of < 0.0001. As seen in Figure 8, there is also a clear ceiling for the amount of feedback received by the control group. Despite a slight overlap between the two groups, the experiment group clearly has a much larger number of substantial suggestions than the control group. Overall, the results show a strong support for hypothesis H3. Overall, only superficial suggestions, borrowed suggestions, and positive comments had p-values above 0.0001. Superficial suggestions and positive comments had p-values less than 0.05 (still making them significant) and borrowed suggestions had a p-value greater than 0.05.

Discussion

Our studies have shown that the addition of a scapegoat in critiquing parallel prototypes increases the amount of feedback amongst almost all types of feedback, most prominently to actionable feedback and substantial suggestions. While conducting our experiments, we noticed several trends that hint at the source of the scapegoat's success that warrant further investigation and discussion. We be-

lieve the positive effects of a scapegoat prototype is primarily attributable to three main reasons.

First, scapegoating can help those not experienced with critiquing designs to boost confidence and feel more comfortable with giving critiques. This was already discussed in-depth in the introduction section of this paper.

Secondly, scapegoating may have given the participant's some useful practice in noticing. There is already some research in how experts get novices to notice[3]. People do not necessarily notice features or errors, but as they develop this expertise, they can begin to start noticing. In this experiment, the scapegoat may have given the user real-world practice in noticing by purposefully showing them the a scapegoat first, allowing them to give more and better feedback when a target that was not intentionally created to be easy to critique is displayed. Likewise, these contrasting cases between the scapegoat and the main designs may have led to users having an easier time noticing what is and what is not important.

Lastly, we believe that the scapegoat may have set a high expectation for the amount and quality of feedback that was expected from the remaining two prototypes. From Table 1, we see that the scapegoat elicited the most feedback, with 11.4 statements. With this expectation, these participants expect themselves to provide a similar amount of feedback for the remaining two prototypes. We see this in Table 1, which shows all of the prototypes in the experiment group eliciting a similar amount of feedback. In other words, not all of the participants' energy went into attacking the scapegoat: if anything, it seemed to fuel more energy into critiquing the remaining two designs. Anecdotally, we saw that participants in the experiment group would examine the main prototypes for a longer period of time before moving on, possibly attempting to give more responses to match the scapegoat. This was not observed in the control group.

Overall, the scapegoat encouraged participants to talk, which accounts for why certain statements did not decrease when another increased. In other words, getting one type of statement did not compromise getting others, as all types of statements except for borrowed increased.

Conclusion

The most important contribution of this research is in its implications regarding usability testing. Usability testing is an essential part of any user-centered design process, and being able to perform it efficiently across a wide range of users is essential for improving the overall quality of the final design. Professionals across a wide number of design disciplines must acknowledge that not all users have impressive technical backgrounds. This work aims to help both designers and these users by ensuring they can still provide high quality feedback: this way, no user subgroup will be excluded from the design process. Furthermore, we feel that having this scapegoat setup is beneficial since it gives the user real-world practice that is built into the experiment, making benefits like the three listed in the discussion section happen.

A logical follow-up experiment would replicate this one, but with experiment groups of two prototypes rather than three. This might be useful since designers can quickly test their single design, but still get the advantages of parallel prototyping without putting much effort into making another design of similar quality. Additionally, further experiments looking into the efficacy of scapegoats in different situations would warrant further investigation, such as different prototype fidelities or participants with different demographics.

References

1. Carter, C. A. (1996). *Kenneth Burke and the scapegoat process*. University of Oklahoma Press.
2. Elder, A. N., & Zhou, E. (2013, October). Brainstorm, define, prototype: timing constraints to balance appropriate and novel design. In *Proceedings of the adjunct publication of the 26th annual ACM symposium on User interface software and technology* (pp. 105-106). ACM.
3. Goodwin, C. (1994). Professional Vision. *American Anthropologist*, 96(3), 606-633. http://doi.org/10.1525/aa.1994.96.3.02a00100
4. Landes, R. (1994). Scapegoating. *Encyclopedia of Social History. New York: Garland Pub. Inc.*
5. Nielsen, Jakob. "10 usability heuristics for user interface design." *Fremont: Nielsen Norman Group.[Consult. 20 maio 2014]. Disponivel na Internet*(1995).
6. Scheidlinger, S. (1982). Presidential address: On scapegoating in group psychotherapy. *International Journal of Group Psychotherapy*, 32(2), 131-143.
7. Tohidi, M., Buxton, W., Baecker, R., & Sellen, A. (2006, April). Getting the right design and the design right. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 1243-1252). ACM