Study Report: Carbon Busters, a New Interface for CO2 Compensation

Yi-Tau Ly, Mert Çökmez, Vikram Thanigaivelan, George Selim, Raffael Kunz (Group 16)
ETH Zürich, Human Computer Interaction

1 Introduction

Greenhouse gas effect has been a rising topic in the last decade and remains important. There are many activities that result in increased CO2 emissions, which include but are not restricted to commuting, training machine learning algorithms, streaming, shopping and so on. For this project, we decided to focus on the emissions caused by online shopping, since consumerism is one of the main driving forces behind increased greenhouse effect.

We predict that by visualizing the user's contribution to CO2 emission caused by the purchases, we will be able to help users understand their impact on the environment and motivate them to be more mindful about their habits. Information provided by e-commerce websites about the emissions caused by shopping are normally sparse; we want to improve on that with our project.

Most online shops only provide a simple checkbox as an option for compensating your purchase emissions. However, this is not very motivating for shoppers since most people tend to see CO2 compensation as just paying more for the same. To improve on this, we decided to develop a plugin for online shops, which should motivate and inform users about compensation while keeping it as simple and quick as possible and in the interest of web shops.

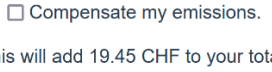

Figure 1: *Existing checkbox option in e-commerce websites*

We came up with the idea of "Carbon Busters" which is inspired by the 1984 hit movie "Ghostbusters".

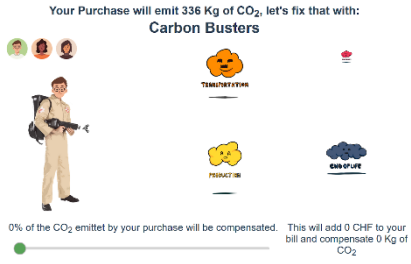

Figure 3: The plugin implemented in checkout pages

The plugin consists of an input method that allows for a selection of compensation between 0% and 125%, clouds representing the relative emission factors of transportation, material, production and end of life emissions of your purchase and a character selection option and some engaging animations.

We tested different input options for our plugin. The first input method we used was a continuous slider. The slider ranged from 0% compensation until 125% of the total CO2 emissions of the shopping cart. The second option was a checkbox layout with six options, which gave the user the choice to compensate 0%, 25%, 50%, 75%, 100% or 125% of the total emissions. And in the third option, we provided a value input box, also ranging between 0% and 125%.

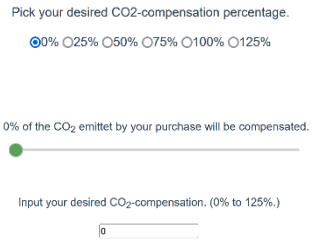

Figure 2: *different input methods of compensation in our plugin (just replacing the slider in Fig. 2).*

2 Study Design

The tests were conducted on a laptop running the hi-fi prototype, simulating a web shop and featuring our plugin or the traditional checkbox.

The independent variables for this test were the method of compensation input, („and" removed) the provided, randomly generated budget, the cart and the emissions in the four sectors.

On the other hand, the dependent variables were the percentage of CO2 compensated, the amount of information retained about emissions, the system usability, the user satisfaction, the number of clicks and the task completion time.

2.1 Null hypothesis:

Test 1: Plugin vs traditional checkbox

- **H1.1** There will be no effect of using our Carbon Busters plugin in comparison to using the traditional checkbox system on the percentage of CO2 compensated by the user.
- **H1.2** There will be no effect of using our Carbon Busters plugin in comparison to using the traditional checkbox system on the CO2-compensation-information taken up by the user.

- **H1.3** There will be no effect of using our Carbon Busters plugin in comparison to using the traditional checkbox system on user satisfaction.
- **H1.4** There will be no effect of using our Carbon Busters plugin in comparison to using the traditional checkbox system on the time spent by the user to compensate.

Test 2: Different inputs to Carbon Buster plugin

- **H2.1** There will be no effect when changing the input methods between a slider, checkboxes or direct value input on the effectiveness (time) of the user in choosing the CO2 to compensate.
- **H2.2** There will be no effect when changing the input methods between a slider, checkboxes or direct value input on the error rate (number of clicks) of the user in choosing the CO2 to compensate.
- **H2.3** There will be no effect when changing the input methods between a slider, checkboxes or direct value input on intuitiveness.
- **H2.4** There will be no effect when changing the input methods between slider, checkboxes or direct value input on the percentage compensated.

2.2 Testing Protocol / Procedure:

First, the participants were asked to fill in a preliminary survey to better approximate their profiles. Then, they were asked to do two tasks which corresponded to our two AB tests.

In the first task, participants were given a budget (enough to compensate between 75 to 125%) and were asked to proceed with the checkout. In addition, we told them that we would provide them with chocolate, equal to the saved budget and compensate for real CO2 equivalent to the compensation chosen. This was repeated five times for each input method in random order. The participants were given two questionnaires on the last two repetitions to assess the retained information.

In the second task, participants were once more given a budget, but this time between 0% and 125%. The task now was to compensate as much as possible in a timely manner. We randomly provided the participants with one of the three input methods for our plugin.

At the end of testing, the participants were asked to fill out a final questionnaire.

2.3 Demographics:

Our participant population of 16 testers was aged between 20 and 63 years, with a median age of 24 years and an average age of 28.06 years. The testing population was almost completely (15 of 16) from academic background.

RESULTS

We will be using a paired Wilcoxon signed rank test to compare the input methods. This is because we are using the same participant group for all tests, which means that they are not independent and that the distributions for the data for the input methods are not normal based upon the Shapiro-Wilk normality test (p-value < 0.05 for distributions tested)

Test 1: Plugin vs traditional checkbox

H1.1:

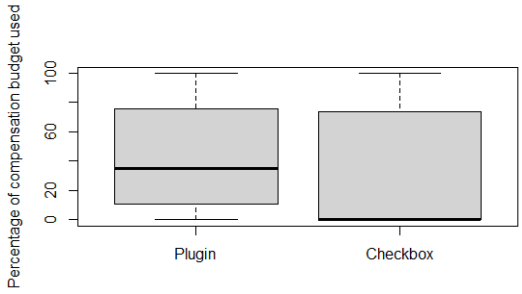

Figure 4: Compensation Budget Percentage used Comparison

To find out if our plugin motivates users to compensate more, we compared the percentage of the budget used to compensate using the checkbox and plugin interfaces. We conducted a Wilcoxon signed rank test. With the traditional checkbox input method, the average budget percentage was 28.4% (SD =39.3%) and for the plugin, the average was 41.3% (SD =34.2%). The mean difference between the plugin and the checkboxes was statistically significant: V = 1823, p-value = 7.227e-05. This indicates that using our plugin, the users ended up compensating more than using the traditional checkbox.

H1.2:

To find out if our plugin motivates users to find out more about their emissions, we looked at two metrics of our test results: the first being the error between the actual compensated emissions and the reported compensated emissions by the user in the questionnaire. We conducted a Wilcoxon signed rank test. With the traditional checkbox input method, the average error was 9.4% (SD =27.2%) and for the plugin, the average was 18.5% (SD =27.3%). The mean difference between the plugin and the checkboxes was statistically significant: V = 50, p-value = 0.1424. Those results indicate that the difference between the error percentage of the two interfaces was not statistically significant. The second metric was a question about the biggest emitter category, these results are presented in Fig.4 in the Appendix.

H1.3:

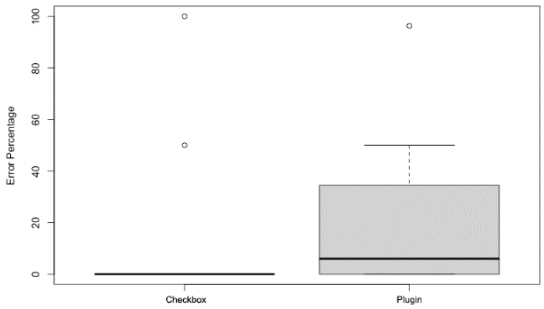

Figure 5: Error Percentage between real and reported

After the completion of the first test, the users were asked to choose a preferred interface between the two. Eight users chose the Carbon Busters plugin while six preferred the traditional checkbox.

We also offered the option to give a reason for the

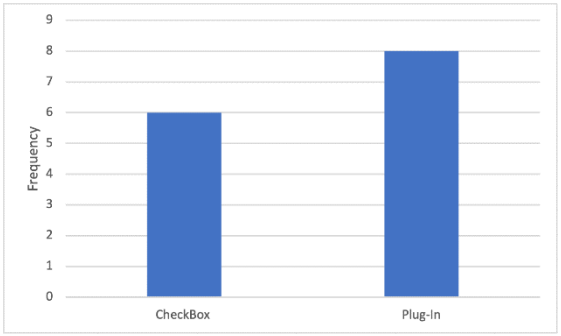

Figure 6: Histogram of user preference

preferred compensation method at the end of the questionnaire. The most common answer for the plugin preferring testers was along the lines of "allows to choose the amount," for the one preferring the checkbox it was usually the higher simplicity compared to the plugin.

H1.4:

To compare the effect of the input method on the

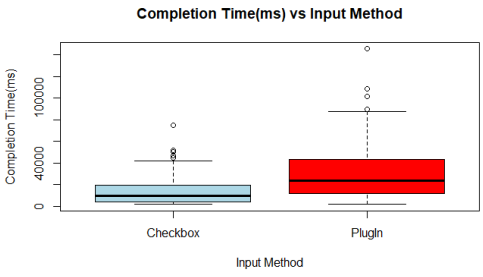

Figure 7: Completion time test 1

completion time, we conducted a Wilcoxon signed rank test. With the traditional checkbox input method, the average time was 15001 ms (SD = 14303) and for the plugin, the average was 33199 ms (SD = 27932.07). The mean difference between the plugin and the checkboxes was statistically significant: V = 197, p-value = 8.933e-12. These results indicate that the input method plugin was slower than the traditional checkboxes. The H0 is rejected. This was probably because the checkboxes are much simpler to use and have less options so it took less time to pick a compensation there, that would maximize the budget. This also made the checkboxes simpler to use and therefore the task was completed faster.

Test 2: Different inputs to Carbon Buster plugin

H2.1:

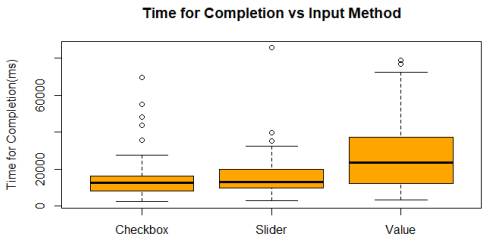

Figure 8: Completion times for test 2

To compare the effect of the input method on the completion time, we conducted a Wilcoxon signed rank test. With the input method checkbox, the average time was 14476 ms (SD = 11052), for slider, the average was 14476 ms (SD = 11019), and for value input, the average was 25984 ms (SD=17069). The mean difference between the slider and the checkboxes were statistically significant: V = 1149 and p-value = 0.02403. Also, the mean difference between the checkboxes and the value was statistically significant: V = 426, p-value = 1.038e-08. Lastly, we also checked the difference between the slider and the value input and found that this was also statistically significant: V = 444, p-value = 1.72e-08. These results indicate that the input method checkboxes was the fastest, next was the slider, and last was the value input. Thus, H0 is rejected. This difference is most likely because it was difficult to choose a value that gave the participants a high compensation while getting close to the total budget with the value input, whereas with the slider, it was easier to change values, and the checkboxes were even simpler as there were only six options.

H2.2:

To compare the effect of the input method on the mouse clicks, we conducted a Wilcoxon signed rank test. With the input method checkbox, the average time was 3.05 ms (SD = 1.99), for slider, the average was 5.15 ms (SD = 2.83), and for value input, the average was 6.788 ms (SD=6.08). The mean difference between the slider and the checkboxes was statistically significant: V = 400.5, p-value = 1.575e-07. Also, the mean difference between the checkboxes and the value input was not statistically significant: V = 1106, p-value = 0.1785. Lastly, we also checked the difference between the slider and the value input and found it

was also statistically significant: V = 297.5, p-value = 6.491e-07. These results indicate that the inout method slider was slower than the checkboxes but faster than the value input. Therefore, H0 is rejected. This was probably because the checkboxes are much simpler to use and have less options so there is less error when choosing the appropriate compensation to maximize the budget.

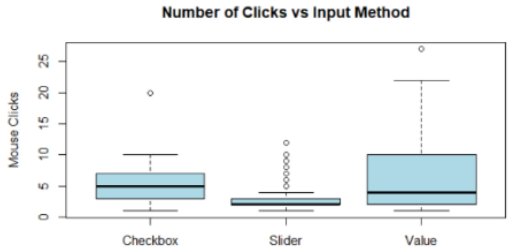

Figure 9: Number of clicks test 2

H2.3:

All graphs in Fig.1-3 in the appendix were measured on a scale of 1 to 5. Where 1 was the least intuitive and 5 was the most intuitive.

To compare the effect of the input method on the intuitiveness score, we conducted a Wilcoxon signed rank test. With the input method checkboxes, the average score was 4.29 (SD = 1.213), slider had an average of 4.588 (SD = 0.62), and value input had an average of 2.47 (SD = 1.07). The test between value input and slider was statistically significant at V = 136, p-value = 0.0003913 and so was the best between value input at checkbox at V = 131, p-value = 0.001085, as both p-values were less than 0.05. These results indicate that the value input method had a worse intuitive score than the checkboxes and the slider. However, no comparison can be made between the checkboxes and the slider, as the p-value was greater than 0.05 (V = 20, p-value = 0.3429). Thus, H0 is partially rejected. This makes sense because the value input entry forced the user to think about how much they wanted to compensate and have to type it in as well. It was more annoying to deal with than the other two input methods, which were easier to simply slide the slider or just check a box.

H2.4:

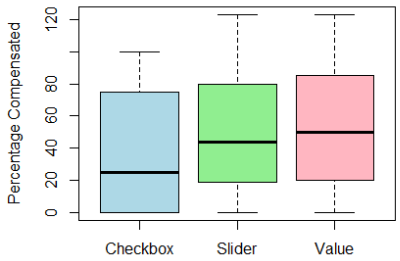

Figure 10: Percentage compensated with input method

To compare the effect of the input method on the percentage compensated, we conducted a Wilcoxon signed rank test. With the input method checkbox, the average percentage compensated was 41.875% (SD = 36.19), for slider, the average was 50.388%(SD = 37.244), and for the value input, the average was 54.11% (SD=37.85). The Wilcoxon test between the slider and checkbox (V = 2690, p-value = 0.0806), slider and value input (V = 3063, p-value = 0.6412), and checkbox and value input (V = 2578.5, p-value = 0.03311) led to p-values greater than 0.05. These results indicate that there was not enough evidence to conclude that there was a difference between the percentage compensated. The H0 is not rejected. This was possible because of there being a large error which led us to not being able to make a better conclusion. More testing would be appropriate.

3 Limitations

The results of our user study need to be interpreted with care, as the testing population was quite small with only 16 participants and the age distribution was heavily biased to the age group of 20 to 30 years old, with only two participants belonging to different age groups. Another bias might have been introduced due to the participants' educational status, as all the participants had completed at least a Matura or bachelor degree. This affects the credibility of our results and leads to a limited representability of the real population. Our questionnaires consisted of questions which might not have been formulated clear enough and also included a lot of subjective ratings. Due to this, the results gained from the questionnaires might not be completely representative of the real values and the number of open-ended questions were quite short.

4 Future Work

One would need to test with a broader population to get more conclusive results. A future prototype could feature a slider with buttons added at certain, often chosen values, such as 0%, 50%, 100% and 125%. This could be interesting, as we observed that 100% is very often or always chosen by some of our participants.

5 Conclusions

Looking back at our results, we conclude that our plugin sadly did not help to inform the users better about their emissions. The line between keeping it simple and providing more information is quite thin. On the other hand, we achieved a higher average compensation percentage with our plugin, as it provides more flexibility to our users.

6 Appendix

All questionnaires and some additional figures not included in the report can be found in the appendix (see GitHub repository inside *Deliverables/Study*) .