Harvard CS279 Tier 2: Comparing When2Meet and Doodle

Simon Alexander Warchol
Harvard John A. Paulson
School of Engineering and
Applied Sciences
Boston, MA USA
simonwarchol@g.harvard.edu

USER STUDY

The goal of this user study is to compare When2Meet and Doodle. In this study, we collected qualitative user feedback about each system as well as quantitative feedback regarding users' performance and accuracy when using each tool to input times and discern the availability of others.

Experiment Conditions

Task 1: Inputting Availability
Given a list of time slots in text form, a user must input these times as available using each system.

Task 2: Reading the Availability of Others
Given randomized availability of 3 simulated system users, the participant must input how many slots exist where all 3 are free.

Experiment Set-Up

All participation in this study is remote; users visit a web address and are prompted to complete each task in both Doodle and When2Meet. As we are measuring the speed and accuracy with which tasks are completed, users are not limited by their unfamiliarity with specific hardware or software, which could impact results if all testing was done on a centralized machine.

Participants

All participants were students currently enrolled in CS279R. We recruited 6 students, who were not compensated for their time.

Limitations

We note several key limitations with our experiment. Most notably, the small number of participants limits the statistical significance of any findings. Moreover, the fact that all 6 participants were classmates implies a level of familiarity with both systems that is not necessarily representative of a broader public. Beyond limitations with the participants, we acknowledge that the order in which users accomplish the tasks may skew results, specifically in Task 1, where the time blocks are static. If a user has already input the same dates into one system, they may be quicker at doing so in another. Thus, randomizing the dates and order of systems are both advisable. While we do randomly generate simulated availability for Task 2, this process may also impact results. We hypothesize that if a pre-attentive number of slots are free, users will be quicker and more accurate in completing the task. In future work, a larger number of participants and more detailed analysis could address this limitation.

ISBN 978-1-4503-2138-9.
DOI: 10.1145/1235

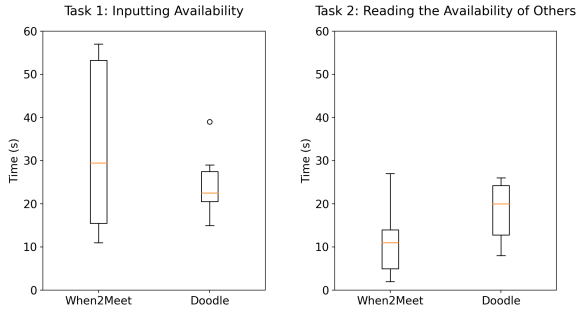

Figure 1: **User Study Results**: We find that users faster at inputting availability into Doodle, but faster at interpreting the availability of others in When2Meet.

RESULTS

Quantitative Results

We evaluate participants in terms of both speed and accuracy. In the case of both Task 1 and Task 2, users were 100% accurate. We visualize the speed with which participants accomplished both tasks in **Fig.** 1. From our limited sample size, users, on average, accomplished Task 1 quicker in doodle than When2Meet. However, they completed Task 2 quicker in When2Meet than in Doodle. Given the small participant pool, it is hard to draw any conclusions from these findings. However, we assert that such an experiment could be conducted at scale and yield statistically meaningful results. The same can be said for all accuracy ratings, as, with a larger sample size, participant performance should vary more.

Qualitative Results

After using each system, participants were asked to evaluate the system. Specifically, using a 5 point Likert scale, users

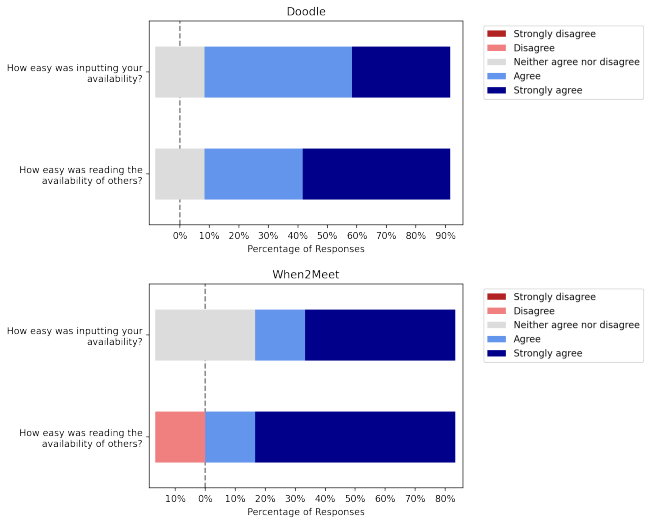

Figure 2: **User Study Results**: We find that users faster at inputting availability into Doodle, but faster at interpreting the availability of others in When2Meet.

answered the following two questions: **(1)** How easy was inputting your availability?, **(2)** How easy was reading the availability of others? It is hard to make much of the results of this survey given the small sample size and lack of a clear distinction between the two systems. One participant, however, really did not like reading the availability of others in When2Meet. As with the quantitative evaluation, we anticipate that this will not necessarily be the case as the participant pool increases dramatically in size.