

## SF2020 Usability Study Summary

### **Objective and Procedure**

The primary objective of the usability study was to determine the extent to which the newly-developed SF2020 replicates the gameplay experience of Revised Space Fortress (RSF). Ten female, right-handed participants played 5 practice games and then 10 test games (Block 1) of one version of the game, then another 5 practice and 10 test games (Block 2) in the alternate version. Assignment to which version was encountered first was handled according to a pre-determined balanced randomized sequence to control for potential order effects.

### **Analysis Plan and Results**

We hypothesized that if SF2020 effectively replicates RSF, we would expect participants' performance and behavior metrics from Block 1 to be predictive of the same metrics in Block 2. Alternatively, If Block 1 metrics are unrelated or negatively related to Block 2 metrics, this might suggest significant differences in how the two versions handle certain gameplay mechanics. To test this, we examined the correlation between participants' Block 1 and 2 metrics for 17 variables recorded by the SF game (one additional variable was examined later, as discussed below). To obtain more robust estimates of each variable, we averaged across the 10 test games for each block and used these mean values in the analyses.

As can be seen in the figure below, the majority of the variables displayed a marked positive relationship between versions. Of the 17 variables examined, 13 showed significant positive correlations. The exceptions were clockwise rotations ( $r = .33, p = .35$ ), fortress collisions ( $r = -.61, p = .06$ ), number of friend mines destroyed ( $r = .61, p = .061$ ), and the Speed score ( $r = -.38, p = .274$ ). Importantly, Total score—the indicator of overall gameplay performance—was significantly correlated across versions ( $r = .66, p = .038$ ), despite the Speed

subscore demonstrating a negative relationship. A new variable, which subtracted the Speed score from the Total score, was computed and, as expected, the correlation for this metric was improved compared to the Total score that included Speed ( $r = .71, p = .018$ ).

### **Conclusions**

Taken together, these analyses suggest that SF2020 and RSF provide quite similar experiences for participants, with a few notable exceptions. The fortress collisions metric for all SF2020 games was exactly zero, so it is possible this might be due to an error in recording. The discrepancy in the number of clockwise rotations is more difficult to explain, but given the strong correlations for ship thrusts and counterclockwise rotations ( $r = .82, p = .004$  and  $r = .83, p = .003$ , respectively), it seems more plausible that this is driven more by insensitivity due to the small sample size rather than an issue with recording joystick inputs or biasing player behavior. Similarly, the number of foe mines destroyed trended toward significance ( $p = .061$ ), and no participants verbalized any observation of differences between the friend mine mechanics across versions, so this might also be attributed to sample size. The negative correlation between players' Speed scores is more concerning, however, given it is one of the primary performance metrics. Examining the average Speed scores across the two versions, it is notable that 8 out of 10 players achieved a negative mean Speed score while playing RSF, whereas all 10 players achieved a positive mean Speed score while playing SF2020. This perhaps suggests a difference in the two versions' scoring methods, but more technical analysis will be required. To facilitate this, the output text files from all the games (organized by participant ID number) can be accessed [REDACTED].

