By: Franklin Li

A Comparison and Contrast of the Parallel Randomized Control Trial (RCT) and Multi-Armed Bandit (MAB) with Thompson Sampling (TS)

Figure 1: Average Playtime per Iteration for Each Variant

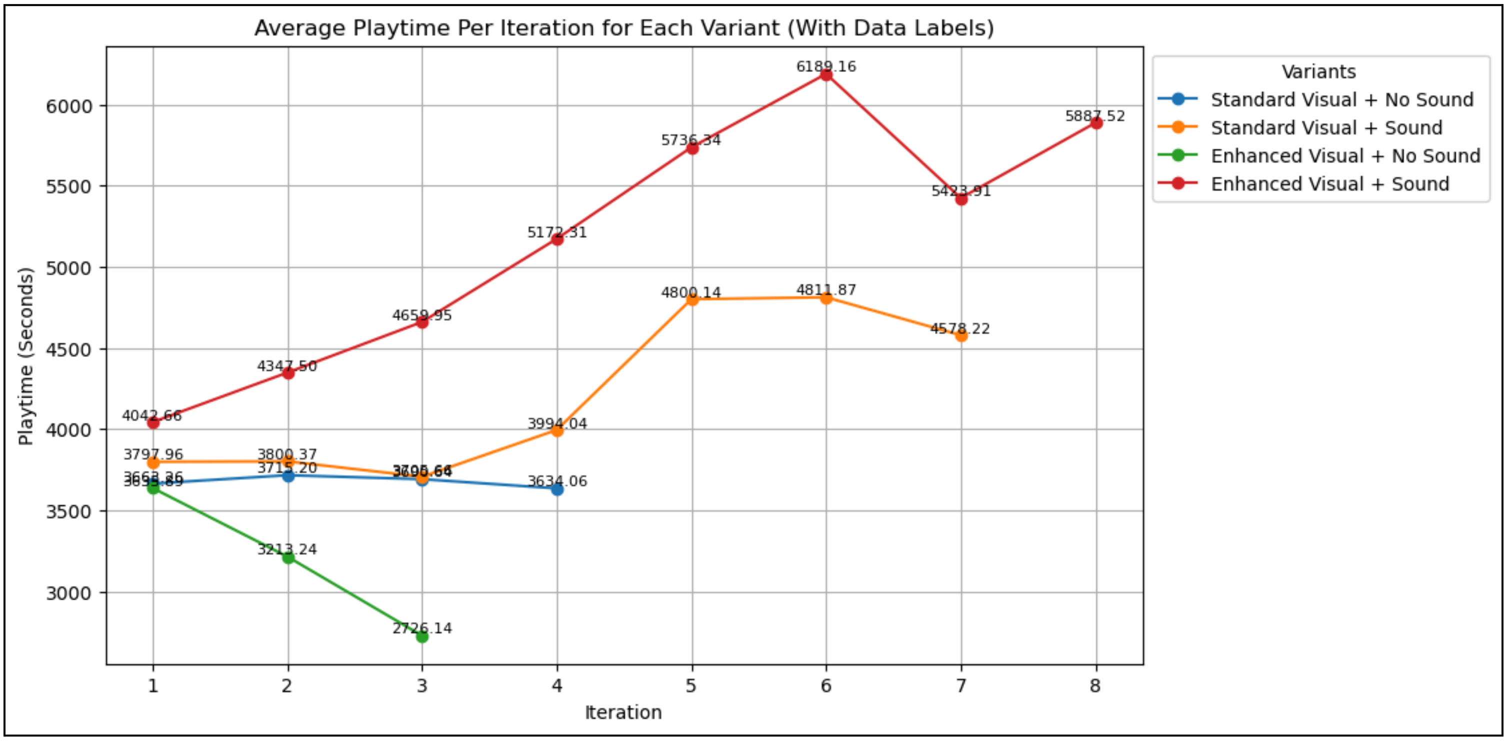

Figure 1. This line graph shows for each iteration, the non-cumulative allocation of average playtime (in seconds) to each variant (for which the allotted sample size will vary according to how well it performed in the previous iteration).

Figure 2: Allocation of Plays Per Iteration for Each Variant

Iteration	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Sum**
Standard Visual + No Sound	40	20	24	28	0	0	0	0	112
Standard Visual + Sound	40	47	57	67	81	40	20	0	352
Enhanced Visual + No Sound	40	47	23	0	0	0	0	0	110
Enhanced Visual + Sound	40	46	56	65	79	120	140	160	706
Sum	160	160	160	160	160	160	160	160	**1280**

Figure 2. This table lays out the dynamics of resource allocation throughout the Thompson Sampling process. The figure shows premature elimination of poorly performing variants (by iteration four for Enhanced Visual + No Sound and by iteration five for Standard Visual + No Sound) and stopping of the algorithm at iteration eight for Enhanced Visual + Sound.

Figure 3: Key Statistical Features and Outputs of Parallel RCT and Multi-Armed Bandit

Result or Feature	Parallel RCT	Context-Independent MAB (Thompson Sampling)
Average Outcome	• Standard Visual (1199.76 s) • Enhanced Visual (3000.00 s)	Stopping/Elimination reached: • Standard Visual + No Sound (3634.06 s at i4) • Standard Visual + Sound (4578.22 s at i7) • Enhanced Visual + No Sound (2726.14 s at i3) • Enhanced Visual + Sound (5887.52 s at i8)
Contextualization	• Unadjusted group means (context-free: not taking into account blocking factor and covariate)	• Context-free: does not account for confounders
Decision-Making	• Pure exploration	• Probabilistic balance between exploration and exploitation
Insights	• Strong causal inference • No backtracking in terms of random assignment	• Reinforcement learning (continual performance optimization) • Backtracking allowed in terms of resource allocation
Sample Size Allocation	• 50/50 across both groups from the start and in terms of cumulative allocations	Standard Visual + No Sound, Standard Visual + Sound, Enhanced Visual + No Sound, Enhanced Visual + Sound: • 25/25/25/25 at iteration 1 • 0/0/0/100 at iteration 8 • 8.75/27.50/8.59/55.16 in terms of cumulative allocations

Inference	• Static as allocation ratio remains constant and independent of past outcomes • Strong causal inference	• Decision-making in real time (allocation ratio adapts according to performance of variants) • more careful analysis required to make causal claims
Inference Accuracy	• Setting significance threshold reduces type I error (false positive) • Conducting power analysis reduces type II error (false negative)	• Inflated type I error without the use of appropriate correction methods • Undersampling can lead to type II error
Internal Validity	• High internal validity due to random assignment (confounding variables are balanced across groups) • Unbiased estimation of treatment effects	• Introduction of potential selection bias (e.g., temporal confounds) • Regret minimization

Figure 3. The above table compares and contrasts the results or key features of the parallel randomized control trial and multi-armed bandit in terms of statistical logistics or output.

Similarities in Participant Allocation and Context Independence

Our parallel randomized control trial (RCT) and multi-armed bandit with Thompson Sampling (MAB with TS) entailed equal random assignment across conditions from the very start of the study; our RCT assigned 50% of participants to either conditions, while our MAB allocated 25% of all resources to each of the four variants. Moreover, both of our models were context-independent. Our parallel RCT entailed raw, unadjusted group means that did not account for our blocking factor (time of day) and covariate (age), and our MAB did not use Bayesian methods to account for blocking factors or covariates.

Similarities and Differences in Decision-Making

Both models entailed some form of exploration. The parallel RCT made full use of exploration, treating each condition equally. The MAB also made use of exploration, but in terms of differences, as a whole, the MAB balanced exploration with exploitation in a probabilistic manner. In other words, a key difference between the parallel RCT and MAB is the fact that the former is purely exploratory, whereas the latter makes greater use of exploration early on (when uncertainty is higher, though exploration happens more so for better-performing variants with high uncertainty than for lower-performing variants with high uncertainty) and greater use of exploitation later on (when uncertainty is lower but better-performing variants are still undergoing iterations).

Differences in Allocation Ratio, Causality, and Outcome Dependence

In terms of differences, our parallel RCT did not allow for backtracking in terms of random assignment, whereas our MAB allowed for reinforcement learning (dynamic allocation across time) and regret minimization (exploitation) and by the final iteration (the point at which our stopping rule was satisfied), all resources were allocated to the highest-performing variant (enhanced visual + sound).

Moreover, our parallel RCT made it possible for us to infer a strong causal effect of visual conditions on playtime; enhanced visuals led to significantly longer average playtime (3000.00 seconds) than standard visuals (1199.76 seconds), and this difference was characterized by a large effect size (Cohen's d). Moreover, with respect to our parallel RCT, setting a significance threshold of $\alpha = 0.05$ and conducting a power analysis beforehand allowed us to minimize type I and II errors. On the other hand, the adaptive nature of the MAB did not allow for such causal inferences to be made. Without the application of correction methods such as Bayesian models, the MAB was more likely to produce type I errors. Additionally, because underperforming variants were undersampled, the MAB was more likely to produce type II errors.

Once again, our parallel RCT entailed equal random assignment (50% to each condition) at a fixed time point (beginning of the study); there was no discrimination between conditions in terms of performance, and no conditions or variants were discarded or punished. Moreover, our parallel RCT design did not implement any stopping rule, nor did it end prematurely based on some arbitrarily imposed criterion; statistical analysis only occurred after all participants were assigned to a condition (treatment or control) and the data from the entire sample was collected. On the other hand, our MAB entailed real-time optimization of resource allocation, implementation of elimination and stopping rules in favour of exploitation, and dynamic changes in sample size allocation across time according to performance. The presence of our elimination rule (10% of our sample size or 16 allocations) enabled the model to discard poorly performing variants. The presence of our stopping rule enabled us to identify the best-performing variant by the point at which the stopping rule was satisfied (95% of our sample size or 152 allocations). In terms of proportions of cumulative allocations, our MAB showed that enhanced visuals and sound (55.16%) led to the longest average playtime, followed by standard visuals + sound (27.50%), standard visuals + no sound (8.75%), and enhanced visuals + no sound (8.59%).

Interestingly, our two approaches led to contradictory results; while our MAB showed that prior to elimination, enhanced visuals + no sound led to shorter average playtime (2726.14 seconds at iteration three) than standard visuals + no sound (3634.06 seconds at iteration four) did, our parallel RCT demonstrated that overall, enhanced visuals (with no sound; 3000.00

seconds) led to significantly longer average playtime than standard visuals (with no sound; 1199.76 seconds) did.

Overall Thoughts

Overall, the results of our parallel RCT and traditional MAB with TS are more different in their key features than they are similar. Both approaches involve the random assignment of an equal number of participants to each condition or variant at the beginning of the study and do not account for blocking factors, such as time of day, or covariates, such as age. However, causal claims can be made more plausibly with respect to the parallel RCT than to the MAB. The parallel RCT involves pure exploration and measures outcomes at a single timepoint for each participant, providing every participant an equalized chance of being assigned to each condition, whereas the MAB entails a progressive shift from exploration to exploitation across multiple iterations, punishing and in our case, even discarding poorly performing variants and rewarding the best-performing variants. That is, the parallel RCT enables us to make more clear-cut claims with statistical precision and certainty (hypothesis testing), whereas MAB emphasizes decision-making in the here-and-now in terms of which variant performs the best. The needs and goals of the researcher may help determine which approach to implement.