Proceedings, The Twelfth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE-16)

Contained: Using Multiplayer Online Games to Quantify Success for Collaborative Behavior

Damian Debkowski,[1] **Andrew Marrero,**[1] **Nicole Yson,**[1] **Li Yin,**[1]
Yichen Yue,[1] **Seth Frey,**[2] **Mubbasir Kapadia**[1]
[1]Rutgers University, [2]Dartmouth College

Abstract

Every day, millions of people gather on online game servers to collaborate in real time toward shared goals. What may seem like frivolous activity may, when investigated more closely, provide revolutionary opportunities to advance the science of teamwork. Teamwork is an important part of modern society, however, collaboration between people is often made difficult due to differing ideals, opinions, and personality types. We leverage a popular self-hosted multiplayer online game environment to design a framework for developing and deploying tasks that elicit different kinds of teamwork. We propose to use these to capture fine-scale details of individual and group performance across environments. The game in which we implement this system, Minecraft, is ideal because it is heavily modifiable and already enjoys a vast user base of surprising gender, age, and ethnic diversity. We heavily modify the game in order to introduce new mechanics that facilitate collaboration, monitor activity, and manipulate group composition, all to provide the groundwork for deeper quantitative insights into effective teams. A supplementary video can be found at https://db.tt/QxOVe2JL.

Introduction

Ideally, the result of effectively working in a team is that the combined work of each group member amounts to more than what could be equivalently achieved by an individual in the same amount of time. For this reason, companies, workplaces, and academic settings find it more advantageous to collaborate in order to reach a desired goal.

Since collaboration by definition is working together to realize shared goals, conflict often arises by something more innate than clashing objectives: personality. Correlating personality and successful collaboration allows for conclusions to be drawn to optimize team content in terms of the properties of individuals. Through an analysis of personality in Minecraft, we aim to analyze group success in terms of group composition. With insight from multiple coordination tasks, results in a popular and easy-to-study environment such as Minecraft are more likely to generalize to the workplace, classrooms, and anywhere people are required to collaborate. Leveraging the full power of a Minecraft setting to study teamwork permit deeper and more efficient insights into questions like team composition. The existing literature on the effects of personality factors on team effectiveness is mostly in social psychology. (Neuman, Wagner, and Christiansen 1999) investigates the relationship between team effectiveness and the average level of a trait given within a team and team personality diversity. In this study, for the traits of conscientiousness, agreeableness, and openness to experience, team personality evaluation was positively correlated to team performance; team personality diversity in regards to extraversion and emotional stability was also positively related to team performance. (Barrick et al. 1998) examines relationships among team composition in terms of ability and personality and team performance. Teams higher in conscientiousness, agreeableness, extraversion, and emotional stability received higher supervisor ratings for team performance. Despite the strength of these studies they often specify one vital issue—the ultimate quantification of team success—in terms of subjective or ill-defined characteristics.

We present "Contained," a set of heavy modifications and additions to Minecraft's base engine. We expand upon previous work by introducing many new Minecraft mechanics and game modes that facilitate or require groups of players to communicate and cooperate together, and by implementing a plug-in that logs behavioral data about players as they play.

We split a Minecraft Forge server into several separate game worlds. The main is the "lobby" world, which like a traditional Minecraft server is intended for extended long-term play. However, the bulk of our experiment is handled through "mini-game" worlds, which are offshoot worlds of the lobby that have groups of players participating in short 40-minute games. If a team is able to accomplish one of the mini-games' winning conditions, then the mini-game can be shorter than 40 minutes. These games are designed so that good strategizing and communication are crucial for each team's success.

The lobby world contains many new features that *facilitate* cooperation between players, while the mini-game worlds *require* cooperation between players. We log gameplay and statistical data for players in all of these. On top of this, each player is required to fill out a short personality survey. All of these together provide a very efficient automated platform for implementing teamwork experiments that can elucidate the interactions between personality, col-

laboration, and group success.
Our proposed solution has the following main benefits over prior work:

1. Previous studies that logged player behavior in online games were forced to tailor their collected data to whatever mechanics the games happen to offer. Instead, we use a modding framework to modify the game, tailoring it to the types of data we wish to collect.
2. Our study implements game mechanics where players are required to work together in a team environment. We also create some tasks that allow this collaboration to be measured and quantified. This gives more useful data than previous studies which studied collaboration in games where teamwork was completely voluntary.
3. By leveraging a popular online platform, we dramatically simplify the task of recruiting subjects. The resulting high-throughput experimental platform can accelerate the pace of discovery in a field that has traditionally been hampered in its ability to quickly iterate through experimental designs.

We present "Contained," a set of heavy modifications and additions to Minecraft's base engine. We implement Contained using the popular Minecraft Forge modding framework, allowing our final mod to be deployed on any public or private Minecraft server that is setup for Forge mods.
We demonstrate the benefit of modifying Minecraft to tailor it to the study by devising two custom mini-games which have groups of players compete in teamwork-oriented tasks. We use these to quantify different facets of collaboration, toward understanding how different personality types contribute to group success.
In order to validate whether logged actions could be mapped directly to personality factors, we implement an established survey in-game to see if the measured activities correlate with the independent measure of personality factors. By laying down personality measures as ground truth. We can then establish correlations between survey data and logged actions together online.

1. User Study: We set up a server instance with Contained. As groups arrive, they complete surveys, and perform "mini-game" teamwork tasks within ad hoc teams.
2. Quantitative Evaluation: We take data logged from groups playing the mini-games, using this as labeled training data to find trends between personality factors and the success of teams in these games.
3. Qualitative Evaluation: We provide several tools to allow administrators to actively monitor the games on the server to obtain qualitative feedback, as well as logging player chat logs to provide some qualitative context for the otherwise quantitative data we store in the database.

Related Work

Video Games as Vehicles of Analysis Researchers have shown that virtual worlds provide the means to implement controlled experiments at the scale of whole societies (Bainbridge 2007; E. Castronova 2013). The advantage to performing studies virtually via video games is that it makes

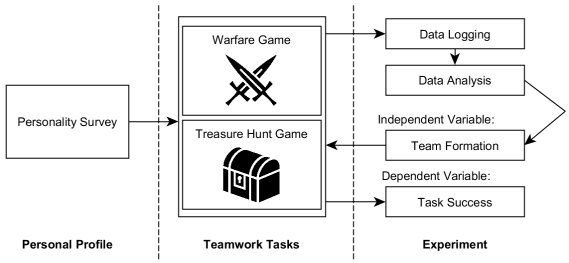

Figure 1: Framework Outline. After completing individual surveys, players are randomly assigned and reassigned to groups that compete in different tests of teamwork.

it easy to research large numbers of people across sociocultural boundaries and to perform large-scale, long-term experiments. The social-scientific value of video game data is becoming increasingly widely recognized across the social sciences. Such work demonstrates the promise of virtual worlds for studying macro-scale phenomena empirically (Castronova et al. 2009; I. Constantiou 2012). We advance this work by using a game environment to generate teams that must work under conditions of resources constraints. Freeman states that teams that work collaboratively can obtain greater resources, recognition and reward when facing competition for finite resources (Reviews 2012).

Minecraft Others have recognized the value of Minecraft for studying human collaboration. HeapCraft (Mueller et al. 2015a; 2015b; 2015c) offered a framework for collecting large volumes of unconstrained game data, although it may not be entirely generalizable to a real-life context, because it analyzes a sandbox world with infinite resources. Analyzing this type of infinite world does not take into consideration Realistic Group Conflict theory, which proposes that prejudice and inter-group conflict stem primarily from competition over finite resources (Sutton 2013). We couch our efforts in terms of the five-factor model of personality, as a framework for understanding individual differences. The five-factor model is ideal because it has been thoroughly validated (McCrae and Costa 1987) and used heavily in and beyond its original field of personality psychology (Judge and Bono 2000).

Framework Overview

The overview of our project is seen in Figure 1. The first step is to gather data at the individual level. Upon first entering the game, players are prompted to fill out a survey that can be used to collect vectors of individual-level features. The data logger also collects all player actions through Forge's event handler, which fires automatically for a large variety of different gameplay related tasks.
The second step is to gather data at the team level. Players can form their own teams which can own territory, and all of this type of data is logged as well; when teams are created, when players join/leave teams, when territory is claimed, when a player is elected to leader status within the team, etc. In addition to these player-created teams, there are also tem-

porary randomly generated teams created during our mod's mini-game sessions. These are short games that have groups of players compete in teamwork related tasks. All of this is explained in more detail in the next section.
All data is time-stamped and stored in a database for later use. Run on a large number of servers, Contained can provide large volumes of team behavior data for analysis. The core of our own user study revolves around looking at the winning and losing teams of each mini-game, and identifying trends in the personality combinations that consistently cause successful or poor team performance.

Minecraft Modification for Collaboration

As our study focuses on studying collaboration between people, it was essential to make sure that we modify Minecraft so that group collaboration becomes a central part of the game, rather than a voluntary decision. We implemented a number of new gameplay mechanics and modifications into Minecraft, all of which share this common goal of helping to facilitate teamwork between players. This list of features is as follows:

1. **Finite World.** "Vanilla" Minecraft has an infinite world which randomly generates more of itself as players explore beyond a server's central spawn location. The problem with this system is that people often travel so far that they become isolated from all other people on the server. We redesigned Minecraft's world generator so that after a finite configurable distance from the spawn point in all directions, all further world generation only creates "wasteland", which is barren and extremely hostile, providing no incentive to explore the world beyond that point. Thus, all players on the server are induced to remain close to each other at all times.
2. **Resource Clustering.** Usually, resources are scattered uniformly throughout the world, making them harder to exclude or deny to other players. We also modified this, so that each individual resource type only appears in a small number of places around the finite world, but appears in a very large quantity as a localized cluster. Consequently, a group of players settled in a strategic location can dominate one or two sets of resources, while lacking others. This enhanced resource clustering is intended to motivate trading and negotiating with neighboring groups.
3. **Player Teams.** At the core of our mod's teamwork system is a system for organizing player teams. Using an in-game GUI, players can create teams and invite other players to join, each team having a maximum occupancy of five people. Many of the gameplay features in our mod revolve around these teams, and they are largely a mandatory part of the game. To enforce this, we've modified some of the default Minecraft mechanics to require teams as well (for example, you can't mine resources until you are a member of a team). This enforces team boundaries and encourages the emergence of shared team objectives. Local team-exclusive chat, and other features, are also introduced to help players work better together.
4. **Territory Acquisition.** Teams can purchase areas of land which they expand over time. These areas become protected, effectively entering a "read-only" mode for anyone not part of the team. Teams can have a location that they can return to, regroup, and securely store property. It also associates each team with a relatively stable place, encouraging collaboration that is syncronous in both time and space.

Figure 2: A player in team territory, highlighted in orange.

5. **Customizable Permissions.** By default, a team's territory is fully protected in every aspect from other teams. However, this is not always ideal, as it may be desirable to permit teams to become allied. We developed a GUI team permissions system to allow team leaders to assign and manage specific permissions of outsiders within the team's territory. Permissions are fine-grained, and include controlling outsiders' ability to break blocks, place blocks, open chests, interact with NPCs, pick up dropped items, harvest crops, etc.
6. **Trading.** Each team has access to a GUI where they can post trade requests (i.e. selling a certain item in exchange for a different specific item) that are visible in a list to other teams, and in a bulletin board. Trading can also be tuned within the permission systems, permitting embargoes against other teams. Trading systems provide a simple resource distribution system that complements' each world's clustered resource structure.
7. **Occupational Classes.** Players in the Contained framework can gain experience among different customizable classes, e.g. Warrior, Wizard, Cook, Builder, and Collector. Gaining enough experience allows the player to choose a "perk" within their class, giving them special upgraded abilities. Within task specialization systems, individual contributions to a group become constrained and easier to measure. It also fosters the strategic division of labor.

Minecraft Mini Games

At any point while on a Contained server, a player can request to join a mini-game. Each mini-game is a short play-session that lasts 40 minutes, pitting two randomly selected teams of any 5 members each against each other. These games are designed so that effective, competitive teamwork demands strong communication and collaboration skills. There are two different types of mini-games available to play: the Warfare game, and the Treasure Hunt game.

Figure 3: A player queuing to join the mini-game using the mini-game panel.

Warfare Game

In the warfare mini-game, the opposed teams spawn on opposite sides of the map, and are given a small area of claimed territory. Unlike in the normal "lobby" environment, it is not possible to expand upon the territory. By being smaller, the Warfare territory also helps facilitate combat.

Upon killing a player from the other team, that player drops a few "anti-territory gems", and is sent back to their territory where they started. These gems can be used to steal pieces of the opponent team's territory, by running into their territory and clicking on the ground with the gem. There are three different ways to win this game:

1. Steal all of the opponent team's territory.
2. Kill all of the opponent players (everyone has a finite number of lives).
3. Be the team with the most kills once the timer runs out.

Teamwork and strategy naturally come up in the Warfare mini-game, as players can stage coordinated attacks, implement task specialization, and organize overt or covert offensive movements. In the domain of task specialization, we support players in adopting high-level roles such as:

- Offensive roles, in which groups of players actively engage the each other in combat, carry out invasions, and steal territory.
- Defensive roles, in which some players hold back, build fortifications, set traps, and otherwise protect their team's territory from the opponent's offense.
- Resource scouting roles. Resources found in the world can be used to craft weapons and armor that will give your teammates a large advantage in combat. Because searching for resources can be time-consuming, and these mini-games are very short, we've created a special "dowsing rod" item specifically for the mini-games to help with this role. This item allows players, with one click, to check all of the resources that are present within a small radius of their current position.

Treasure Hunting Game

In the Treasure Hunt mini-game, each team gets an area of territory, which contains four special blocks called altars. These altars start in an inactivated state, but can be activated by inserting an emblem of the same color. These emblems

Figure 4: Screenshots from the mini-games: the left column illustrates how teammates collaborate to collect resources and prepare during the Warfare game; the right column shows how the players in the Treasure Hunt game explore and collect chests.

are constructed by assembling together fragments that are found in the chests around the world. There are two different ways to win this game:

1. Be the first team to activate 3 out of the 4 altars.
2. Be the team that collected the most chests once the timer runs out.

The chests containing emblems spawn in random locations, highlighted by beacons that are always visible regardless of where in the game world one is. Chests spawn for the most part on the surface. Any time a chest is collected, the team gains a point, and a new chest spawns in a random location elsewhere on the map. Treasure chests also contain a variety of other items useful for survival and exploration.

One way that the game encourages team work is that the most efficient way to win would be coordinating team members to split up and cover as much ground as possible. It is also essential for players to frequently communicate with each other, so that they can keep track of which fragments the team has found so far, and which are remaining. Once it has been determined that the team has gathered the needed fragments, players must consolidate back at the territory, in order to share the fragments among each other and activate the altars. Direct player-versus-player combat is disabled in the Treasure Hunt in order to preserve a clear distinction between the two mini-games.

These two mini-games are very different in the playstyles. The treasure game is a passive game focusing on exploration and coordination of resources. The warfare game is an active, aggressive game focusing on combat. This was a conscious choice, to illustrate a flexible framework in which the task-dependency of group composition can be established with a variety of mini-game environments. In the case of the two games introduced here, it may be that the personality types that excel in an aggressive group task are

different than the personality types that excel in a passive group task.

Evaluation

User Study

Our user study seeks to take the logged data we collect from users playing within the Contained framework, and use this data to discover relationships between personality types, and how successfully people work together in a team. Thus, the important quantitative variables we are concerned with are as follows:

Dependent Variables

1. Group Success: The performance of a team. We use our mini-game system to quantify and measure this value. It is quantified in two ways. Since each mini-game consists of exactly two teams, we can have a simple Boolean measurement as to whether or not the team won the game. However, to be able to additionally perform ranking within these winning teams, we also have a second numerical quantification. In the Warfare mini-game, this would be the number of kills the team achieved, while in the Treasure Hunt mini-game, this would be the number of treasure chests the team collected.
2. Task Independence: It's important to see whether the trends we identify are dependent on the task or not. This is the primary reason our study introduces two different teamwork-oriented tasks, both with very different affinities. For the sake of the analysis, our data logger makes sure to log the distinction between data pertaining to teams in the Warfare game, versus data pertaining to teams in the Treasure Hunt game.

Independent Variables

1. Individuals' personality factors.
2. The set of individuals cooperating in a group.

The first step of our study is to collect personality data on each of the users that decides to play in our Minecraft server. We achieved this by implementing an in-game GUI interface that administers the survey. This GUI appears automatically the first time the player enters the server, and they must complete it before they can begin participating. We took this path of making the survey mandatory, because the player's data will be rendered unusable for our study if the survey data is not collected. The questionnaire is an in-game implementation of the standard 50-question Big Five personality test (McCrae and Costa 1987).

Quantitative Analysis. In the first pass of our experiment, teams are assigned randomly to their groups. The variation in group composition due to random assignment and reassignment to teams permits controlled variation ideal for training classifiers that can predict the success of a group based on the personality vectors of its members. In a second pass of the experiment, we will stop assigning teams randomly. Instead, we will use our trained classifier to assign half of the teams in ways that are predicted to be very successful, and half of them in ways that are predicted to be unsuccessful. Questions we seek to answer with this quantitative analysis:

Figure 5: A player filling out a survey in-game.

1. Do the teams built to be successful consistently perform better than those built to be unsuccessful?
2. Is the average performance of the classifier's successful teams consistently higher than the average performance of the randomly assigned teams?
3. Does team performance degrade if we use the Warfare classifier to assign Treasure Hunt teams, and vice verse?

If it turns out that group success is actually dependent on the task, then the idea of finding a "one-size-fits-all" set of successful personality make-ups is not feasible. Thus, if the second pass fails to provide comprehensive results, we have devised a third pass of the experiment to further test the dependency between personalities, task orientation and group success. This is divided into two scenarios. In the first scenario, we assign groups such that each group only contains players of the same personality. In the second scenario, we assign groups with exclusively mixed personalities. We want to observe the level of team success these team formations achieve in the different mini-game tasks. We hypothesize that groups of a more diverse makeup will be more successful in the Warfare Game, since it lends itself to a variety of roles (offensive, defensive, scouting). On the other hand, in the Treasure Hunt, players' tasks are largely identical, so we hypothesize that similar personalities might contribute to better communication and collaboration.

Qualitative Analysis. We have implemented an in-game admin console, which allows admins to monitor, teleport to, and spectate any of the online players, in any of the game worlds (lobby and mini-game worlds). Admins are set to be invisible, so users are not aware when an admin is spectating. This offers a useful way for admins to actively observe qualitative information from the participating users, without interfering with the study.

Proof of Concept As a first step, we demonstrate the stability of our platform with a series of pilot tests. In the most successful pilot, we ran two rounds of each mini-game type, each game being assigned two teams of three people, randomly selected from a pool of 8 online players, and collected all performance data. Before play, we also used the in-game survey mechanic to collect personality and other

Figure 6: An administrator of the game uses the admin panel to monitor players' behavior during the mini-game.

data from each player. We ensured the quality of this survey data with help from a simple system for detecting survey spam. This effort establishes the stability of our platform for experiments of group composition and team effectiveness in online tasks.

Conclusion

We present a framework for researchers to quickly iterate and implement diverse "mini-game" teamwork tasks for small distributed groups. We started with a popular, heavily modifiable, self-hosted sandbox video game, Minecraft, and built upon its existing mechanics to fashion it into a tool appropriate for high-throughput group research. We implemented two sample mini-games which serve the purpose of helping to quantify collaboration, by having groups of players compete in teamwork-oriented tasks. The mod can be built and deployed on public servers, and data is passively collected from players and stored in a database for later analysis. This provides a powerful tool for use by behavioral scientists and other researchers.

Future Work

Even with the advantages of a popular online game, the recruitment still requires an active management effort. We have attempted to incentivize collaboration by couching it within familiar game mechanics, possibly at the cost of generalizable insights.

Future work will continue to scale up our experimental system, and apply more groups to a greater variety of tasks. Larger, more heterogeneous populations will allow greater power to generalize to the real world, even allowing for the limits of simple game environment. Moving beyond the tools of personality psychology to other kinds of individual-level features will also permit us to ask a wide variety of social scientific questions within the Contained framework.

References

Bainbridge, W. S. 2007. The scientific research potential of virtual worlds. *Science* 317:472–476.

Barrick, M. R.; Stewart, G. L.; Neubert, M. J.; and Mount, M. K. 1998. Relating member ability and personality to work-team processes and team effectiveness. *Journal of applied psychology* 83(3):377.

Castronova, E.; Williams, D.; Shen, C.; Ratan, R.; Xiong, L.; Huang, Y.; and Keegan, B. 2009. As real as real? macroeconomic behavior in a large-scale virtual world. *New Media & Society* 11(5):685–707.

E. Castronova, T. L. Ross, I. K. 2013. Designer, analyst, tinker: How game analytics will contribute to science. *Game Analytics* 665–687.

I. Constantiou, M. F. Legarth, K. B. O. 2012. What are users' intentions towards real money trading in massively multiplayer online games? *Electron Markets* 22:105–115.

Judge, T. A., and Bono, J. E. 2000. Five-factor model of personality and transformational leadership. *Journal of applied psychology* 85(5):751.

McCrae, R. R., and Costa, P. T. 1987. Validation of the five-factor model of personality across instruments and observers. *Journal of personality and social psychology* 52(1):81.

Mueller, S.; abd Mubbasir Kapadia, B. S.; Frey, S.; Klingler, S.; Mann, R.; Sumner, R. W.; and Gross, M. 2015a. Heapcraft: Interactive data exploration and visualization tools for understanding and influencing player behavior in minecraft. In *Motion in Games)*, MIG'15.

Mueller, S.; Kapadia, M.; Frey, S.; Klingler, S.; Mann, R. P.; Solenthaler, B.; Sumner, R. W.; and Gross, M. 2015b. Heapcraft social tools: Understanding and improving player collaboration in minecraft. In *Foundations of Digital Games*, FDG'15.

Mueller, S.; Kapadia, M.; Frey, S.; Klingler, S.; Mann, R. P.; Solenthaler, B.; Sumner, R. W.; and Gross, M. 2015c. Statistical analysis of player behavior in minecraft. In *Foundations of Digital Games*, FDG'15.

Neuman, G. A.; Wagner, S. H.; and Christiansen, N. D. 1999. The relationship between work-team personality composition and the job performance of teams. *Group & Organization Management* 24(1):28–45.

Reviews, C. 2012. *e-Study Guide for: Clinical Psychology by Arthur Freeman, ISBN 9780471414995*. Cram101.

Sutton, R. 2013. *Social Psychology*. Palgrave Macmillan.