Escape to Reality: Gauging the Performance of Escape Room Groups

Danielle Lynn Curet

Follow this and additional works at: https://louis.uah.edu/honors-capstones

Recommended Citation
https://louis.uah.edu/honors-capstones/280

This Thesis is brought to you for free and open access by the Honors College at LOUIS. It has been accepted for inclusion in Honors Capstone Projects and Theses by an authorized administrator of LOUIS.
Escape to Reality: Gauging the Performance of Escape Room Groups

by

Danielle Lynn Curet

An Honors Capstone
submitted in partial fulfillment of the requirements
for the Honors Diploma
to
The Honors College
of
The University of Alabama in Huntsville
May 9, 2017
Honors Capstone Director: Dr. Allen W. Wilhite

Property rights reside with the Honors College, University of Alabama in Huntsville, Huntsville, AL
Honors Thesis Copyright Permission

In presenting this thesis in partial fulfillment of the requirements for an Honors Diploma from The University of Alabama in Huntsville, I agree that the Library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by my advisor or, in his/her absence, by the Chair of the Department, Director of the Program, or the Dean of the Honors College. It is also understood that due recognition shall be given to me and to The University of Alabama in Huntsville in any scholarly use which may be made of any material in this thesis.

Danielle Curet  
Student Name

Danielle L. Curet  
Student Signature

5/10/17  
Date
# Table of Contents

Acknowledgements .............................................................. 4  
Abstract ................................................................. 5  
I. Introduction ............................................................... 7  
II. Literature Review ..................................................... 11  
III. Observations of a Real-World Escape Room .................. 11  
   A. Data Collection and Methods ................................. 12  
   B. Analysis .......................................................... 16  
   C. Results ........................................................... 19  
IV. A Computational Model of Escape Rooms ...................... 21  
   A. Description of the Computational Model .................. 21  
      1. Parameters .................................................. 23  
      2. Objective .................................................... 24  
      3. Game Flow ................................................... 25  
   B. Results .......................................................... 25  
V. Conclusions & Findings .............................................. 26  
References ............................................................... 27
Acknowledgements

This project could not have been possible without the support and patience of so many respected individuals. I would like to acknowledge Dr. Allen Wilhite, the College of Business’ Department Chair of Accounting, Finance, and Economics and my project mentor, who provided guidance long before I was a student in his class; Dr. Robert F. Scherer, Interim Dean of the College of Business, who granted our group research funding; Jason and Jennifer Steinhauser, Escape Pod Owners and Designers, who were interested in our project and let my group observe players in the escape room without hesitation; Erin Conner, Sean Dafter, Ben Dissette, and Jameson Dozier, my group members for the observational study, who tolerated my incessant need to go above and beyond and spent many weekend hours observing the escape room and beta-testing; U.G. Wilson, brilliant Escape Pod Engineer; John Anglin, Katherine Johnson, and Charity Pounders, Managers and Room Technicians; Zac Dickson and Lexie Gilmore, Technicians, Emily Bearden, Interior Designer; David Mayo, beta-tester; and last, but certainly not least, Niall W. Rogers, beta-tester and helpful brainstormer.
Abstract

This Capstone project explores the behavior of players in a live escape room both empirically and computationally. Live escape rooms began as virtual game rooms where players could enter, hunt for clues, and solve puzzles. Today, there are over 6,881 live escape rooms worldwide and 2,958 in the U.S. alone, offering players the chance to lock themselves in a room for an hour with either strangers or people they know to try to beat the clock and escape. This project began as an observational study in order to discern the optimal group composition for solving and escaping a room based on a group's average age, male-to-female ratio, connectivity, and the technician's behavior. Although observational errors abound, groups of females of above average age that did not know each other prior to the game tended to have the fastest solve times.

Using the information gained from that empirical investigation, I constructed a computational model of an escape room to explore the relationship between puzzle type and difficulty and the creativity and expertise of the players. In the sections that follow, the theoretical, methodological, and practical considerations for this model will be discussed in turn, culminating in a presentation of results. In summary, groups with an above-average creativity level, a below-average expertise level, and a below-average stubbornness level had the quickest solve times, but further testing is necessary.
I. Introduction

“We have received intelligence from a reliable source that Soviet spies have infiltrated Redstone Arsenal and are attempting to locate Doctor Wernher Von Braun’s latest rocket schematics. If they are successful, I fear our arms race with the Soviets will escalate to doomsday levels... Please hurry, agents. They could strike at any moment, killing anyone who dares get in their way.” Such was then FBI director J. Edgar Hoover’s call to action inside Von Braun’s Office, a live escape room located at the Escape Pod in Huntsville, Alabama. My group and I were interested in finding out how we should compose our team in order to have the best chance of escaping, so we played the aforementioned escape room to familiarize ourselves with the game’s intellectual property and design our observation to answer this question.

It became clear after completing the empirical study that we were more interested in discerning whether players and groups with certain types of abilities solve puzzles faster than others. Unlike demographic traits such as age, gender, and relationships, ability is a trait that one cannot directly observe. Therefore, I constructed an agent-based model that allowed me to assign players and puzzles with certain ability and difficulty levels, respectively. An agent-based model (ABM) is “one of a class of computational models for simulating the actions and interactions of autonomous agents (both individual or collective entities such as organizations or groups) with the goal of assessing their effects on the system as a whole” (ScholarMap, FSU). Thanks to the increases in computing power and speed, we can now run flexible experiments as many times as is necessary for testing. Agent-based models differ from classical economic models in that they are built from the “bottom-up” rather than from the “top-down.” Because of this, we can observe how individual agents’ actions
change over time and maybe even discover some emergent properties that could not have been visible from a system containing aggregates. The ABM for this computational study was programmed in the open-source software NetLogo, and will be discussed following the presentation of the empirical study.

II. Literature Review

Modern escape room games are a relatively recent phenomenon that has become popular in the U.S. in the last ten years (Nicholson, 2016). While the origins and motivation for their emergence are unclear, it seems they are the physical version of mystery computer games in which players are caught in a room or must solve a series of puzzles before advancing through the game. Live escape rooms often involve a particular time period and theme, and as described in the introduction, the room we observed was set in the 1950s/60s, and we were FBI agents during the nuclear arms race against the Soviet Union. Scott Nicholson, Professor of Game Design and Development, wrote a paper entitled “The State of Escape: Escape Room Design and Facilities,” which presents some of the basic design features of escape rooms and trends in the industry. According to Nicholson’s survey, 25% of the escape rooms surveyed were set in present day (2000 - 2015), and the theme of 30% of the rooms was escaping a specific unpleasant place, such as a dungeon, prison, preschool, etc.

This project studies two crucial attributes of a successful escape: (i) teamwork and (ii) matching puzzle type with player abilities. Escape rooms are designed to be a team effort: individuals do not enter alone. Rather, players either book a full room with people they know, or individuals can sign up and will be placed on a team with people they do not
know. The necessity of teamwork to escape a room is one reason why corporate groups are a target market for escape room owners. In addition to being an entertaining activity, corporate testimonials claim that escape rooms serve as ice-breakers for their team and foster communication and teamwork. In practice, however, the level of teamwork is observed to initially be low as players begin to solve puzzles on their own. As the game progresses, however, instances arise in which two or more individuals will work together on a single puzzle, a situation that we explore in our computational model. This dynamic between familiar and unfamiliar players is interesting and, at times, unpredictable. It’s not always clear whether a group of strangers will communicate and work together more or better than a group of players that know each other. It is because of this interesting dynamic of familiarity (or lack thereof) that we devise the concept of a “connectivity score” that is described in much greater detail in section A. Data Collection and Methods.

Perhaps because it is still an emerging form of entertainment, very little academic literature has been written about escape rooms to date. In his survey of 175 escape room facilities worldwide, Professor Nicholson generated a list of the common escape room puzzle types, shown in Figure 1.

More on the topic of puzzle type, Business e-Coach Vadim Kotelnikov claims that all problems can be classified into one of four types:

1. The problem is evident and known to exist, but action is required: e.g. studying or losing weight
2. The problem is evident and known to exist, but expertise is required: e.g. math problems, spotting poisonous plant or animal species, etc.
3. The problem is evident and known to exist, but creativity is required: e.g. computer programming, other types of critical thinking
<table>
<thead>
<tr>
<th>What puzzle types are in the room?</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Searching for physical objects hidden in the room</td>
<td>78%</td>
</tr>
<tr>
<td>Team Communication</td>
<td>58%</td>
</tr>
<tr>
<td>Light</td>
<td>54%</td>
</tr>
<tr>
<td>Counting</td>
<td>53%</td>
</tr>
<tr>
<td>Noticing something &quot;obvious&quot; in the room</td>
<td>49%</td>
</tr>
<tr>
<td>Symbol substitution with a Key (such as looking symbols up in a book)</td>
<td>47%</td>
</tr>
<tr>
<td>Using something in an unusual way (Out-of-the-box thinking)</td>
<td>47%</td>
</tr>
<tr>
<td>Searching for objects in images</td>
<td>43%</td>
</tr>
<tr>
<td>Assembly of a Physical object (such as a jigsaw puzzle)</td>
<td>40%</td>
</tr>
<tr>
<td>Algebra and other Mathematics</td>
<td>39%</td>
</tr>
<tr>
<td>Pattern identification (such as visualizing a shape in a set of dots)</td>
<td>38%</td>
</tr>
<tr>
<td>Riddles</td>
<td>37%</td>
</tr>
<tr>
<td>Ciphers without a Key (such as letter substitution)</td>
<td>35%</td>
</tr>
<tr>
<td>Hearing</td>
<td>26%</td>
</tr>
<tr>
<td>Mirrors</td>
<td>26%</td>
</tr>
<tr>
<td>Abstract logic (such as Sudoku)</td>
<td>22%</td>
</tr>
<tr>
<td>Research using information sources</td>
<td>20%</td>
</tr>
<tr>
<td>Strategic thinking (such as Chess)</td>
<td>20%</td>
</tr>
<tr>
<td>Hand-eye Coordination (such as shooting a target)</td>
<td>17%</td>
</tr>
<tr>
<td>Rope or chains (such as undoing complex knots)</td>
<td>16%</td>
</tr>
<tr>
<td>Traditional Word Puzzles (such as crosswords or word search)</td>
<td>14%</td>
</tr>
<tr>
<td>Mazes</td>
<td>14%</td>
</tr>
</tbody>
</table>
4. The problem is hidden (not known to exist) and needs to be identified before it can be addressed, e.g. process inefficiencies, miscommunication, etc.

Once the types of puzzles in a room have been identified, they can be organized in the room in different ways. Escape room puzzle organization can be categorized in four ways: open, sequential, path-based, or a pyramid-like structure [see Figure 2].

Puzzles in a room with open puzzle organization can be solved in any order, and once all of them are solved, the game is complete. Puzzles laid out sequentially require the preceding puzzle be solved before the next one can be. Room designers occasionally use sequential puzzle organization to slow groups down early on in the game with difficult puzzles, but as the sequence continues, the puzzles become progressively easier. Path-based puzzle organization involves several independent processes that together provide the solution to the entire game or send the group to the final puzzle. Lastly, rooms with a pyramid-like puzzle organization have multiple path-based portions that lead to other path-based or sequential processes,
culminating in a final puzzle.

Both puzzle type and organization become important design parameters in the computational section of this paper (Section IV), but first, we focus on an actual escape room.

**III. Observations of a Real-World Escape Room**

**A. Data Collection and Methods**

The initial phase of this project was to get some hands-on experience with an escape room and to observe others make their way through such rooms. Our first stop was to a local escape room facility where my group and I played the game and debriefed each other on the particular puzzles, strategies, and our successes and failures while playing the game. Then, with permission from the owner, we set up a mechanism to observe and collect data on players. As customers entered the facility for their booking, we presented them with a Non-Disclosure Agreement asking for their consent to view their live video feeds along with the technician. (All players are monitored by a technician who watches for safety reasons, to prevent cheating or foul play, and to give hints as needed.) We also asked our participants to fill out a short survey and through these two means we were able to collect information on the length of time it took to solve each puzzle, the number of people that meaningfully contributed to solving each puzzle, whether they were male or female, and their age. We also used the survey information to measure the groups’ “connectivity,” a network measure that uses weighted links to calculate how closely the team members are related. Our link weighting system was as follows:
(a) This group is comprised of 3 couples that are all friends (like a triple date), with a connectivity score of \(\frac{3(3) + 12(2)}{6} = 5.5\) out of 10.

(b) This group is comprised of one couple whose husband knows three other women in the group, and one of those women had her husband with her as well. This group’s connectivity score is \(\frac{4(0) + 2(1) + 7(2) + 2(3)}{6} = 3.66667\) out of 10.

Figure 3: Two example network diagrams for groups we observed. Blue nodes are males and pink nodes are females. Players are labeled A through E. By our measure, Group (a) is more connected than Group (b) because they have a higher connectivity score. This makes sense intuitively and graphically because Group (b) has a player that only one person knows.

1 = acquaintances  
2 = friends  
3 = couples.married  
4 = family

To calculate their degree of closeness we then added up the weighted links and divided by the number people in the group to give us that group’s connectivity score. Figure 3 displays two examples of network diagrams and how they were used to calculate a group’s connectivity.

B. Analysis

The Von Braun Office is divided into three rooms, the Lobby, Office, and Back Room. Using the Nicholson (2016) puzzle organization classifications, this was a pyramid structure with independent paths in each room that can be solved in any order with an ultimate checkpoint puzzle that cannot be solved until all of the paths are complete. Figure 4 shows
the completion times for each puzzle in each of the three rooms and as one can see, the Office was the most challenging. Although it may not be entirely evident from the average solution times, the Back Room was easier than the Lobby. In fact, the benchmark most predictive of a group’s progress was whether they made it to the Bookcase puzzle (the last puzzle in the Office). A point of clarification: I say “made it to the bookcase” because it’s technically not a puzzle itself but rather a checkpoint for the entire game. In general, if groups made it to the bookcase puzzle with more than 10 minutes to spare, they are obviously skilled and are likely to breeze through the final puzzles.

![Average Solution Times by Puzzle](image)

Figure 4: Each unit on the x-axis lists the puzzles in order of appearance, and the dashed vertical lines separate puzzles into one of three rooms.

The primary measures of success defined in this study are the time it takes to reach the Bookcase puzzle and the overall percentage of puzzles solved (% Solve). We are interested in how those success measures are affected by team characteristics: mostly male or mostly female, the group’s connectivity, and the group’s age range. Table 1 displays our observations:
Average Time It Takes to Reach the Bookcase Puzzle by Group Demographic and Technician Gender

<table>
<thead>
<tr>
<th>Demographic</th>
<th>Female Tech</th>
<th></th>
<th>Male Tech</th>
<th></th>
<th>Overall</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Avg. Time</td>
<td>% Solve</td>
<td>Avg. Time</td>
<td>% Solve</td>
<td>Avg. Time</td>
<td>% Solve</td>
</tr>
<tr>
<td>Majority Female</td>
<td>0:50:06</td>
<td>71.43%</td>
<td>0:59:38</td>
<td>33.33%</td>
<td>0:51:42</td>
<td>60.00%</td>
</tr>
<tr>
<td>Equal M:F</td>
<td>0:26:49</td>
<td>100.00%</td>
<td>x</td>
<td>0.00%</td>
<td>0:26:49</td>
<td>20.00%</td>
</tr>
<tr>
<td>Majority Male</td>
<td>0:52:12</td>
<td>100.00%</td>
<td>0:38:52</td>
<td>50.00%</td>
<td>0:47:45</td>
<td>75.00%</td>
</tr>
<tr>
<td>High Connectivity</td>
<td>0:43:24</td>
<td>50.00%</td>
<td>0:49:15</td>
<td>33.33%</td>
<td>0:49:16</td>
<td>37.50%</td>
</tr>
<tr>
<td>Low Connectivity</td>
<td>0:47:29</td>
<td>87.50%</td>
<td>x</td>
<td>0.00%</td>
<td>0:47:29</td>
<td>63.64%</td>
</tr>
<tr>
<td>High Age Range</td>
<td>0:53:34</td>
<td>100.00%</td>
<td>0:49:15</td>
<td>40.00%</td>
<td>0:51:25</td>
<td>57.14%</td>
</tr>
<tr>
<td>Low Age Range</td>
<td>0:45:46</td>
<td>62.50%</td>
<td>x</td>
<td>0.00%</td>
<td>0:45:46</td>
<td>50.00%</td>
</tr>
<tr>
<td>Overall</td>
<td>0:48:07</td>
<td>80.00%</td>
<td>0:49:15</td>
<td>22.22%</td>
<td>0:48:01</td>
<td>52.63%</td>
</tr>
</tbody>
</table>

Table 1: The values in Column 1 show the average elapsed time to reach the bookcase puzzle for groups of a given demographic and technician gender. The percentages in Column 2 show the percentage of total puzzles solved by the group, i.e. for the 'High Connectivity' row, it took the groups with high connectivity and a female technician 43 minutes and 24 seconds on average to make it to the bookcase puzzle. Despite the short length of time it took to reach the bookcase puzzle, groups of this type only solved 50% of the puzzles in the game on average.

The bright yellow cells shown in Table 1 indicate inconclusive data, due to little or no data for groups of that type. This can be attributed to either not having observed groups of a particular composition at all or who failed to make it to the Bookcase Puzzle. For example, there were five groups that had an equal Male-to-Female Ratio. However, of those five, only one made it to the Bookcase. It also happened to be the only group of that composition to have a female technician. This may indicate that such groups are more likely to succeed with a female technician than a male technician, but more data would be required to draw
this conclusion. The cells highlighted to the right and bottom of the table represent broader conclusions which may be used as a point of reference for the remainder of the data. These represent wider categories with more data per composition group.

The first team characteristic deals with the male to female ratio (M:F), which we later converted to “male percentage.” Simply put, if there were more females in the group, the group was considered majority female, and the opposite for majority male. If the group had an equal number of males and females, they were put in the equal M:F group (50% male percentage). As explained above, of the few groups that had equal composition, only one reached the bookcase puzzle. As such, the conclusions for that group may not be terribly accurate. The next team characteristic relates to the connectivity score explained previously. Due to our method of calculating the Connectivity Score explained in A. Data Collection and Methods, the score fell in the range of 0 to 10. Groups with scores less than or equal to five were considered low connectivity, and those greater considered high connectivity. It may be concluded from this table that groups with low connectivity are more likely to solve puzzles up to this point than those with high connectivity, contrary to our hypothesis. We see the low connectivity result crop up again, and we provide some interpretations of that fact in Conclusions & Findings.

Age range was chosen as the final team characteristic because of its robustness against outliers. With groups as small as four and as large as seven, those outliers are extremely influential. The average between the highest age range and lowest was taken to form the median between low and high age ranges. Therefore, groups with ranges below 25 were placed in the lower category and those above in the higher. While groups with higher age ranges were slightly more likely to reach the Bookcase, groups with lower ranges tended to
solve it faster.

C. Results
In addition to analyzing the performance of groups of varying composition and assigned technician, we investigated the winning groups and the time it took them to escape the room. We began by plotting the independent variables for each group (mean age, standard deviation in age, connectivity score, and male percentage) against the dependent variable, puzzle duration, for each puzzle in seconds. This process yielded a total of 60 scatter plots (15 puzzles x 4 independent variables) that we could draw conclusions from based on their $\beta_1$ values (slopes of the trendlines). We summed up the slopes for all 15 puzzles separately for the 4 independent variables and plotted the total area per puzzle in Figure 5. This allowed us to compare the area above to the area below the x-axis to determine whether a given team characteristic had an overall positive or negative effect on puzzle duration. If the area above the x-axis is greater than the area below the x-axis, that signifies an overall positive trend for a given team characteristic correlated with puzzle duration, and vice versa for a greater sum below the x-axis.

![Sum of Mean Age Slopes](image)
Figure 5: Plotted above are slopes from simple linear regression of mean age, male percentage, connectivity score, and standard deviation in age, respectively, on puzzle duration in seconds. The positive and negative sums in the chart titles refer to whether the slope area (shown in blue) is primarily above or below the x-axis.

Our ultimate goal was to take these theoretical sums and compare them to the slopes of the “experimental” winning groups to determine if the trends for all groups are consistent with the trends of the winners. We hypothesize that groups with a high connectivity score
and a low standard deviation in age will possess the most advantageous initial strategies to solve and escape the room. We rationalize that well-connected groups will know more about each other and be able to ask the right person for help than a group of strangers. Our slope aggregates gave us values of -0.013 for mean age, 0.00296 for male percentage, 0.0266 for connectivity score, and 0.0839 for standard deviation in age. As you can see, the only team characteristic with an overall negative trend was mean age. An overall negative trend when plotted against puzzle duration indicates that groups with a higher mean age tend to solve a puzzle in a shorter amount of time, owing to a high y-intercept. The opposite is the case for an overall positive trend, giving us a prediction that groups with a high mean age, low male percentage, low connectivity score, and low standard deviation in age will be most likely to escape the room.

In case summing the slopes was skewed towards a large area for any given puzzle, we verified that we would get the same results if we had used the average of the slopes rather than just the sum. We know that the closer the average is to zero, the less of a pronounced trend we can conclude, either positive or negative. Our slope averages are converging on zero (mean age: -0.00086666, male percentage: 0.00197333, connectivity score: 0.00177333, and standard deviation in age: 0.00559333), but still illustrate the same findings found by summing the slopes. Included below are the graphs for the winning groups to show that the theoretical prediction matched the experimental results:
Below is the % error formula we used to compare the theoretical and experimental results:

\[
\% \text{ error} = \frac{|\text{theoretical value} - \text{experimental value}|}{\text{theoretical value}}
\]

<table>
<thead>
<tr>
<th></th>
<th>Exp</th>
<th>Theo</th>
<th>% Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Age</td>
<td>-0.0267</td>
<td>-0.013</td>
<td>-105.385</td>
</tr>
<tr>
<td>Male %</td>
<td>0.0016</td>
<td>0.00296</td>
<td>45.94595</td>
</tr>
<tr>
<td>Connect</td>
<td>0.0114</td>
<td>0.0266</td>
<td>57.14286</td>
</tr>
<tr>
<td>Std. dev.</td>
<td>0.0052</td>
<td>0.0839</td>
<td>93.80215</td>
</tr>
</tbody>
</table>

Table 2: Comparing Theoretical slopes (trend for all groups) to Experimental slopes (trendline slopes of the winners’ graphs above). The probability that the sign of the theoretical and experimental slopes were the same for all four team characteristics is \((\frac{1}{2})^2 = \frac{1}{16} = 6.25\%\).

IV. A Computational Model of Escape Rooms

The empirical study focused on the relationship between player demographics and team attributes on puzzle solving time. The computational section of this paper uses the information
gleaned from my observations of an actual escape room to create an agent-based computational model of escape rooms which can then be manipulated to explore the relationship between player abilities, team abilities and the puzzles in such rooms. A computational model is a computer program that creates a virtual environment (in this case an escape room) with artificial decision makers who react to that environment.

When conceptualizing an agent-based model, the most difficult task is finding an empirical basis for the parameters one is setting. One must take caution not to “program-in” the effects they wish to observe from an ABM. For example, since my empirical results suggested that females have an advantage in that particular escape room, does that mean I should explicitly assign a female advantage to agents in the model? As it turns out, this is erroneous practice and will ultimately bias the best solve times in favor of groups with a low male-to-female ratio. In order to create a theoretically sound and practically significant model, I needed to reference a combination of related empirical results as well as the ability to simplify while replicating reality as close as possible.

Although my empirical study was demographically focused, most of my research for constructing the ABM has been puzzle-centric: determining the type, difficulty, layout, and number of puzzles to include in the programmatic room. I used Kotelnikov’s problem types as a basis for creating the two types of puzzles in the model: creativity and expertise. Problems that simply require action, like those described in item 1 on page 8, were assumed to take a negligible amount of time to solve (less than one round), so they were not included in the model. The hidden problems described in item 4 on page 9 can be identified through teamwork and communication, a decision players are allowed to make.

Taking into consideration the various puzzle organizations mentioned in Nicholson’s sur-
vey, I chose to organize the puzzles openly for simplicity. An “open” puzzle organization means that puzzles can be solved simultaneously and do not become more challenging as the game progresses. The group’s solve rate increases over time due to the lower number of puzzles that remain to be solved each round and collaboration, which only occurs once the number of players exceeds the number of puzzles. Also, the puzzle difficulty and type are randomly generated each run so groups face a changing set of puzzles.

A. Description of the Computational Model

1. Parameters
The observer chooses an initial number of puzzles in the room between 7 and 30 which appear in fixed physical coordinates for every game. Puzzles have a type (creativity or expertise) and a difficulty level, which is a randomly generated value between 1 and 5. There are 5 randomly-spaced decoy puzzles in the room each game. Figure 6 depicts the computational model with 15 puzzles and 6 players as an example prior to any movement. The 6 players in the game (maximum team size for a live escape room) are endowed with levels of creativity and expertise that are randomly generated values between 0 and 1. These two parameters are set independently – thus, someone with a high creativity score could either have a high expertise score (and consequently be proficient at both kinds of puzzles), or a low expertise score and would therefore be somewhat of a specialist in solving creativity-type puzzles. Similarly, some individuals might be specialists in expertise-type puzzles, while others might have both low creativity and low expertise scores (and consequently be poor puzzle solvers overall).

Players also have a randomly-generated level of stubbornness that takes on a value be-
Figure 6: An example of how the interface looks prior to any movement. Yellow squares are the puzzles (there are 15 in this view), labeled either “C” or “E” to indicate puzzle type; the red squares are the decoy puzzles, labeled according to their “trick time,” i.e. the number of rounds a player works on that decoy puzzle until they realize it is a decoy and move on to the next puzzle. Players are labeled in white text with the coordinates of the patch (unit square) nearest them, what I’ve programmatically called their “target puzzle.”

The purpose of the stubbornness level is to model the players that refuse to ask for help and ultimately make the group worse off. The magnitude of the stubbornness level represents the amount of time that player will wait before he or she asks for help. For example, a stubbornness level of zero means a player will ask for help before even attempting to solve a puzzle, while a stubbornness level of 10 means that player will work on his or her own for ten rounds before calling someone over to help.

It is important to note that the stubbornness level does not become a factor in the model until the number of players exceeds the number of puzzles. Players solve puzzles independently until there is not a puzzle left for each person to solve on their own, at which point they ask for help according to their stubbornness level. Granted, if a player is the most qualified of the group to be working on a given puzzle (i.e. has the highest creativity
level and is working on a creativity puzzle, or vice versa), another person’s help will not be beneficial and the solver’s stubbornness doesn’t end up being a detriment to the group. An example of this in action is if a stubborn player is working on an algebra problem, but they are the best at algebra (could be the reason they are being stubborn), so that arrangement is most efficient. However, as we saw in Nicholson’s escape room survey, it is often the case that team communication simplifies an otherwise difficult puzzle, and a stubborn person would be reluctant to share clues and information – thus, would not benefit from teamwork.

2. Objective

A puzzle is solved by decrementing the puzzle’s difficulty each round by the player’s ability level for that puzzle type. For example, if a player with a creativity level of 0.25 begins solving a creativity puzzle with a difficulty of 1, it will take 4 rounds for that player to solve that puzzle on their own. A player’s proficiency in the other puzzle type (expertise to continue our example) is not an advantage for the type of puzzle they’re currently working on. A player remains at a puzzle until it is solved and then targets another unoccupied puzzle and begins moving toward it, one step per round.

Depending on the group, an individual player’s objective may conflict with that of the group. The group’s objective is to solve all puzzles as quickly as possible. This means that the team favors teamwork and information sharing. A stubborn individual, by contrast, is concerned with how many puzzles he can solve by himself. In pursuit of pride and glory, the stubborn individual impedes the group’s progress, misses out on any immediate help, and prevents his or her teammates from benefiting from his or her abilities. In general, when a
player is asked to help, they immediately leave the puzzle they are working on and all their progress is erased. Theoretically, a couple of players should be able to solve a puzzle more quickly than one player alone, but the maximum number of puzzles are not being worked at once. This trade-off is an inherent effect of the model, but players do not explicitly have the option to refuse to help a fellow player.

3. Game Flow

A player targets the puzzle nearest them and moves toward it one step each round. If one’s target puzzle is occupied, the player picks the next closest puzzle. Once a player has reached their target, they determine (sooner or later) if that puzzle is a decoy. Some decoys take longer to identify than others, but once a decoy has been discovered, it is erased from the game space, never to be mistaken for a real puzzle again. While the number of puzzles exceeds the number of players in the room, players solve puzzles on their own. This strategy is most efficient, allows players to physically “spread out,” and is generally in accordance with group behavior in practice.

Once there are no longer enough puzzles to go around, collaboration is necessary and inevitable. Players ask for help when their stubbornness dissipates, and progress is made toward the puzzle based on the highest ability level of the players for that puzzle type. The trouble is, group work is beneficial only 50% of the time, detrimental 40% of the time, and neutral 10% of the time (values to be tested). This implies that the effectiveness of the same pair or trio collaborating on a puzzle can vary from round to round. If group work is beneficial, a factor is subtracted from the puzzle’s difficulty. Conversely, if group work is detrimental, that same value is added to the puzzle’s difficulty. And finally, if group work is
neutral, nothing is added nor subtracted to the puzzle's difficulty and the puzzle is essentially solved by the person with the highest ability level present. Figure 7 displays an example of the model after a majority of the puzzles have been solved.

![Figure 7: An example of how the interface looks after 50 rounds. 6 puzzles remain, so players have started to collaborate, as exhibited by the 2 players with target puzzle 2, 10. One decoy puzzle remains that will delay an unlucky player by 4 rounds.]

**B. Results**

The model was run 10 times and the 3 groups with the fastest solve times had an average creativity level of 0.53, an average expertise level of 0.46, and an average stubbornness level of 4.2. Compared to the midpoint, these groups had an above-average creativity level, a below-average expertise level, and a below-average stubbornness level. The fast solve time for the below-average expertise level could be due to the initial distribution of creativity vs. expertise puzzles. The rest of the results make sense, however, should be tested using one-way ANOVA to determine if the variable values across the 3 winning groups are significantly different. Further, this model should be run several more times with different initial numbers of puzzles and the average puzzle difficulty should be calculated for each run.
V. Conclusions & Findings

In regards to the empirical study, we turned out to be correct in our hypothesis that groups with a low standard deviation in age would be the most strategically composed to escape the room, but our analysis showed that a low connectivity score was more advantageous than a high score. One possible interpretation of these findings is that diversity aids in problem solving, a message urged in virtually every aspect of daily life as well. Further statistical analysis could ascribe magnitudes to the “low” and “high” results, giving us a better picture with exact numbers. Multiple linear regression should also be explored to estimate a group’s solve time given their average age, male percentage, connectivity score, technician gender, and number of hints and type, to name a few.

We encourage our readers to remember that the empirical results contained herein are for the Von Braun Office located at the Escape Pod in Huntsville, AL and may not necessarily hold true for other escape rooms. This research can be applied to managerial decisions regarding project team compositions. A more serious application would be to predict how individuals and teams would behave in controlled simulation rooms in chaotic and/or traumatic situations such as search and rescue missions, live disease environments, or natural disaster response.

The computational model introduced some interesting hypotheses to test as well: How will different levels of stubbornness affect solve time given the same number of initial puzzles? What would a group’s optimal ability mix be if the room had a complex puzzle organization instead of an open one? Future research into modeling live escape rooms could explore puzzle organization, network effects, and/or the effects of different relationships of people working together.
References


The Escape Pod, live escape room at 3322 S Memorial Pkwy 705, Huntsville, AL 35801

The Escape Room Directory: http://escaperoomdirectory.com/


