The Role of Teamwork in Predicting Movie Earnings

by

Verica Buchanan

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved July 2016 by the
Graduate Supervisory Committee:

Nancy Cooke, Chair
Ross Maciejewski
Scotty Craig

ARIZONA STATE UNIVERSITY

August 2016

ABSTRACT

Intelligence analysts' work has become progressively complex due to increasing security threats and data availability. In order to study "big" data exploration within the intelligence domain the intelligence analyst task was abstracted and replicated in a laboratory (controlled environment). Participants used a computer interface and movie database to determine the opening weekend gross movie earnings of three pre-selected movies. Data consisted of Twitter tweets and predictive models. These data were displayed in various formats such as graphs, charts, and text. Participants used these data to make their predictions. It was expected that teams (a team is a group with members who have different specialties and who work interdependently) would outperform individuals and groups. That is, teams would be significantly better at predicting "Opening Weekend Gross" than individuals or groups. Results indicated that teams outperformed individuals and groups in the first prediction, under performed in the second prediction, and performed better than individuals in the third prediction (but not better than groups). Insights and future directions are discussed.

DEDICATION

I dedicate this work to my family. To my wonderful children, Lidija, Melissa, and Michael. May you never stop wondering.

To my encouraging and supporting husband, I could have not done this without your help!

ACKNOWLEDGMENTS

I sincerely thank my chair, Dr. Nancy Cooke for all her continuous support, guidance, and help. I thank my committee member Dr. Ross Maciejewski for his assistance with the movie interface and Dr. Scotty Craig for his help with the methods section.

I sincerely thank Dr. Nathan McNeese for being my mentor.

I sincerely thank Yafeng Lu and Michael Steptoe for their assistance with the movie interface and answering my numerous emails.

Finally, I sincerely thank Laís Gonçalves de Lima and Nikhil Patel for their help with data transcription and coding.

TABLE OF CONTENTS

Page

APPENDIX Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Intelligence analysts' work is becoming of increasing importance to our safety as citizens and security as a nation. Constantly shifting security needs coupled with the need for immediate intelligence production has analysts working around the clock at a faster pace. Analysts also face a challenging working environment due to purposeful deception and constant ambiguity. These challenges are amplified by the fact that analysts are bombarded with an onslaught of data. Sources of data vary greatly and can include, but are not limited to, social media, text documents, newspapers, audio, video, photographic images, maps, graphs, diagrams, equations, and others. Having to process all of this information is difficult to say the least. In order to develop new computational tools to better aid the analyst, experimental studies need to investigate ways to display information that is more intuitive, easier to understand, and quicker to process.

Intelligence analysis is also marked by an overall lack of teamwork and poor communication and coordination within and between intelligence organizations. Reasons for this vary, but contributions include organizational stove-piping, ignorance, and rivalry among intelligence agencies and analysts (Cooper, 2005; Treverton & Gabbard, 2008; Swenson, 2003). Additionally, current reward structures are centered around individual analysts' contributions such as the number of daily briefs produced, and not teamwork contributions (Johnston, 2005). However, teamwork in intelligence has been advocated as a means to reduce biases (Heuer, 1999; Straus, Parker, & Bruce, 2011). Teamwork also leads to more accurate forecasting of events (Mellers et al., 2015), and improves overall intelligence analysis (Richards & Pherson, 2010). However, there is a lack of empirical studies investigating team data exploration and predictions. This study examined how individuals, groups, and teams differ in "big" data exploration. In

particular, what are some of the benefits of working in teams (compared to working in groups or individually), and in what circumstances is the analyst better off working alone versus in a group or team (a team is a group with members who have different specialties and who work interdependently)? Previous work has indicated that in the domain of cyber security teamwork is superior to group or individual work (Cooke, Champion, Rajivan, & Jariwala, 2013; Rajivan et al., 2013). This study examined whether these results extend to intelligence analysis more generally.

BACKGROUND

Teamwork in Intelligence

Complex problems require teamwork, and failure to work together may lead to serious consequences. For instance, it is generally acknowledged that systematic communication, collaboration, and coordination failures among analysts, and as a whole, have contributed to some of the most recent intelligence failures – the 9/11terrorist attack and Iraqi weapons of mass destruction (Roth, Greenburg, & Wille, 2004; Senate Select Committee, 2004). However, even though the world has dramatically changed over the past 25 years, most of today's intelligence analysis is still conducted individually (Straus et al., 2011). This is an outdated technique given the large amount of data that analysts must cover. Individual analysts are expected to sift through an onslaught of data and understand intricate relationships between multiple countries, tribes, and even small isolated groups. Analysts must also consider other emerging factors, such as the proliferation of internet access, personal technology, and recently the influence of social

media on information propagation. Many more issues play a role in today's intelligence analysis forming a complex intertwined web of questions to be answered by the analyst. In light of this complexity, the world of intelligence is surpassing individual analysis and should move toward teamwork analysis (National Research Council, 2011).

Defining Teams

It is helpful to begin our discussion of teamwork in the context of intelligence analysis by defining what teams are, and essentially, what they are not. Literature reporting on intelligence issues often depicts analysts as socially interacting with one-another, working together, collaborating, and corroborating each other's analysis before the results are moved up the chain of command (Connors et al., 2004; Trent, Patterson, & Woods, 2007). Although these tasks can be performed by teams, they are not what defines them. Teams are more than a collection of analysts (individuals) working together (Paris, Salas, & Cannon-Bowers, 2000) because team members have specific roles, responsibilities, and work towards a common goal interdependently, adaptively, and dynamically (Salas, Dickinson, Converse, & Tannenbaum, 1992). Incorporating teamwork into the intelligence process would mean that analysts work on intelligence problems together, from the start, with each analyst fulfilling a different role, working on certain pieces of an analysis, developing expertise within a given problem statement, collaborating and coordinating through communication to assess intelligence issues. The added benefit of teamwork in the intelligence process is that it extends individual's capabilities by adding various abilities, skills, experiences and knowledge to the intelligence process (National Research Council, 2011).

Team Cognition

Intelligence analysts' work is predominantly cognitive in nature, and though most analysis is still conducted individually, teams of analysts need to be emerging due to an ever increasing complexity of security issues. Team cognition, defined as cognitive processes occurring at the team level (Salas, Cooke, & Rosen, 2008), is thus an important aspect of team intelligence analysis. In turn, gaining an understanding of teams' cognitive processes requires that we observe how teams learn, reason, problem solve, decide, or make judgments (Cooke, Gorman, Myers, & Duran, 2013). This means that in the context of intelligence analysis, team cognition provides us with valuable insights on how teams approach intelligence issues and analyze them. More importantly, by assessing the teams' cognitive processes we can parse out which processes lead to better assessments of intelligence issues and which ones do not. Furthermore, team cognition is closely tied to team performance (Cannon-Bowers & Salas, 2001; Cooke, Gorman & Winner, 2007). Hence, better cognitive processes on a team level should equate to superior intelligence analysis.

Interactive Team Cognition

Interactive Team Cognition (ITC), a theory proposed by Cooke, Gorman, Myers, & Duran (2013) states that team cognition is not a property or product, needs to be measured at the team level, and is closely tied to context. This view, contrary to the more traditional theories of team cognition, emphasizes cognition that develops through interactions among team members. Interactions, particularly team communication, is team cognition. Although this theory does not dismiss the importance of individual-

based knowledge, which is necessary, it does dismiss the notion that the sum of all individual-based knowledge equals team cognition. Through interactions, team members bounce ideas back-and-forth, learn from one another, gain new insights, and, more importantly, develop solutions to intricate problems. Therefore, team cognition unfolds in real time, as members communicate with one-another. In turn, measuring team cognition equates to monitoring team communication in context and, preferably, in real time.

OVERVIEW

There has been some progress towards teamwork within the Intelligence Community, but overall this effort seems to be limited. For instance, intelligence colleges may require their students to work together analyzing intelligence issues as part of their assignments (Landon-Murray, 2013; Wheaton, 2011); however, this generally does not translate to real world situations. Once graduated, most of the intelligence work is still conducted individually (National Research Council, 2011). Nonetheless, simply advocating the need for teamwork in Intelligence is not sufficient; one needs to demonstrate its associated benefits during the analytical process first.

For example, Mellers et al. (2015) conducted a large-scale, two year, geopolitical forecasting tournament in which 743 forecasters participated. These "analysts" generated more than 150,000 forecasts on 199 events, ranging from North Korea security concerns to the Greece debt crisis. In the first year of the experiment, the researchers manipulated group composition (independent forecasters, crowd-belief forecasters, and team forecasters) as well as training (probabilistic-reasoning training, scenario training, and no

training). During the second year, similar manipulations were conducted with again group compositions being one (team forecasters vs. independent forecasters) and training (probabilistic-reasoning training and no training) the other. One of their main findings during both years was that forecasting teams outperformed all other forecasters (independent forecasters and crowd-belief forecasters). The researchers attributed the teams' success to collaboration and knowledge sharing. They also hypothesized that working in teams increased motivation partly because team members may have not wanted to let each other down. Whereas other competencies played a significant part in ensuring successful analysis, it showed that, with all else being equal, collaborative teams outperformed all other forecasters. However, the researchers did not measure team cognition in order to understand precisely why teams outperformed independent forecasters or crowd-belief forecasters.

Another domain closely related to intelligence is cybersecurity. Cyber analysts, much like intelligence analysts, experience a high cognitive workload due to large amounts of data, inadequate teamwork, and time pressures (Champion, Rajivan, Cooke & Jariwala, 2012). Rajivan and colleges used a synthetic task environment, CyberCog, to replicate the tasks of cyber defense analysts (Rajivan et al., 2013). Participants were given real world intrusion alerts and asked to classify alerts as suspicious or benign. Some of these alerts were easier to analyze, whereas others were harder. In order to complete these tasks, the groups/teams had to access various information sources such as websites to accurately classify 225 alerts during a 30 minute session (two sessions total). The researchers found that teams significantly outperformed groups in classifying the difficult alerts only. Both, groups and teams, performed equally well at classifying the

easier alerts. Hence, this experiment illustrated that teamwork is especially beneficial when the task is difficult, problems are interrelated, and workload is high.

Similarly, the intelligence domain is full of complex issues with interrelated relationships combined with big data, much like the above mentioned study investigated. The current study used a team testbed to investigate whether the results from our previous studies can be extended to intelligence analysis more generally. Here, the researcher abstracted the analysts' tasks and replicated them in the lab (controlled environment). This allowed the researcher to investigate the domain of intelligence analysis without having to gain security clearances and access to analysts. Furthermore, intelligence literature indicates the need for more empirical studies within the intelligence domain, particularly when it comes to what role teamwork plays within the intelligence process (Folker Jr, 2000; Straus et al., 2011; Treverton & Gabbard, 2008).

In this study, the participants used visual analytics of social media data to determine the opening weekend gross movie earnings of four pre-selected movies developed by the Visual Analytics and Exploration Research (VADER) team at Arizona State University (Lu et al., 2014; Lu, Wang, & Maciejewski, 2014). As previously mentioned, analysts' tasks were abstracted. For example, participants were able to use their own knowledge about movies to start their investigation, much like analysts would start their intelligence analysis with previously acquired knowledge. Participants also needed to determine the significance of information, such as whether it mattered that the movie was a remake, part of a series, released on a holiday or during a peak season, has a trending star, etc. This part of the movie task would equate to analysts making judgments of what information to incorporate/exclude from their analysis. More importantly,

participants were expected to sift through a large amount of data. For example, participants examined the market cap and the movies released for the previous year of the week under analysis. They inspected the other weekends for similar movie releases to gauge the accuracy of the prediction range. They also evaluated the sentiment polarity of each movie's tweets to indicate a positive or negative reaction to the movie. Additionally, they determined which movies are most similar to the movie under analysis based on multiple indicators (Number of Tweets per day, Release Date & MPAA Rating, and Release Date & Movie Genre) for each sorting option and determined if movies released in the last three weeks were relevant predictors of the movie under analysis. Furthermore, these data were displayed in various formats such as graphs, charts, and text. Participants used these data to make their predictions. Therefore, the participants, much like 'real' analysts, explored "big" data, synthesized it, and predicted an outcome.

Other concepts that were explored were whether predictive models are useful during analysis, as well as meaningful ways to visualize "big" data. However, the main goal of this study was to investigate the role of teamwork during analysis. More accurately, how individuals, groups, and teamwork shape data exploration and accuracy of prediction. Hence, the study was directed at answering the questions: what are some of the benefits of working in teams (compared to working in groups or individually), and; in what circumstances is the analyst better off working alone versus in a group or team? The researcher hypothesized that teamwork would result in more accurate predictions than group and individual work.

METHOD

Design

The experiment used a 3x3 Mixed Factor design with Movie as a within-subject factor (Movie 1, 2, and 3) and Type of Unit as a between-subject factor (Individual, Group, and Team). The Movie task consisted of making predictions on three different movies. The primary dependent variable was performance. That is, accuracy of prediction; the difference between participants' predictions and the actual "Opening Weekend Gross" earnings amount. The Type of Unit consisted of three individuals working either individually, in a group, or as a team.

Participants

Ninety participants were recruited (thirty groups of three participants each). Participants were recruited through advertising by posting fliers on campus (ASU and local community colleges) and by listing the study on local listserve sites. Participants were compensated $10 per hour for their time. The experiment lasted around two hours. Out of the 90 participants, 68 were males (76%) and 22 females (24%). The average age of the participants was 24.1 years old (SD 6.2).

Equipment and Materials

Standard desk top computers were used for this study (see Appendix Q for a detailed description of equipment used). The training PowerPoint presentation was pre-loaded on the participants' computers. The movie software interface was accessed using

the internet browser 'Mozilla Firefox'. The movie interface consisted of the following main menus: the "Home", "Weekend Prediction", "Weekend Market Share", "Sentiment Analysis", "Movie Similarity", and the "Make Prediction" menu. The data presented in the movie interface included formerly gathered Twitter tweets of previously released movies and built in computational models. Movie data were represented visually (graphs, charts) as well as in text format (see Appendix J). Additionally, microphones attached to the participants' desks were used to record all verbal communication for further analysis.

The movies were presented in the same order for each experimental condition (Individual/Group/Team) throughout the entire experiment. First, the Disney movie *Frozen*, 2013 was used to illustrate interface capabilities during the training session. Second, the participants analyzed the data for the movie *Jack Ryan: Shadow Recruit*, 2014 in the practice session. Third, *The Hobbit: The Desolation of Smaug*, 2013 movie data was loaded automatically after the participants submitted their practice prediction. Forth, *About Last Night*, 2014 movie data automatically loaded once the prediction of the previous movie (Hobbit) was submitted by the participants. Lastly, the *Robocop*, 2014 movie data was automatically loaded after the participants submitted their previous prediction (About Last Night).

Procedure

Conditions

In the *Individual* condition the participants were trained on the whole interface: Home, Weekend Prediction, Weekend Market Share, Sentiment Analysis, Movie

Similarity, and Make Prediction. They were given all of the information and asked to make their predictions individually. The individual participants provided one estimate for each task, including the practice session, for a total number of four estimates for each participant (see appendix, P for the detailed experimental procedure).

In the *Group* condition the participants were trained on the whole interface: Home, Weekend Prediction, Weekend Market Share, Sentiment Analysis, Movie Similarity, and Make Prediction. They were given all of the information and asked to collaborate with each other to make their predictions as a group. The group provided one estimate for each task, including the practice session, for a total number of four estimates per group.

In the *Team* condition, participants were randomly assigned roles. Roles consisted of "Weekend Market Share Specialist" (computer station 1), "Sentiment Analysis Specialist" (computer station 2), and "Movie Similarity Specialist" (computer station 3). Training corresponded to their assigned role (expert domain). However, overall training content was not manipulated. That is, the same PowerPoint training material was used as in the individual/group condition, just broken up into three parts. Lastly, all team members were trained on the common features which are: "Home", "Weekend Prediction" and "Make Prediction" page. As in the group condition, the participants were asked to collaborate to make their predictions of opening gross movie earnings. The team provided one estimate for each task, including the practice session, for a total number of four estimates per team.

Training

Training was presented using PowerPoint (slides) and lasted approximately 30 minutes (self-paced). The training material covered the purpose of the study (to accurately predict gross movie earnings), how to navigate through the movie interface, what information is available to them and why this information is useful when predicting opening gross movie earnings. The Disney movie, *Frozen*, was used to illustrate all points made during training. A short quiz was administered after the training (10 questions, 5 minutes). The quiz was meant to test a few crucial pieces of information presented during training. The quiz was reviewed as a group giving the participants a chance to ask questions about the tested material. At the same time, the researcher provided feedback on the recently tested material.

Measures

Performance. The main dependent variable is accuracy of prediction; that is, how close are the participants' predictions to the actual "Opening Weekend Gross" earnings amount for the selected movies. The researcher hypothesized that teams will be more accurate in their "Opening Weekend Gross" earnings predictions than groups which should be better than individuals.

Hypothesis 1: Teamwork will result in more accurate predictions than individual and group work.

Team Process Ratings. Communication was coded in real time and audio recorded. Two process ratings sheets were used for the team and group conditions. The first one, Team Process Rating (Appendix, B), was used to assess the following: Are

participants actively sharing information, giving updates, and helping each other out. In particular, raters coded for: Information Seeking, Sharing of Information, Clarify/Explain Information, Acknowledgement, Giving Updates, Agreement/Consensus, Present Suggestions, Feedback, Planning, Help out, Praise, Repeated Requests (for same information), Negative Comments, and Argument. Additionally, team interactions were coded for "Organizational Strategies" (whether participants are organizing information presented in any particular way). The second process rating sheet, Team Communication Process Rating (please see Appendix A), was completed after each session and assessed the following: Overall quality of communication, group/team leader emergence, and agreement/disagreement on "Opening Weekend Gross" earnings amount. Finally, raters noted briefly how the group/team made decisions. That is, what strategies, if any, did the group/team use to make their decisions (weighing pros and cons, majority vote, compromising, etc.).

It was predicted that teams would exchange greater amounts of relevant information. For one, team members need to work together to cast their prediction because no one member has all the data. On the other hand, group members may spend more time 'chatting' about unrelated things because they can access all the data at the same time, and ,therefore, do not need to communicate crucial information to one another. Furthermore, team members should not be as cognitively overloaded because they do not need to process all data, hence they should have additional cognitive resources to collaborate, discuss, and estimate "Opening Weekend Gross" earnings amount.

Hypothesis 2: Teams will engage in significantly higher information exchanges than groups. In particularly, they will seek and share significantly more relevant information than groups.

Communication. The communication between the participants in the team and group conditions was coded for positive and negative exchanges of information relevant to team behavior found in the team literature (Burke, Salas, Wilson-Donnelly, & Priest, 2004; Cooke, & Gorman, 2009; Salas, Cooke, & Rosen, 2008;). For example, the following was noted: who talked to whom and why (Gorman, & Cooke, 2011), close-loop communication (exchanges of information and acknowledgment of received information; Cannon-Bowers, Tannenbaum, Salas & Volpe, 1995), shared understanding for the situation, feedback, and conflict resolution (Baker, 2003). When participants share information with their group/team the responses were coded by tallying these responses in the appropriate box, "Share Information". Thus, verbal communication was coded for team competencies and their frequency of occurrence. Additionally, an overall assessment of group/team communication was made at the end of each experimental session using a 7-point Likert scale (Team Communication Process Rating score sheet, appendix A for additional detail). Lastly, some inter-personal/team behaviors were also of interest. For example, it was noted whether groups/teams engage in supportive backup behavior and whether a team leader is emerging (Salas, Sims, & Burke, 2005) and why (due to better grasp of the data, ability to organize information, outgoing nature of the participant, etc.). The researcher also noted if the participants were gathered around a particular participant, and if so, why (e.g. because they are looking at a graph together, gathering around the team leader, etc.). These interactions, and other unexpected

behaviors, were noted in the "Additional Comment" section of the "Team Processes" rating score sheet (Appendix, B). The raters coded communication in real time, the audio recording was used as a back-up only.

Semi-structured Interviews. The experimenter interviewed participants at the end of the experiment and these discussions were recorded and transcribed. The participants were asked what data they used to make their decisions and predictions and why. Asking the participants directly led to a deeper understanding of how participants used the presented data to make their predictions. However, the interviews were seen as exploratory and no particular hypotheses are offered.

The interviews were audio recorded and coded using thematic analysis (Braun & Clarke, 2006). The following six-phases will be applied to the interview data: familiarization of the data, generation of initial codes, searching for themes, reviewing of themes, defining and naming themes, and producing the report (McNeese & Reddy, 2015).

Workload. NASA TLX (workload measure, please see Appendix D) – was administered twice during the experiment. It was administered the first time after the practice session (after the first prediction) and the second time after the experiment (after the fourth prediction was made). It was expected that teams will perceive the tasks as easier to complete than groups and individuals because team members did not have to

process all the data by themselves. Hence, the workload measure scores should reflect this.

Hypothesis 3: Teams will have significantly lower task demand scores and, at the same time, higher emotional scores (more secure, relaxed, etc.) than groups and individuals.

Demographics. Demographics questions were administered after NASA TLX was completed (see Appendix, E). Demographics questions covered participants' background (age, gender, education), domain knowledge (movie familiarity, frequency of "going to the movies"), social media usage (frequency), and knowledge of predictive analytics (familiarity with mathematical models).

Table 1. Summary of Measures

Measure	Indiv.	Group/Team	Means of Collection	Example
Performance (measured by accuracy of prediction)	yes	yes	Directly typed into the movie data base	$155, 789,654
Team & Group Interaction	no	yes	Team Process Rating	Information Seeking Sharing of Information Clarify/Explain Information Acknowledgement Giving Updates Agreement/Consensus
Team & Group Interaction	no	yes	Team Communication Process Rating	Overall quality of communication Group/team leader emergence Decision making process
Data Exploration	yes	yes	Semi-structured Interviews	Weekend Prediction Weekend Market Share Sentiment Analysis Movie Similarity
Workload	yes	yes	NASA TLX	Task was easy Task was demanding

Demographics	yes	yes	Questionnaire	Age Gender Background

Experimental Protocol

After obtaining informed consent, participants were randomly assigned to one of the three conditions (individual, group, or team). Also, they were randomly assigned to sit at one of the three computer stations (work station 1, 2, or 3). Following a brief overview of the study the participants were trained on the movie interface according to their assigned condition. During the training the researchers answered questions about the interface and content of the training, if asked by the participants. A short quiz (10 questions, 5 minutes, please see Appendix F) was administered after the training. The quiz was reviewed as a group once all participants completed it. Then, the participants were given a 10 minute break. During the break, the researchers loaded the movie interface on all participants' computers according to condition and opened the "calculator". The training PowerPoint remained open for the rest of the experiment so that the participants could refer to it throughout the experiment.

Participants were instructed to access the movie "Tutorial" once they returned from their break to familiarize themselves with the references available to them. Next, the participants took part in a 15-minute practice session. During the practice session participants were given the movie *Jack Ryan: Shadow Recruit* to research. At the end of the practice session the participants were given five minutes to finalize their findings and discuss it with their group/team members. In the individual condition the participants

were also be given five minutes to finalize their findings and estimation, but were not be prompted to collaborate with each other. Participants were prompted to give their estimate of "Gross Opening Weekend" earning once the 5 minutes were up. In the Individual condition, participants gave their movie earning estimate individually (three individual estimates), in the group condition as a group (one group estimate), and in the team condition as a team (one team estimate). The first workload measure (NASA TLX) was administered after the practice session. Additionally, participants were given feedback on how close they were to the actual "Gross Opening Weekend" earning's amount for the practice session only. Next, the participants completed three additional sessions lasting 20 minutes each (15 minutes to research the given movie and 5 minutes to finalize and discuss findings). The following movies were presented to the participants in the following order: *The Hobbit*, *About Last Night*, and *RoboCop*. After each scenario the participants were prompted to make their predictions.

The participants were asked to complete the second NASA TLX workload measure as soon as the last prediction was given. The demographic questionnaire was administered after the workload measure was completed. Next, the researcher conducted the short interviews. The interviews were conducted individually, one participant at the time. Once completed, the participants were paid for their time and fully debriefed on the purpose of the study (See Appendix, H). The researchers addressed participants' questions and then thanked them for their participation.

RESULTS

Performance Score

Performance on each movie prediction was quantified in terms of 1) the deviation from actual box office income (in terms of millions of dollars), and 2) relative absolute error calculated as the proportion of deviation from the actual income. The RAE values were used to determine if there was a significant difference in prediction among the three different groups (Individual/Group/Team) because these values best represent how much over/under the predictions were. Specifically, the RAE values were calculated as:

$$RAE = \left| \frac{Prediction - RealValue}{RealValue} \right|$$

Table 2. Summary statistics of the performance in terms of actual earnings amount and relative absolute error

		Difference to Actual OWG			RAE Actual OWG		
Subject Factor	**Movie**	**Mean**	**STDEV**	**STDERR**	**Mean**	**STDEV**	**STDERR**
Group	Hobbit	27.10	27.85	5.08	37.7%	37.0%	6.8%
Individual	Hobbit	13.32	29.30	5.35	33.7%	27.3%	5.0%
Team	Hobbit	8.14	15.15	2.77	19.2%	13.1%	2.4%
Group	About Last Night	-15.73	2.91	0.53	61.4%	11.5%	2.1%
Individual	About Last Night	-11.96	9.20	1.68	54.5%	21.6%	4.0%
Team	About Last Night	-19.25	2.06	0.38	75.0%	8.0%	1.5%
Group	Robocop	5.68	6.15	1.12	28.5%	25.9%	4.7%
Individual	Robocop	7.49	12.33	2.25	52.7%	40.1%	7.3%
Team	Robocop	0.51	6.72	1.23	24.5%	18.2%	3.3%

Table 2 provides summary statistics of the performance of each team that participated in the movie experiment. All three groups (Individual/Group/Team) over predicted the "Opening Weekend Gross" earnings amount for the first movie (Hobbit) by

$8.1 to $27.1 Mil, under predicted the second movie (About Last Night) by $12 to $19.3Mil, and over predicted the third movie (RoboCop) by $0.5 to $7.5Mil.

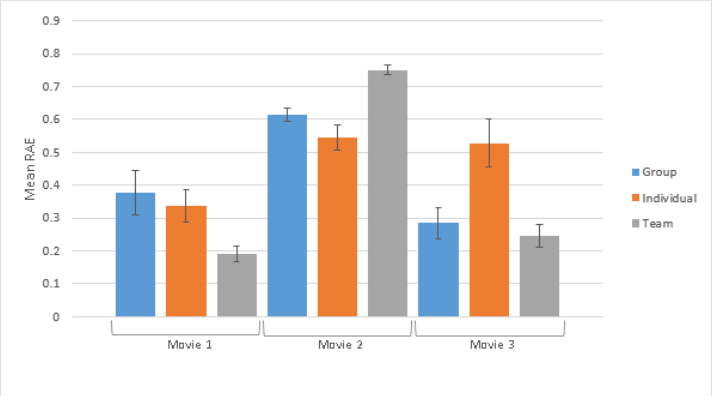


Figure 1. Mean RAE scores for the three groups by movie

Figure 1 shows the mean RAE scores for the three groups (Individual/Group/Team) by Movie (prediction of movie 1, 2, and 3). The biggest error rate was Movie 2 (prediction for About Last Night) at 0.64, followed by the *RoboCop* prediction (Movie 3) at 0.35. The lowest error rate was at 0.3 for the first Movie (the Hobbit prediction).

The Mauchly's Test of Sphericity test indicated that the assumption of sphericity have been violated, $x^2(2) = 10.371$, p=.006. The Greenhouse-Geisser correction was applied to the data. The 3x3 mixed ANOVA performed on these data revealed a significant main effect of Movie [$F(1.8, 156.25)=45.99$, $MSe= 3.26$, $p=.00$, partial $\eta_p^2 =.35$], but not a significant main effect of Type of Unit ($p>.1$). The main effect of Movie occurred because the RAE scores were significantly higher for Movie 2 (About Last

Night). The Movie x Type of Unit [$F(3.59, 156.25)=9.34$, $MSe= .66$, $p=.00$, partial $\eta_p^2 =.18$], interaction was significant.

To further understand the significant two-way interactions a one-way ANOVA was conducted. The assumption of homogeneity of variance was violated; therefore the Brown-Forsythe F-ratio was reported. There was a significant effect of Type of Unit for Movie 1 (Hobbit) RAE scores, $F(2, 61.57)= 3.73$, $p=.03$, a significant effect of Type of Unit for Movie 2 (About Last Night) RAE scores, $F(2, 53.22)=14.76$, $p=.00$, and a significant effect of Type of Unit for Movie 3 (RoboCop) RAE scores, $F(2, 62.94)=8.05$, $p=.00$. This occurred because the three groups (Individual/Group/Team) varied in accuracy of predictions for each movie.

In order to understand whether teams were more accurate than individuals or groups independent t-test comparisons were conducted. The p-value was corrected to $p<.025$. For the first "Opening Weekend Gross" prediction (Hobbit), there was a significant difference in RAE scores between Team (M=.19, SD=.13) and Individual (M=.34, SD=.27) condition, $t(41.63)=-2.62$, SE=.06, $p=.012$, d=-.71 as well as between Team (M=.19, SD=.13) and the Group (M=.38, SD=.37) condition, $t(36.12)=-2.58$, SE=.07, $p=.014$, d=-.69. The significant differences in teams' performance was observed with a large effect size. Lastly, the difference in RAE scores occurred because teams outperformed individuals and groups. That is, teams were closest to the actual "Opening Weekend Gross" earnings amount for the movie *Hobbit* whereas individuals and groups overestimated by a greater margin.

For the second "Opening Weekend Gross" prediction (About Last Night), there was a significant difference in RAE scores between Team (M=.75, SD=.08) and Individual (M=.55, SD=.22) condition, t(36.75)=4.88, SE=.04, *p*=.00, d=1.21 as well as between Team (M=.75, SD=.08) and the Group (M=.61, SD=.12) condition, t(51.57)= 5.31, SE=.03, *p*=.00, d=1.37. This difference in RAE scores occurred because both, individuals and groups, outperformed teams. That is, individuals and groups were closest to the actual "Opening Weekend Gross" earnings amount of the movie *About Last Night*. Overall, all three groups (Induvial/Group/Team) underestimated the actual "Opening Weekend Gross" earnings amount, however, teams significantly underestimated the earnings amount compared to the other two groups. Lastly, these significant differences were observed with a large effect size.

For the third "Opening Weekend Gross" prediction (RoboCop), there was a significant difference in RAE scores between Team (M=.25, SD=.18) and Individual (M=.53, SD=.40) condition, t(40.52)=-3.51, SE=.08, *p*=.001, d=-.90, but not between Team (M=.25, SD=.18) and the Group (M=.29, SD=.26) condition, t(52.04)=-.69, SE=.06, *p*=.49, d=-.18. The differences in scores occurred because both, groups and teams, outperformed individuals in this third, and last prediction. Here, groups and teams were closest to the actual "Opening Weekend Gross" earnings amount for the movie 'RoboCop'. Additionally, the significant differences between teams and individuals were observed with a large effect size.

Team Communication

Group and team communication was coded using the Team Process and Team Communication Process rating sheets. In order to ensure interrater reliability, a random sample of 5% coded data was used to run a Cohen's kappa. Interrater reliability for coding verbal data indicated a good agreement among the two raters with a Cohen's kappa of K=.76. However, out of all the communication measures, only one was significant. That is, teams (M=4.2, SD=2.74) sought out information significantly more than the Group (M=1.3 SD=1.98) condition, t(16.42)=2.71, SE=1.07, p=.01, d=1.21. Lastly, these results were achieved with a relative high effect size.

Workload

The NASA TLX workload measure of mental effort and task demands significantly differed across Type of Unit. There was a significant difference between Individual (M=-.55, SD=2.41) and Team (M=.62, SD=1.84) conditions, t(52.3)=2.08, SE=.56, p=.02, d=-.55 in their mental effort rating. Overall, individuals found that predicting movie earning amounts required more mental effort than teams. There was also a significant difference between Individual (M=-.59, SD=2.73) and Team (M=.48, SD=1.74) , t(47.6)=1.78, SE=.60, p=.04, d=-.47 in their task demands' rating. In this case, individuals found the tasks more demanding than participants working in teams.

Interview Question

There was a significant statistical difference in interface usage of the 'Movie Similarity' page between Team (M=3.59, SD=1.09) and Group (M=2.90, SD=1.59)

condition, t(49.48)=1.93, SE=.36, p=.03, d=.51 as well as between Team (M=3.59, SD=1.09) and Individual (M=2.31, SD=1.89) condition, t(44.65)=1.93, SE=.41, p=.00, d=.83. This difference was due to teams using the 'Movie Similarity' page more often than either the individuals or groups.

There was also a significant statistical difference in interface usage of the 'Weekend Market Share' page between Individual (M=1.45, SD=1.78) and Team (M=2.66, SD=1.76) condition, t(55.99)=2.59, SE=.47, p=.01, d=-.68. This difference was due to the fact that teams used the 'Weekend Market Share' page more often than the individual participants.

DISCUSSION

This study set out to explore whether teamwork led to more accurate assessments of data, and more importantly, to more accurate predictions. Although the experiment yielded mixed results–not all hypotheses were fully supported–it showed that working in teams has its benefits. For one, it can lead to more accurate predictions with less mental effort. Specially, teamwork resulted in more accurate predictions for *The Hobbit* (more accurate than individuals and groups) and *RoboCop* (more accurate than individuals only) thus partially supporting hypothesis 1. Additionally, Hypothesis 3 which predicted that teams would have significantly lower task demand scores was also partially supported. This indicated that teams found this task easier than individuals. However, teams did not differ in difficulty ratings between the groups. Lastly, there were no differences in how the three groups felt emotionally during the experiment.

A possible explanation for why teamwork did not result in better predictions during all three tasks could be that the interface did not fully support teamwork. That is, some of the data were easier to analyse than others. In particular, "Sentiment Analysis" (participant station 2), was easier to analyse than "Weekend Market Share" (station 1) and "Movie Similarity" (station 3). The "Sentiment Analysis" interface page consisted primarily of displaying Tweeter tweets (words), but did not have any additional graphs compared to "Weekend Market Share" (station 1) or "Movie Similarity" (station 3). This speculation is supported through some of the comments captured during the interviews. Participants in the team condition, station 2, stated that they felt as if they could not contribute enough to the team discussion compared to the other two participants. This

may have influenced the final data analysis. Future movie prediction studies should investigate how differences in task difficulty affect data exploration.

Another possible explanation for why teams underperformed predicting the second movie could be due to social decision making biases. In general, all participants were less familiar with the second movie, *About Last Night*, and in general, had a negative view towards the movie. Participants in the team condition may have relied more on each other's input to determine "Opening Weekend Gross" than the actual data. Moreover, outspoken team members, especially the ones who spoke negatively about this movie, may have gained a bigger audience than normal because individual team members lacked a "base" knowledge for this movie. However, it is not to say that participants in the Individual and Group condition knew more about the second movie, just that they may have relied more on the data to fill their knowledge gap. This of course was possible because individuals and group members were able to access all the data whereas individual team members could not. Lastly, participants in this experiment were overwhelmingly male (76%) with the smallest proportion of females in the Team condition. Again, this could have influenced data exploration for a romantic comedy such as *About Last Night*. In particular, male participants may have had a greater bias towards *About Last Night* across all three conditions, but due to the least number of females in the Team condition it may have happened to a higher degree.

Hypothesis 2 predicted that teams would engage in significantly more information exchanges than groups. In particular, they would seek and share relevant information significantly more than groups. This was only partially supported because teams sought

significantly more information, but did not share information significantly more. The communication data from this experiment revealed that in order for team members to work effectively together they must be able to request information from each other. More importantly, the communication data also revealed that group and team members were equally willing (or unwilling) to share information with each other. However, individual team members should have ideally shared more information with each other because they knew that not everyone had the same data. That extra step, sharing additional information, was not taken. Perhaps, fostering information sharing among participants could have led to an increase in information exchanges. One possible solution could be to provide participants with the ability to share (and seek) their data through some sort of collaborative tool.

Again, the communication results from this experiment indicated that it may be necessary to facilitate team data exploration through collaborative tools. This may be especially important in the intelligence domain in which analysts operate under constant time pressure. Working with others often requires additional time which analysts do not have. Collaborative tools, on the other hand, could make the process of sharing and requesting information among team members easier. However, it would be imperative to design collaborative tools that are easy, intuitive, and fluid to use because otherwise it would just add to the analysts' workload.

The interviews yielded another interesting result. As already stated, there was a significant statistical difference in interface usage of the 'Movie Similarity' page because teams used the 'Movie Similarity' page more often than either the individuals or groups. There was also a significant statistical difference in interface usage of the 'Weekend

Market Share' page. This difference was due to teams using the 'Weekend Market Share' page more often than the individual participants. It should be noted that both of these interface pages contain relatively large amounts of data compared to the rest of the interface. Hence, teams did not shy away of using these data sets during their analyses, whereas individuals and groups did. One possible explanation for these results could be that participants in the team condition were cognitively less overloaded than participants in the individual and group condition. In this case, fewer cognitive demands lead team members to analyse more complex data sets suggesting that working in teams can be beneficial during "big" data exploration.

Lastly, this study was developed with the theory of Interactive Team Cognition in mind. Particularly, that monitoring team/group communication would allow the researcher to gain an understanding of how groups/teams reasoned through their data analysis and how they decided on the "Opening Weekend Gross" prediction amount. For example, numerous groups/teams reached consensus by working together on their prediction estimate whereas others met in the middle (between each individual's estimated amount). An example, of a poor performing team leads us to see how some of the worst performing teams made decisions. Here, team members spent a considerable amount of time arguing with each other, insisting that each one's prediction was correct, missing out on the actual data analysis. This poor performing team provided a prime example of how being focused on the wrong thing, insisting they individually are right, lead to poor prediction estimates. However, one of the most interesting findings was that only one team (out of all teams and all groups) planned their "attack". That is, these team members decided ahead of time the sequence of their data discussion (what data to

discuss first, second, etc.), and how they will work together and share information. This finding becomes even more interesting when considering that this team was the best team in predicting "Opening Weekend Gross" for movie 2, *About Last Night.* This seems to suggest that in "big" data exploration planning matters, especially when things are unknown (in this case, unfamiliar movie). All in all, communication data can be a useful tool in assessing group/team decision making processes and in turn, gaining an understanding of what particular behaviours lead to more accurate predictions.

Study Limitations

A major limitation of this study is that the movies were presented in the same, fixed order. That is, the first movie was always The *Hobbit: The Desolation of Smaug*, the second *About Last Night*, and the third *RoboCop*. This may have influenced the results. For example, participants' fatigue could have played a role during data analysis. During the first analysis participants could have been more rested, therefore paying closer attention to the data, and hence, making a better prediction. During the second session participants could have experienced fatigue resulting in a worse prediction. Lastly, participants could have had a last surge in energy resulting in focused data analysis and, at the same time, leading to a better prediction.

CONCLUSION

This thesis work has multiple implications. Above all, this study illustrated that it is difficult to make predictions. In light of this, one can only imagine what today's analysts face. Surrounded by an ever-increasing availability of data coupled with an ever-shrinking time frame for their analysis, analysts are still expected to accurately assess and predict future security concerns. One would think that even the best analysts struggle under such constraints. Therefore, it is imperative for researchers to continue exploring avenues of predictive analytics. Future direction should focus on increasing the complexity of data to replicate more accurately what analysts are faced day-to-day. Additionally, this work illustrates that teamwork can be beneficial during data exploration and can lead to more accurate predictions. Lastly, this research adds to the overall knowledge of team science as well as predictive analytics.

REFERENCES

Barshi, I. (2015). From Healy's Training Principles to Training Specifications: The Case of the Comprehensive LOFT. The American journal of psychology,128(2), 219-227.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101.

Burke, C. S., Salas, E., Wilson-Donnelly, K., & Priest, H. (2004). How to turn a team of experts into an expert medical team: guidance from the aviation and military communities. Quality and Safety in Health Care, 13(suppl 1), i96-i104.

Butler, A. C., Karpicke, J. D., & Roediger III, H. L. (2008). Correcting a metacognitive error: feedback increases retention of low-confidence correct responses. Journal of Experimental Psychology: Learning, Memory, and Cognition, 34(4), 918.

Cannon-Bowers, J. A., & Salas, E. (2001). Reflections on shared cognition. Journal of Organizational Behavior, 22(2), 195-202.

Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E., & Volpe, C. E. (1995). Defining competencies and establishing team training requirements. Team effectiveness and decision making in organizations, 333, 380.

Champion, M. A., Rajivan, P., Cooke, N. J., & Jariwala, S. (2012). Team-based cyber defense analysis. *Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), 2012 IEEE International Multi-Disciplinary Conference on,* 218-221.

Connors, E. S., Craven, P. L., McNeese, M. D., Jefferson, T. Jr., Bains, P., & Hall, D. L. (2004). An Application Of The Akadam Approach To Intelligence Analyst Work. Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting, 627-628.

Cooke, N. J., Champion, M., Rajivan, P., & Jariwala, S. (2013). Cyber Situation Awareness and Teamwork. EAI Endorsed Transactions on Security and Safety. Special Section on: The Cognitive Science of Cyber Defense, 13.

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive team cognition. Cognitive Science, 37(2), 255-285.

Cooke, N. J., & Gorman, J. C. (2009). Interaction-based measures of cognitive systems. Journal of Cognitive Engineering and Decision Making,3(1), 27-46.

Cooke, N. J., Gorman, J. C., & Winner, J. L. (2007). Team cognition. Handbook of Applied Cognition, , 239-268.

Cooper, J. R. (2005). Curing analytic pathologies: Pathways to improved intelligence analysis. CENTRAL INTELLIGENCE AGENCY WASHINGTON DC CENTER FOR STUDY OF INTELLIGENCE.

Folker Jr, R. D. (2000). Intelligence analysis in theater joint intelligence centers: An experiment in applying structured methods. JOINT MILITARY INTELLIGENCE COLL WASHINGTON DC CENTER FOR STRATEGIC INTELLIGENCE RESEARCH.

Gorman, J. C., & Cooke, N. J. (2011). Changes in team cognition after a retention interval: the benefits of mixing it up. Journal of Experimental Psychology: Applied, 17(4), 303.

Hattie, J., & Timperley, H. (2007). The power of feedback. Review of educational research, 77(1), 81-112.

Hattie, J. (1999). Influences on student learning. Inaugural lecture given on August, 2, 1999.

Healy, A. F., Kole, J. A., & Bourne Jr, L. E. (2014). Training principles to advance expertise.

Heuer, R. J. (1999). Psychology of intelligence analysis. Lulu. com.

Richards, H. J., & Pherson, R. H. (2010). Structured analytic techniques for intelligence analysis. CQ Press.

Johnston, R. (2005). Analytic culture in the US intelligence community: An ethnographic study. Washington, DC: Center for the Study of Intelligence.

Krause, U. M., Stark, R., & Mandl, H. (2009). The effects of cooperative learning and feedback on e-learning in statistics. Learning and Instruction,19(2), 158-170.

Landon-Murray, M. (2013). Moving US Academic Intelligence Education Forward: A Literature Inventory and Agenda. International Journal of Intelligence and Counter Intelligence, 26(4), 744-776.

Lu, Y., Kruger, R., Thom, D., Wang, F., Koch, S., Ertl, T., & Maciejewski, R. (2014, October). Integrating predictive analytics and social media. In Visual Analytics Science and Technology (VAST), 2014 IEEE Conference on (pp. 193-202). IEEE.

Lu, Y., Wang, F., & Maciejewski, R. (2014). Business intelligence from social media: A study from the vast box office challenge. Computer Graphics and Applications,

IEEE, 34(5), 58-69.

McNeese, N. J., & Reddy, M. C. (2015). The role of team cognition in collaborative information seeking. Journal of the Association for Information Science and Technology.

Mellers, B., Stone, E., Atanasov, P., Rohrbaugh, N., Metz, S. E., Ungar, L., ... & Tetlock, P. (2015). The psychology of intelligence analysis: Drivers of prediction accuracy in world politics. Journal of experimental psychology: applied, 21(1), 1.

National Research Council. (2011). Intelligence Analysis for Tomorrow: Advances from the Behavioral and Social Sciences. Committee on Behavioral and Social Science Research to Improve Intelligence Analysis for National Security, Board on Behavioral, Cognitive, and Sensory Sciences, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Paris, C. R., Salas, E., & Cannon-Bowers, J. A. (2000). Teamwork in multi-person systems: a review and analysis. Ergonomics, 43(8), 1052-1075.

Rajivan, P., Champion, M., Cooke, N. J., Jariwala, S., Dube, G., & Buchanan, V. (2013). Effects of Teamwork versus Group Work on Signal Detection in Cyber Defense Teams. In Foundations of Augmented Cognition (pp. 172-180). Springer Berlin Heidelberg.

Roth, J., Greenburg, D., & Wille, S. (2004). National Commission on Terrorist Attacks Upon the United States. Monograph on Terrorist Financing: Staff Report to the Commission.

Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. Human Factors: The Journal of the Human Factors and Ergonomics Society, 50(3), 540-547.

Salas, E., Dickinson, T. L., Converse, S. A., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. Teams their Training and Performance, 3-29.

Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a “Big Five” in teamwork?. Small group research, 36(5), 555-599.

Straus, S. G., Parker, A. M., & Bruce, J. B. (2011). The group matters: A review of processes and outcomes in intelligence analysis. Group Dynamics: Theory, Research, and Practice, 15(2), 128.

Swenson, R. G. (2003). Bringing intelligence about: Practitioners reflect on best practices. Lulu. com.

Trent, S. A., Patterson, E. S., & Woods, D. D. (2007). Challenges for cognition in

intelligence analysis. Journal of Cognitive Engineering and Decision Making, 1(1), 75-97.

Treverton, G. F., & Gabbard, B. C. (2008). Assessing the tradecraft of intelligence analysis. Rand Corporation.

Wheaton, K. J. (2011). Teaching strategic intelligence through games. International Journal of Intelligence and CounterIntelligence, 24(2), 367-382.

United States Congress, Intelligence Reform and Terrorism Prevention Act of 2004, 108th Congress, 2nd Session, 20 January 2004.

APPENDIX A

3X3 MIXED FACTOR ANOVA ANALYSIS RESULTS

3x3 Mixed Factor ANOVA Analysis Results

General Linear Model

Within-Subjects Factors

Measure: RAE

Movie	Dependent Variable
1	Task_1_RAE
2	Task_2_RAE
3	Task_3_RAE

Between-Subjects Factors

		N
Type	Group	30
	Individual	30
	Team	30

Descriptive Statistics

	Type	Mean	Std. Deviation	N
Task_1_MRAE	Group	.3770	.37003	30
	Individual	.3367	.27291	30
	Team	.1920	.13063	30
	Total	.3019	.28430	90
Task_2_MRAE	Group	.6140	.11545	30
	Individual	.5447	.21646	30
	Team	.7500	.07983	30
	Total	.6362	.17042	90
Task_3_MRAE	Group	.2850	.25936	30
	Individual	.5273	.40065	30
	Team	.2450	.18236	30
	Total	.3524	.31748	90

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[d]
Movie	Pillai's Trace	.614	68.450[b]	2.000	86.000	.000	.614	136.899	1.000
	Wilks' Lambda	.386	68.450[b]	2.000	86.000	.000	.614	136.899	1.000
	Hotelling's Trace	1.592	68.450[b]	2.000	86.000	.000	.614	136.899	1.000
	Roy's Largest Root	1.592	68.450[b]	2.000	86.000	.000	.614	136.899	1.000
Movie * Type	Pillai's Trace	.391	10.585	4.000	174.000	.000	.196	42.339	1.000
	Wilks' Lambda	.630	11.195[b]	4.000	172.000	.000	.207	44.781	1.000
	Hotelling's Trace	.555	11.799	4.000	170.000	.000	.217	47.195	1.000
	Roy's Largest Root	.487	21.181[c]	2.000	87.000	.000	.327	42.363	1.000

a. Design: Intercept + Type
Within Subjects Design: Movie

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

d. Computed using alpha = .05

Mauchly's Test of Sphericity[a]

Measure: RAE

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Movie	.886	10.371	2	.006	.898	.937	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + Type
Within Subjects Design: Movie

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tests of Within-Subjects Effects

Measure: RAE

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Movie	Sphericity Assumed	5.846	2	2.923	45.990	.000	.346	91.980	1.000
	Greenhouse-Geisser	5.846	1.796	3.255	45.990	.000	.346	82.597	1.000
	Huynh-Feldt	5.846	1.874	3.120	45.990	.000	.346	86.167	1.000
	Lower-bound	5.846	1.000	5.846	45.990	.000	.346	45.990	1.000
Movie * Type	Sphericity Assumed	2.374	4	.593	9.338	.000	.177	37.352	1.000
	Greenhouse-Geisser	2.374	3.592	.661	9.338	.000	.177	33.542	.999
	Huynh-Feldt	2.374	3.747	.634	9.338	.000	.177	34.991	.999
	Lower-bound	2.374	2.000	1.187	9.338	.000	.177	18.676	.975
Error(Movie)	Sphericity Assumed	11.059	174	.064					
	Greenhouse-Geisser	11.059	156.250	.071					
	Huynh-Feldt	11.059	163.002	.068					
	Lower-bound	11.059	87.000	.127					

a. Computed using alpha = .05

Tests of Within-Subjects Contrasts

Measure: RAE

Source	Movie	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Movie	Linear	.115	1	.115	1.358	.247	.015	1.358	.211
	Quadratic	5.731	1	5.731	135.124	.000	.608	135.124	1.000
Movie * Type	Linear	.599	2	.300	3.538	.033	.075	7.077	.644
	Quadratic	1.775	2	.887	20.921	.000	.325	41.841	1.000
Error(Movie)	Linear	7.369	87	.085					
	Quadratic	3.690	87	.042					

a. Computed using alpha = .05

Tests of Between-Subjects Effects

Measure: RAE

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Intercept	49.966	1	49.966	857.930	.000	.908	857.930	1.000
Type	.249	2	.124	2.136	.124	.047	4.273	.427
Error	5.067	87	.058					

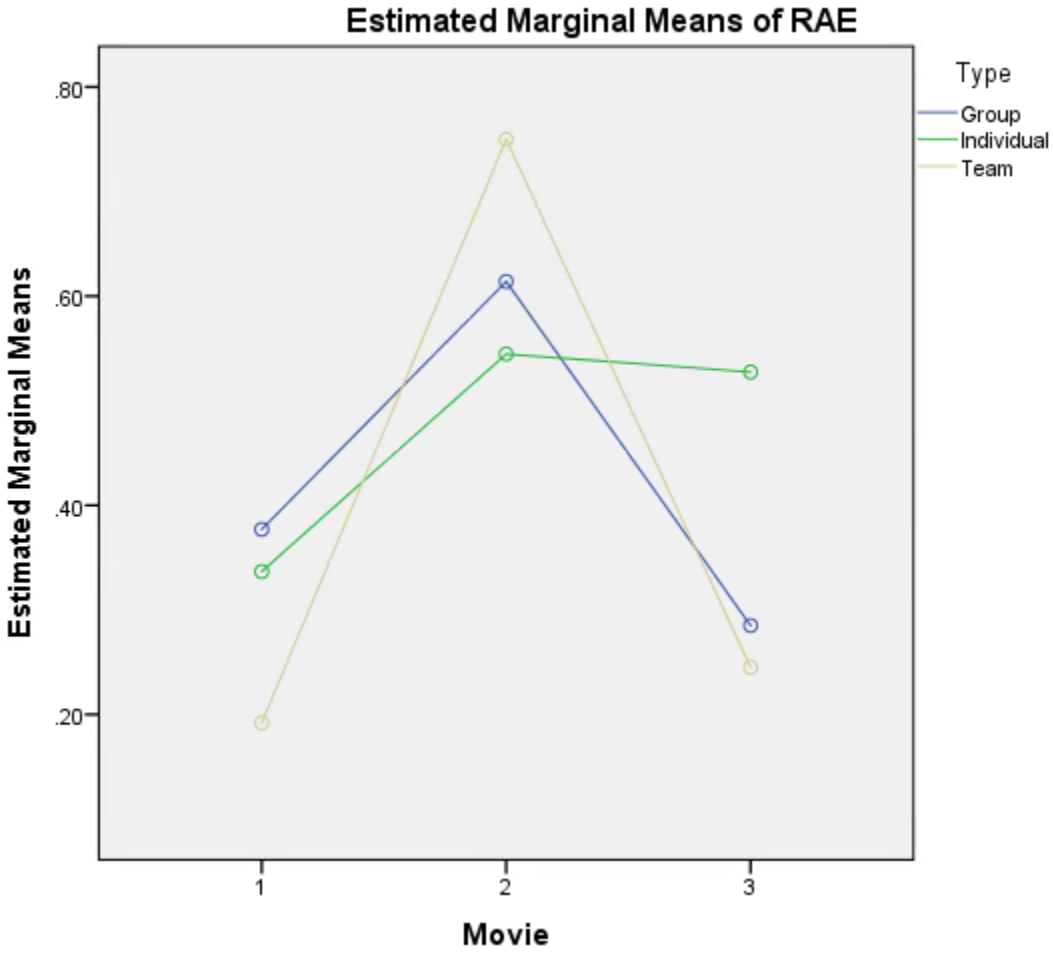
a. Computed using alpha = .05

Estimated Marginal Means

Grand Mean

Measure: RAE

Mean	Std. Error	95% Confidence Interval	
		Lower Bound	Upper Bound
.430	.015	.401	.459


Estimated Marginal Means of RAE
Estimated Marginal Means
.80
.60
.40
.20
1
2
3
Movie
Type
Group
Individual
Team

APPENDIX B

ONE-WAY ANOVA ANALYSIS RESULTS

One-Way ANOVA Analysis Result – Movie 1

ANOVA

Task_1_RAE

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.568	2	.284	3.728	.028
Within Groups	6.626	87	.076		
Total	7.193	89			

Robust Tests of Equality of Means

Task_1_RAE

	Statistic[a]	df1	df2	Sig.
Welch	5.834	2	48.859	.005
Brown-Forsythe	3.728	2	61.567	.030

a. Asymptotically F distributed.

Oneway

Descriptives

Task_1_RAE

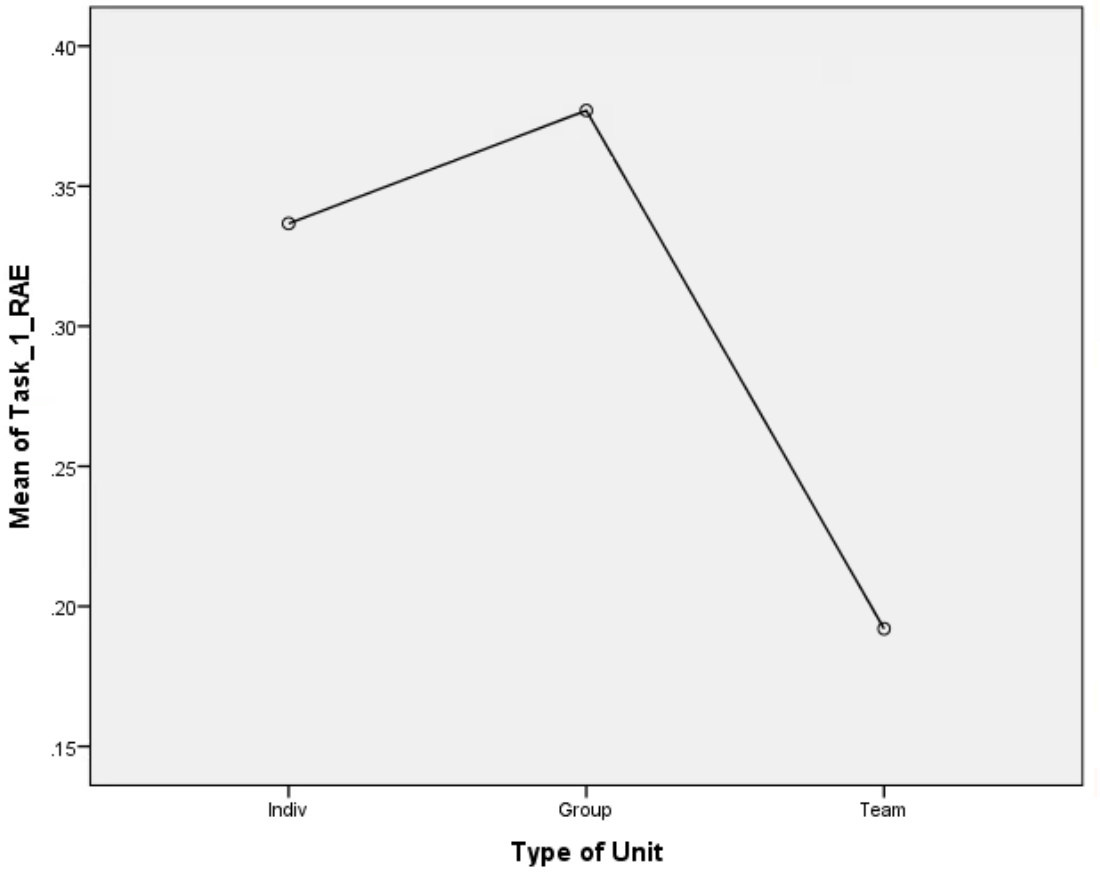
	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	30	.3367	.27291	.04983	.2348	.4386	.02	.85
2	30	.3770	.37003	.06756	.2388	.5152	.05	1.04
3	30	.1920	.13063	.02385	.1432	.2408	.03	.44
Total	90	.3019	.28430	.02997	.2423	.3614	.02	1.04

Test of Homogeneity of Variances

Task_1_RAE

Levene Statistic	df1	df2	Sig.
18.082	2	87	.000

Means Plots


.40
.35
.30
.25
.20
.15
Mean of Task_1_RAE
Indiv
Group
Team
Type of Unit

One-Way ANOVA Analysis Result – Movie 2

Oneway

Descriptives

Task_2_RAE

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	30	.5447	.21646	.03952	.4638	.6255	.07	.93
2	30	.6140	.11545	.02108	.5709	.6571	.44	.79
3	30	.7500	.07983	.01457	.7202	.7798	.59	.86
Total	90	.6362	.17042	.01796	.6005	.6719	.07	.93

Test of Homogeneity of Variances

Task_2_RAE

Levene Statistic	df1	df2	Sig.
15.645	2	87	.000

ANOVA

Task_2_RAE

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.655	2	.327	14.755	.000
Within Groups	1.930	87	.022		
Total	2.585	89			

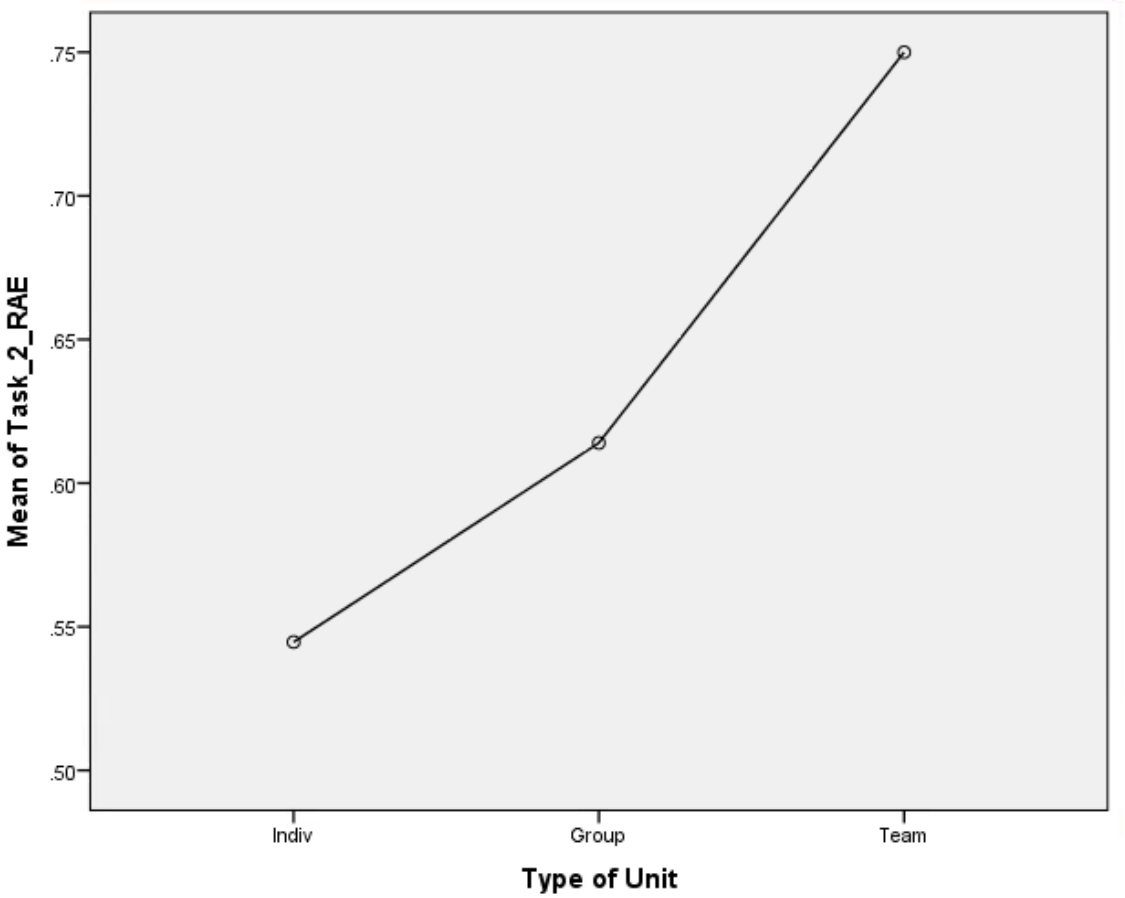
Robust Tests of Equality of Means

Task_2_RAE

	Statistic[a]	df1	df2	Sig.
Welch	21.441	2	52.300	.000
Brown-Forsythe	14.755	2	53.223	.000

a. Asymptotically F distributed.

Means Plots


.75
.70
.65
.60
.55
.50
Mean of Task_2_RAE
Indiv
Group
Team
Type of Unit

One-Way ANOVA Analysis Result – Movie 3

ANOVA

Task_3_RAE

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.400	2	.700	8.047	.001
Within Groups	7.570	87	.087		
Total	8.970	89			

Robust Tests of Equality of Means

Task_3_RAE

	Statistic[a]	df1	df2	Sig.
Welch	6.106	2	53.531	.004
Brown-Forsythe	8.047	2	62.940	.001

a. Asymptotically F distributed.

Oneway

Descriptives

Task_3_RAE

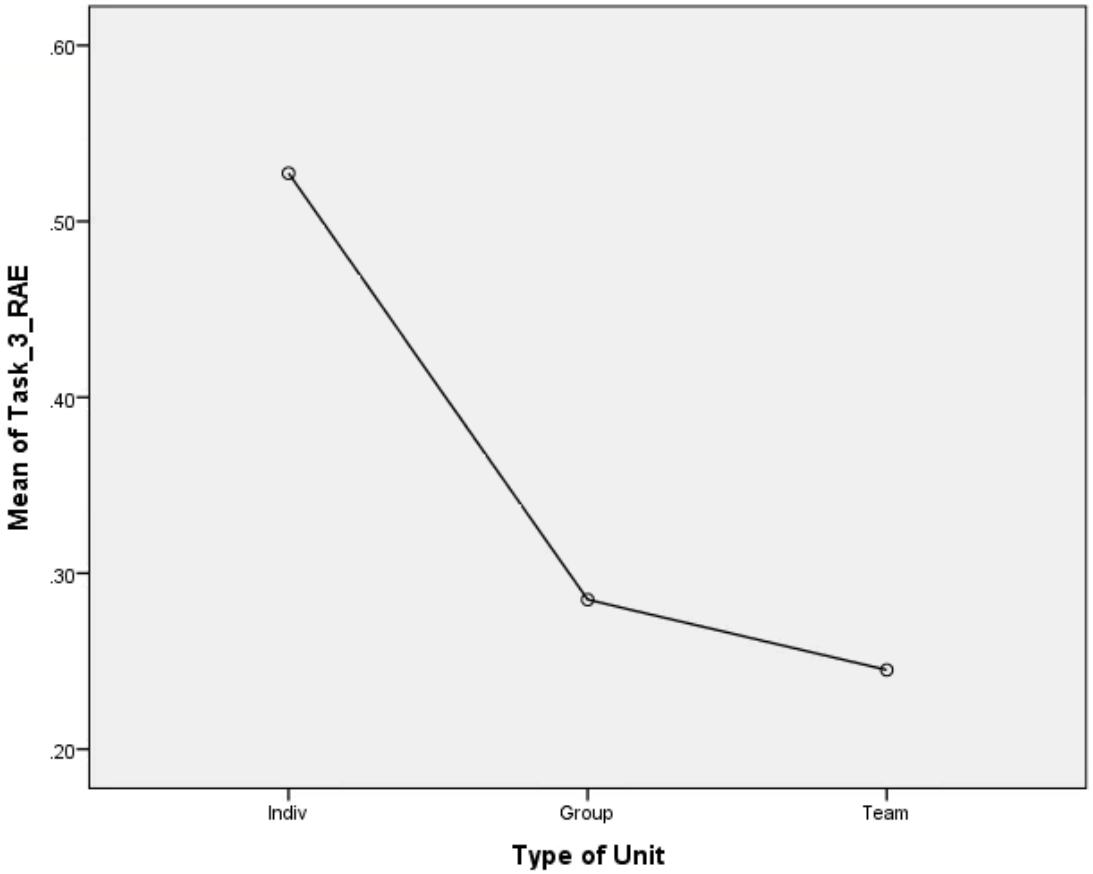
	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	30	.5273	.40065	.07315	.3777	.6769	.00	1.83
2	30	.2850	.25936	.04735	.1882	.3818	.01	.94
3	30	.2450	.18236	.03329	.1769	.3131	.01	.61
Total	90	.3524	.31748	.03347	.2860	.4189	.00	1.83

Test of Homogeneity of Variances

Task_3_RAE

Levene Statistic	df1	df2	Sig.
5.683	2	87	.005

Means Plots



APPENDIX C

INDEPENDENT T-TEST ANALYSIS RESULTS

Independent t-Test Analysis Results

T-Test

Group Statistics

	Group	N	Mean	Std. Deviation	Std. Error Mean
Task_1_RAE	3	30	.1920	.13063	.02385
	1	30	.3367	.27291	.04983

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Task_1_RAE	Equal variances assumed	23.257	.000	-2.619	58	.011	-.14467	.05524	-.25524	-.03409
	Equal variances not assumed			-2.619	41.626	.012	-.14467	.05524	-.25617	-.03316

T-Test

Group Statistics

	Group	N	Mean	Std. Deviation	Std. Error Mean
Task_1_RAE	3	30	.1920	.13063	.02385
	2	30	.3770	.37003	.06756

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Task_1_RAE	Equal variances assumed	33.635	.000	-2.582	58	.012	-.18500	.07164	-.32841	-.04159
	Equal variances not assumed			-2.582	36.118	.014	-.18500	.07164	-.33029	-.03971

T-Test

Group Statistics

	Group	N	Mean	Std. Deviation	Std. Error Mean
Task_2_RAE	3	30	.7500	.07983	.01457
	1	30	.5447	.21646	.03952

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Task_2_RAE	Equal variances assumed	24.092	.000	4.875	58	.000	.20533	.04212	.12102	.28965
	Equal variances not assumed			4.875	36.745	.000	.20533	.04212	.11997	.29070

T-Test

Group Statistics

	Group	N	Mean	Std. Deviation	Std. Error Mean
Task_2_RAE	3	30	.7500	.07983	.01457
	2	30	.6140	.11545	.02108

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Task_2_RAE	Equal variances assumed	5.594	.021	5.307	58	.000	.13600	.02563	.08470	.18730
	Equal variances not assumed			5.307	51.571	.000	.13600	.02563	.08457	.18743

T-Test

Group Statistics

	Group	N	Mean	Std. Deviation	Std. Error Mean
Task_3_RAE	3	30	.2450	.18236	.03329
	1	30	.5273	.40065	.07315

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Task_3_RAE	Equal variances assumed	9.737	.003	-3.513	58	.001	-.28233	.08037	-.44321	-.12146
	Equal variances not assumed			-3.513	40.521	.001	-.28233	.08037	-.44470	-.11997

T-Test

Group Statistics

	Group	N	Mean	Std. Deviation	Std. Error Mean
Task_3_RAE	3	30	.2450	.18236	.03329
	2	30	.2850	.25936	.04735

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Task_3_RAE	Equal variances assumed	.655	.422	-.691	58	.492	-.04000	.05789	-.15587	.07587
	Equal variances not assumed			-.691	52.041	.493	-.04000	.05789	-.15615	.07615

APPENDIX D

TEAM COMMUNICATION PROCESS RATINGS

Date:_____ Team: _____ Session _____
Condition _____

Team Communication Process Ratings

Complete after each Session

Overall, how well did the members communicate with each other (group/team)?

Communication: Verbal and non-verbal passing of relevant and necessary information and the recipients' acknowledgement of understanding information. Example: Discussing meaning of graphs, the potential opening weekend gross earnings amount, etc.

Good Communication: staying on task, acknowledging information when received, not interrupting others as they speak…

Poor Communication: off-task conversation, interrupting others, not acknowledging received information/input…

Poor 1	2	3	Average 4	5	6	Excellent 7

Did the group/team members take turns speaking?

Not talking over each other, taking turns speaking, waiting for a response and then answering.

Never 1	2	3	Sometimes 4	5	6	Always 7

Were there off topic conversations?

Anything that is unrelated to the experimental tasks.

Never 1	2	3	Sometimes 4	5	6	Always 7

Is there are group/team leader emerging?

Examples of leadership: taking charge of the discussion (can be pos. or neg.), making decisions for the group/team, directing others, coordinating analysis, dividing up tasks, guiding the group during the analysis, deciding how information should be interpreted, etc.

Yes – If yes, then mark who? Station 1, 2, or 3 **No**

Did the group/team verify and agree upon the "Gross Opening Weekend Earnings Amount" before submitting?

Acknowledgement of ALL group/team members and acceptance by ALL team members.

Yes **No**

How did the group/team make decisions? Describe briefly (weighing pros and cons, majority vote…).

APPENDIX E

TEAM PROCESS RATING

Date:_____ Team: _____ Session _____
Condition _____

Team Process Rating – complete <u>during</u> each session.

Information Seeking Team/group members ask for information (active seeking) pertaining to the interface, information about movies, etc.	
Sharing of Information Team/group members share relevant information (anything related to the task – movie earnings, graphs, reviewing who the actors are, movie release times, etc.)	
Clarify/Explain Information Team/group members explain given information to each other, clarify misconceptions (e.g. I think this graph means …).	
Acknowledgement Team/group members acknowledge each other's input, suggestions, questions, etc.	
Giving Updates Team/group members update each other on what they are currently doing (e.g. what graph/information they are currently working on/looking at, etc.)	
Agreement/Consensus Team members agree upon meaning of graphs, information presented, prediction amount, etc.	
Present Suggestions Team/group members present possible solutions, suggestions on how to analyze data, ways of looking at the graphs, etc.	
Feedback Team/group members provide input on analysis to each other (e.g. what is going well/poorly).	
Planning What to do next, what strategies to use in subsequent analysis, etc.	
Help out Could be in form of assisting with navigation through the interface (computer assistance) or understanding data/information.	
Praise	
Repeated Requests (for same information) Having to ask repeatedly for a response, explanation, input, etc.	
Negative Comments	
Argument	
Did the team/group organize the given information in any way? (besides what was dictated by the experimental condition) If applicable,	

describe briefly (Did they divide their tasks up? What information/data each of them will be looking at?)	

Additional Comments:

APPENIX F

INTERVIEW QUESTIONS

Interview Questions

Decision Making

Individual:
What data did you use to make your decisions and predictions and why?

Group:
What data did your group use to make decisions and predictions and why?

Team:
What data did your team use to make decisions and predictions and why?

APPENDIX G

NASA TLX

TLX Report

Instructions:

Below you will be asked some questions about the task you just completed. Please read each question and think about the information being requested. Then, respond on each scale about how you felt or what you experienced within the task. Please consider each scale independent of the previous or following scales. If you have any questions, please ask the experimenter.

2.How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)?

The task was easy	**1 2 3 4 5 6 7 8 9 10**	The task was demanding
The task was simple	**1 2 3 4 5 6 7 8 9 10**	The task was complex
The task was forgiving	**1 2 3 4 5 6 7 8 9 10**	The task was exacting
The task was mentally effortless	**1 2 3 4 5 6 7 8 9 10**	The task was mentally difficult

2. How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred?

The task was slow	**1 2 3 4 5 6 7 8 9 10**	The task was rapid
The task was leisurely	**1 2 3 4 5 6 7 8 9 10**	The task was frantic

3. How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)?

Unsuccessful **1 2 3 4 5 6 7 8 9 10** Successful

4. Please rate the following emotional dimensions felt during the task

Insecure	**1 2 3 4 5 6 7 8 9 10**	Secure
Discouraged	**1 2 3 4 5 6 7 8 9 10**	Gratified
Irritated	**1 2 3 4 5 6 7 8 9 10**	Content
Stressed	**1 2 3 4 5 6 7 8 9 10**	Relaxed
Annoyed	**1 2 3 4 5 6 7 8 9 10**	Complacent

APPENDIX H

DEMOGRAPHIC QUESTIONNAIRE

Demographics Questionnaire

Date: _____ Team #: _____ Role/Station: _____

Please answer the following to the best of your ability. All answers will be kept confidential and will only be reported statistically (grouped with others' responses). Please feel free to leave a question blank if you feel uncomfortable answering it.

1. What is your age? _____

2. What is your gender? (circle):
   a. Male
   b. Female

3. What is your current level of education?
   a. Less than High School
   b. High School/GED
   c. Some College
   d. 2 year degree
   e. 4 year degree
   f. Master's
   g. Doctoral
   h. Professional (MD, JD, etc.)

4. If you have been or are enrolled in a post high school institution, what is your major?
   _____

5. Are you currently employed?
   a. Yes
   b. No

6. If yes to #5, what is your job title?
_____

7. Are you a native English speaker?
   a. Yes
   b. No
   c. If No, then what is your native language?
   _____

8. How long have you lived in the United States?
   a. Native (all my life)
   b. Less than 1 year
   c. 1 year
   d. 2 years
   e. 3 years
   f. 4 years
   g. Greater than 5 years

9. Within a month: On average, how often do you watch movies (Theater, TV, Internet)?
   a. 0
   b. 1-2
   c. 3-4
   d. 5-6
   e. 7-8
   f. 9-10
   g. 11 or more

10. Were you familiar with any of the movies presented today?
    a. Yes
    b. No
    c. If, yes which one(s)?
    _____
    _____
    -
    _____
    _____
    _____
    _____
    _____
    _____

11. If you answered Yes to #10: Did you know how much money the movie(s) had made during the opening weekend?

    a. Yes
    b. No
    c. If, yes which one(s)?
    _____
    _____
    _____

_____
_____

12. Were you familiar with how popular any of these movies were opening weekend?
    a. Yes
    b. No
    c. If, yes which one(s)?
    _____
    _____
    -
    _____
    _____
    _____
    _____

    _____
    _____

13. Do you follow new release movies on Social Media?

    a. Yes
    b. No

14. How often do you use Social Media?

    a. Several times an hour
    b. Hourly
    c. Daily
    d. Every couple days
    e. Once a week
    f. Every couple weeks
    g. Less than once a month
    h. Every couple of months
    i. Once or twice a year
    j. Never

15. Were you familiar with predictive modelling prior to this experiment?

    a. Yes
    b. No
    c. If yes, how? Please describe/list briefly (For example: learned in class, part of my degree program, part of my job, etc.)

    _____
    _____
    _____
    _____
    _____

16. How often have you used predictive modeling prior to this experiment?

    a. Never
    b. Rarely
    c. Occasionally
    d. Frequently
    e. Very Frequently

17. Do you have experience planning or coordinating events?
    a. Yes
    b. No

18. If yes to #17, please elaborate:
_____
_____
_____
_____
_____
_____

19. How often do you use a computer?

    a. Hourly
    b. Daily
    c. Every couple days
    d. Once a week
    e. Every couple weeks
    f. Less than once a month
    g. I do not use computers

20. Please rate the degree to which you agree with the following statement: I am proficient with computers.
    a. Strongly Agree
    b. Slightly Agree
    c. Neutral
    d. Slightly Disagree
    e. Strongly Disagree

21. In what way do you use computers? (Circle **all** that apply)
    a. I do not use computers
    b. Internet
    c. Email
    d. Word processing
    e. Spreadsheets
    f. Computer Games
    g. Other

22. Do you work with a **team** (a team is a group with members who have different specialties and who work interdependently) on a **regular basis** (2-3 times a week)?
    a. Yes
    b. No

23. If yes to #22, in what context do you work with a team and how many individuals make up this team? (Circle all that apply)
    a. Work-related *If circled, provide number of individuals* _____
    b. Sports *If circled, provide number of individuals* _____
    c. Other Recreation *If circled, provide number of individuals* _____
    d. Other *If circled, provide number of individuals* _____
       Please specify other: _____

Please rate the degree to which you agree with the following statements. Consider your *team* to be made up of the other people in this study.

24. I feel like my individual contribution to the team was important.
    a. Strongly Agree
    b. Slightly Agree
    c. Neutral
    d. Slightly Disagree
    e. Strongly Disagree

25. Regardless of outcome, I feel like we performed well overall.
    a. Strongly Agree
    b. Slightly Agree
    c. Neutral
    d. Slightly Disagree
    e. Strongly Disagree

26. The other people on my team were good members.
    a. Strongly Agree
    b. Slightly Agree
    c. Neutral
    d. Slightly Disagree
    e. Strongly Disagree

27. If I were asked to participate in another project like this one, I would like to be with the same team members.
    a. Strongly Agree
    b. Slightly Agree
    c. Neutral
    d. Slightly Disagree
    e. Strongly Disagree

28. This task was complicated.
    a. Strongly Agree
    b. Slightly Agree
    c. Neutral
    d. Slightly Disagree
    e. Strongly Disagree

29. The strategy we employed was the most effective way to complete the task.
    a. Strongly Agree
    b. Slightly Agree
    c. Neutral
    d. Slightly Disagree
    e. Strongly Disagree

30. This task was boring.
    a. Strongly Agree

b. Slightly Agree
c. Neutral
d. Slightly Disagree
e. Strongly Disagree

31. The way we made decisions was the best way to make decisions.
    a. Strongly Agree
    b. Slightly Agree
    c. Neutral
    d. Slightly Disagree
    e. Strongly Disagree

32. Our group could have done better if we had worked together more as a team.
    a. Strongly Agree
    b. Slightly Agree
    c. Neutral
    d. Slightly Disagree
    e. Strongly Disagree

33. I did **not** like the way our team made decisions.
    a. Strongly Agree
    b. Slightly Agree
    c. Neutral
    d. Slightly Disagree
    e. Strongly Disagree

34. I was motivated to help the group complete our tasks.
    a. Strongly Agree
    b. Slightly Agree
    c. Neutral
    d. Slightly Disagree
    e. Strongly Disagree

35. This task was easy.
    a. Strongly Agree
    b. Slightly Agree
    c. Neutral
    d. Slightly Disagree
    e. Strongly Disagree

36. I liked interacting with other members of the group.
    a. Strongly Agree
    b. Slightly Agree
    c. Neutral
    d. Slightly Disagree
    e. Strongly Disagree

37. The user-computer interface was easy to use.
    a. Strongly Agree
    b. Slightly Agree
    c. Neutral
    d. Slightly Disagree
    e. Strongly Disagree

38. What features of the movie interface were easy to use/understand? Please list below.
    _____
    _____
    _____
    _____
    _____
    _____
    _____
    _____
    ____

39. What features of the movie interface were difficult to work with/understand?
    Please list below.
    _____
    _____
    _____
    _____
    _____
    _____
    _____
    _____
    _____
    _____
    _____

40. At any time during this experiment did you guess the "Opening Gross Movie Earnings" amount?
    a. Yes
    b. No

41. If yes, how frequently did you guess?
    a. Practice Prediction – *Jack Rayan: Shadow Recruit*
    b. First Prediction – *The Hobbit*
    c. Second Prediction – *About Last Night*
    d. Third Prediction – *RoboCop*

42. I would use this interface again.
    a. Strongly Agree
    b. Slightly Agree
    c. Neutral
    d. Slightly Disagree
    e. Strongly Disagree

43. I enjoyed participating in this study.
    a. Strongly Agree
    b. Slightly Agree
    c. Neutral
    d. Slightly Disagree
    e. Strongly Disagree

44. Would you like to share anything else about this experiment with us? Please list/describe below._____
    _____
    _____
    _____
    _____

Thank You!

APPENDIX I

MOVIE INTERFACE QUIZ

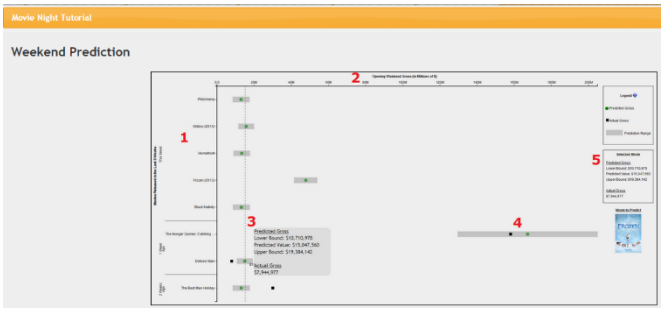
Movie Interface Quiz

1) I have access to the tutorial page throughout the entire study.
    a. True
    b. False
2) The goal of this task is to predict as accurately as possible how much the movies are going to make opening weekend.
    a. True
    b. False
3) I can get the actual gross amount for all movies displayed.
    a. True
    b. False
4) Each of the small green squares represent a week in a calendar year.
    a. True
    b. False
5) Why are some of the last squares in 'Weekend to Predict' colored black?
    a. I can't remember.
    b. They are colored black because that is the weekend to be predicted.
    c. They are colored black because some of those values are not available.
6) The 'Sentiment Analysis' displays tweets about the given movie selected.
    a. True
    b. False
7) I can access the following data. Mark all that apply.
    a. MPAA Rating
    b. Genre/Category of movie
    c. AMGA
    d. Number of tweets per day
    e. Release date
    f. Other
8) My prediction has to be within the given prediction range.
    a. True
    b. False
9) The 'Total Opening Weekend Gross' is:
    a. Predicted amount for the movie under investigation.
    b. Predicted amount for all movies to be released that weekend.
    c. Not sure.
10) My prediction amount can be given with commas, periods, asterisks, etc.
    a. True
    b. False

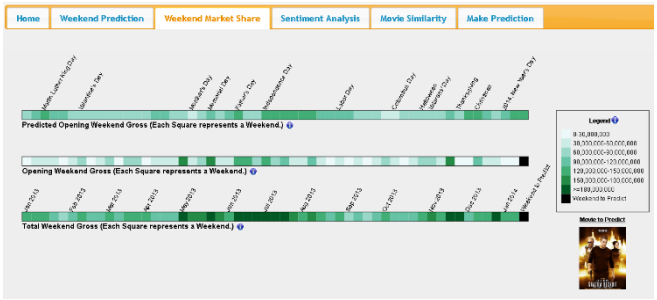
APPENDIX J

MOVIE INTERFACE SCREEN SHOTS

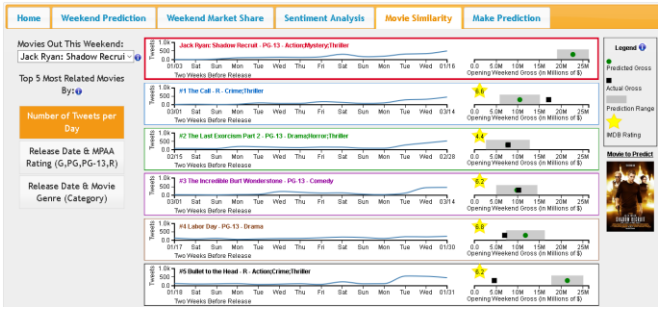
Screen Shot of “Weekend Prediction” page



Screen Shot of “Weekend Market Share” page



Screen Shot of “Movie Similarity” page



APPENDIX K

DEBRIEFING

Debriefing for "Predicting Movie Earnings" Study

The experiment that you have just participated in is part of an Arizona State University project to evaluate various information displays and their effectiveness. Additionally, we are evaluating the effects of team communication has on team performance and team situation awareness for intelligence related tasks (e.g. shifting through vast amounts of data, synthesizing data, making predictions).

Our research will look how the condition you were in affect team communication. At the same time, we will look at how levels of team communication relate to team performance on the task.

Our research will validate measures of team communication, and also improve our understanding of team cognition and team performance in the context of the intelligence analysis. The results of this study should provide methods useful for the study of team performance in other settings as well.

We ask that you do not discuss the details of this experiment with anyone who might participate in the future. If you are interested in being invited back to participate in the future experiments like this, please inform the experimenter at this time.

If you have any questions, comments, or concerns regarding this experiment please do not hesitate to contact the experimenter present today or contact Dr. Nancy Cooke (480) 988-2173.

Thank you for your participation and for your patience.

APPENDIX L

KAPPA SUMMARY

Cohen's Kappa Summary

Cohen's Kappa	Researcher 1			
Researcher 2	Over	Same	Under	Total
Over			26	26
Same		149		149
Under	20			20
	20	149	26	195
Agreement	20	149	26	195
By Chance	2.67			
Kappa	0.76			

Cohen's Kappa by Observation

Question	**Kappa**	**Agreement**
Information Seeking	0.71	Moderate
Sharing of Information	0.47	Fair
Clarify/Explain Information	0.39	Fair
Acknowledgement	0.23	Fair
Giving Updates	0.78	Moderate
Agreement Consensus	0.47	Fair
Present Suggestions	0.78	Moderate
Feedback	1.00	Very Good
Planning	0.93	Good
Help out	0.78	Moderate
Praise	1.00	Very Good
Repeated Requests	0.93	Good
Negative Comments	1.00	Very Good
Argument	1.00	Very Good

Cohen's Kappa Details

Information Seeking

Researcher 2	Researcher 1 Over	Same	Under	Total
Over			2	2
Same		10		10
Under	2			2
	2	10	2	14
Agreement	0	10	0	10
By Chance	0.29			
Kappa	0.71			

Giving Updates

Researcher 2	Researcher 1 Over	Same	Under	Total
Over			1	2
Same		11		10
Under	2			2
	2	11	1	14
Agreement	0	10	0	10
By Chance	0.14			
Kappa	0.78			

Sharing of Information

Researcher 2	Researcher 1 Over	Same	Under	Total
Over			5	2
Same		7		10
Under	2			2
	2	7	5	14
Agreement	0	10	0	10
By Chance	0.71			
Kappa	0.47			

Agreement Consensus

Researcher 2	Researcher 1 Over	Same	Under	Total
Over			4	2
Same		7		10
Under	3			2
	3	7	4	14
Agreement	0	10	0	10
By Chance	0.86			
Kappa	0.47			

Clarify/Explain Information

Researcher 2	Researcher 1 Over	Same	Under	Total
Over			2	2
Same		6		10
Under	6			2
	6	6	2	14
Agreement	0	10	0	10
By Chance	0.86			
Kappa	0.39			

Present Suggestions

Researcher 2	Researcher 1 Over	Same	Under	Total
Over			2	2
Same		11		10
Under	1			2
	1	11	2	14
Agreement	0	10	0	10
By Chance	0.14			
Kappa	0.78			

Acknowledgement

Researcher 2	Researcher 1 Over	Same	Under	Total
Over			7	2
Same		4		10
Under	2			2
	2	4	7	14
Agreement	0	10	0	10
By Chance	1.00			
Kappa	0.23			

Help out

Researcher 2	Researcher 1 Over	Same	Under	Total
Over			1	2
Same		11		10
Under	2			2
	2	11	1	14
Agreement	0	10	0	10
By Chance	0.14			
Kappa	0.78			

Feedback

Researcher 2	Researcher 1 Over	Same	Under	Total
Over				2
Same		14		10
Under				2
	0	14	0	14

Agreement	0	10	0	10
By Chance	0.00			

Kappa 1.00

Praise

Researcher 2	Researcher 1 Over	Same	Under	Total
Over				2
Same		14		10
Under				2
	0	14	0	14

Agreement	0	10	0	10
By Chance	0.00			

Kappa 1.00

Planning

Researcher 2	Researcher 1 Over	Same	Under	Total
Over			1	2
Same		13		10
Under				2
	0	13	1	14

Agreement	0	10	0	10
By Chance	0.00			

Kappa 0.93

Repeated Requests

Researcher 2	Researcher 1 Over	Same	Under	Total
Over			1	2
Same		13		10
Under				2
	0	13	1	14

Agreement	0	10	0	10
By Chance	0.00			

Kappa 0.93

Argument

Researcher 2	Researcher 1 Over	Same	Under	Total
Over				2
Same		14		10
Under				2
	0	14	0	14

Agreement	0	10	0	10
By Chance	0.00			

Kappa 1.00

Negative Comments

Researcher 2	Researcher 1 Over	Same	Under	Total
Over				2
Same		14		10
Under				2
	0	14	0	14

Agreement	0	10	0	10
By Chance	0.00			

Kappa 1.00

APPENDIX M

TABLE OF PERFORMANCE BY TEAM VERSES ACTUAL OWG

		Difference to Actual OWG			RAE Actual OWG		
Team	Subject Factor	Mean	STDEV	STDERR	Mean	STDEV	STDERR
1	Individual	-12.0	5.4	3.1	0.4	0.2	0.1
2	Group	1.7	14.6	8.4	0.3	0.2	0.1
3	Team	-4.7	17.1	9.9	0.5	0.4	0.2
4	Individual	1.0	14.4	8.3	0.3	0.2	0.1
5	Group	20.7	49.3	28.5	0.6	0.4	0.3
6	Team	1.0	14.9	8.6	0.3	0.2	0.1
7	Individual	15.8	38.7	22.4	0.4	0.4	0.2
9	Team	-5.8	16.3	9.4	0.4	0.3	0.2
10	Individual	21.9	38.2	22.0	0.8	0.2	0.1
11	Group	-5.0	14.0	8.1	0.4	0.4	0.2
12	Team	-5.8	11.8	6.8	0.3	0.4	0.2
13	Individual	2.5	17.4	10.1	0.5	0.3	0.2
14	Group	5.8	17.9	10.3	0.5	0.4	0.2
15	Team	-12.2	13.6	7.8	0.4	0.3	0.2
16	Individual	1.1	14.5	8.4	0.4	0.2	0.1
17	Group	-2.3	13.5	7.8	0.3	0.3	0.2
18	Team	1.7	23.2	13.4	0.4	0.3	0.2
19	Individual	-14.6	22.9	13.2	0.5	0.1	0.1
20	Group	11.8	30.5	17.6	0.4	0.3	0.2
21	Team	0.9	27.8	16.0	0.6	0.2	0.1
22	Individual	9.7	19.9	11.5	0.6	0.4	0.2
23	Group	1.7	15.6	9.0	0.3	0.2	0.1
24	Team	-7.7	8.0	4.6	0.3	0.3	0.2
25	Individual	-16.8	21.3	12.3	0.4	0.3	0.2
26	Group	0.7	10.7	6.2	0.3	0.2	0.1
27	Team	-2.7	19.3	11.1	0.4	0.4	0.2
28	Individual	11.1	25.9	15.0	0.5	0.0	0.0
29	Group	0.0	17.0	9.8	0.3	0.3	0.2
30	Team	-0.1	15.0	8.6	0.4	0.3	0.2
31	Group	21.8	49.5	28.6	0.8	0.3	0.2

APPENDIX N

TABLE OF PERFORMANCE BY TEAM VERSES MODEL

		Difference to Model			RAE to Model		
Team	Subject Factor	Mean	STDEV	STDERR	Mean	STDEV	STDERR
1	Individual	-13.6	9.4	5.4	0.3	0.2	0.1
2	Group	0.0	8.8	5.1	0.2	0.1	0.0
3	Team	-6.3	5.0	2.9	0.3	0.3	0.2
4	Individual	-0.6	8.4	4.8	0.2	0.1	0.0
5	Group	19.0	46.4	26.8	0.5	0.3	0.2
6	Team	-0.6	10.2	5.9	0.2	0.1	0.0
7	Individual	14.1	37.0	21.4	0.4	0.3	0.2
9	Team	-7.5	14.7	8.5	0.5	0.3	0.2
10	Individual	20.3	32.4	18.7	0.3	0.3	0.2
11	Group	-6.6	0.9	0.5	0.3	0.2	0.1
12	Team	-7.5	6.4	3.7	0.3	0.3	0.1
13	Individual	0.8	7.6	4.4	0.2	0.2	0.1
14	Group	4.2	4.5	2.6	0.1	0.0	0.0
15	Team	-13.8	9.2	5.3	0.4	0.1	0.1
16	Individual	-0.5	6.9	4.0	0.2	0.1	0.0
17	Group	-4.0	12.2	7.1	0.3	0.2	0.1
18	Team	0.0	20.0	11.5	0.4	0.1	0.1
19	Individual	-16.2	21.3	12.3	0.3	0.2	0.1
20	Group	10.1	29.1	16.8	0.4	0.2	0.1
21	Team	-0.8	26.6	15.3	0.5	0.1	0.1
22	Individual	8.1	7.6	4.4	0.2	0.1	0.1
23	Group	0.0	11.7	6.8	0.2	0.1	0.1
24	Team	-9.3	7.1	4.1	0.3	0.2	0.1
25	Individual	-18.4	30.2	17.4	0.8	0.3	0.2
26	Group	-1.0	5.5	3.2	0.1	0.1	0.0
27	Team	-4.3	15.0	8.6	0.4	0.3	0.2
28	Individual	9.5	20.5	11.9	0.2	0.2	0.1
29	Group	-1.6	12.9	7.4	0.3	0.1	0.1
30	Team	-1.8	5.7	3.3	0.2	0.1	0.1
31	Group	20.2	45.3	26.1	0.5	0.4	0.2

APPENDIX O

TABLE OF RELATIVE ABSOLUTE ERROR BY TEAM AND MOVIE

Hobbit

Team #	Min RAE	Mean RAE	Subject Factor	1	2	3
12	3%	3.0%	Team	Male	Male	Male
24	4%	4.0%	Team	Male	Male	Male
11	5%	5.0%	Group	Male	Male	Male
3	9%	9.0%	Team	Female	Male	Female
26	9%	9.0%	Group	Male	Female	Female
4	2%	9.3%	Individual	Male	Female	Male
30	12%	12.0%	Team	Female	Female	Male
9	15%	15.0%	Team	Female	Male	Male
14	15%	15.0%	Group	Male	Male	Male
17	15%	15.0%	Group	Male	Female	Male
2	18%	18.0%	Group	Male	Female	Female
22	6%	18.7%	Individual	Male	Female	Male
6	19%	19.0%	Team	Female	Male	Male
16	15%	19.3%	Individual	Male	Female	Male
1	2%	21.7%	Individual	Male	Female	Male
23	22%	22.0%	Group	Male	Female	Male
27	22%	22.0%	Team	Male	Male	Male
29	22%	22.0%	Group	Female	Female	Male
28	5%	22.3%	Individual	Male	Male	Male
15	28%	28.0%	Team	Male	Male	Male
7	5%	34.0%	Individual	Male	Male	Male
18	36%	36.0%	Team	Male	Male	Male
13	18%	39.0%	Individual	Male	Male	Female
21	44%	44.0%	Team	Male	Male	Male
25	23%	47.3%	Individual	Male	Male	Male
10	9%	59.0%	Individual	Female	Female	Female
20	63%	63.0%	Group	Male	Male	Male
19	50%	66.0%	Individual	Male	Male	Female
5	104%	104.0%	Group	Male	Male	Male
31	104%	104.0%	Group	Female	Male	Male

About Last Night

Team #	Min RAE	Mean RAE	Subject Factor	1	2	3
13	17%	38.7%	Individual	Male	Male	Female
19	30%	44.0%	Individual	Male	Male	Female
20	44%	44.0%	Group	Male	Male	Male
26	46%	46.0%	Group	Male	Female	Female
10	30%	46.3%	Individual	Female	Female	Female
1	30%	52.0%	Individual	Male	Female	Male
22	39%	53.7%	Individual	Male	Female	Male
14	55%	55.0%	Group	Male	Male	Male
2	57%	57.0%	Group	Male	Female	Female
23	57%	57.0%	Group	Male	Female	Male
16	25%	57.7%	Individual	Male	Female	Male
25	7%	58.3%	Individual	Male	Male	Male
6	59%	59.0%	Team	Female	Male	Male
28	42%	60.0%	Individual	Male	Male	Male
17	61%	61.0%	Group	Male	Female	Male
24	66%	66.0%	Team	Male	Male	Male
4	59%	67.0%	Individual	Male	Female	Male
7	49%	67.0%	Individual	Male	Male	Male
30	68%	68.0%	Team	Female	Female	Male
5	69%	69.0%	Group	Male	Male	Male
29	69%	69.0%	Group	Female	Female	Male
12	75%	75.0%	Team	Male	Male	Male
11	77%	77.0%	Group	Male	Male	Male
15	77%	77.0%	Team	Male	Male	Male
18	77%	77.0%	Team	Male	Male	Male
21	79%	79.0%	Team	Male	Male	Male
31	79%	79.0%	Group	Female	Male	Male
3	81%	81.0%	Team	Female	Male	Female
9	82%	82.0%	Team	Female	Male	Male
27	86%	86.0%	Team	Male	Male	Male

Robocop

Team #	Min RAE	Mean RAE	Subject Factor	1	2	3
7	22%	0.0%	Individual	Male	Male	Male
12	1%	1.0%	Team	Male	Male	Male
20	1%	1.0%	Group	Male	Male	Male
22	59%	5.0%	Individual	Female	Male	Female
29	6%	6.0%	Group	Male	Female	Female
25	82%	7.0%	Individual	Male	Female	Male
18	8%	8.0%	Team	Female	Female	Male
27	10%	10.0%	Team	Female	Male	Male
17	12%	12.0%	Group	Male	Male	Male
24	14%	14.0%	Team	Male	Female	Male
5	15%	15.0%	Group	Male	Female	Female
23	15%	15.0%	Group	Male	Female	Male
15	16%	16.0%	Team	Female	Male	Male
1	23%	19.0%	Individual	Male	Female	Male
6	19%	19.0%	Team	Male	Female	Male
28	37%	27.0%	Individual	Male	Female	Male
4	31%	28.0%	Individual	Male	Male	Male
2	29%	29.0%	Group	Female	Female	Male
10	76%	31.0%	Individual	Male	Male	Male
26	32%	32.0%	Group	Male	Male	Male
9	35%	35.0%	Team	Male	Male	Male
16	50%	36.0%	Individual	Male	Male	Male
11	38%	38.0%	Group	Male	Male	Female
30	38%	38.0%	Team	Male	Male	Male
19	64%	41.0%	Individual	Male	Male	Male
21	43%	43.0%	Team	Female	Female	Female
31	43%	43.0%	Group	Male	Male	Male
13	83%	54.0%	Individual	Male	Male	Female
3	61%	61.0%	Team	Male	Male	Male
14	94%	94.0%	Group	Female	Male	Male

APPENDIX P

DETAILED PROCEDURE FOR THE STUDY

Detailed Procedure for Study:
"The Role of Teamwork in Predicting Movie Earnings"

The following procedure was used during all the experiments.

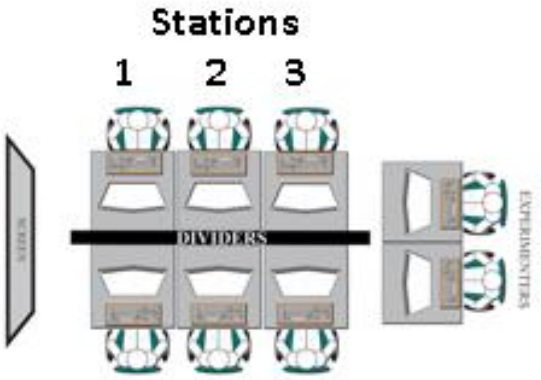
Preparation

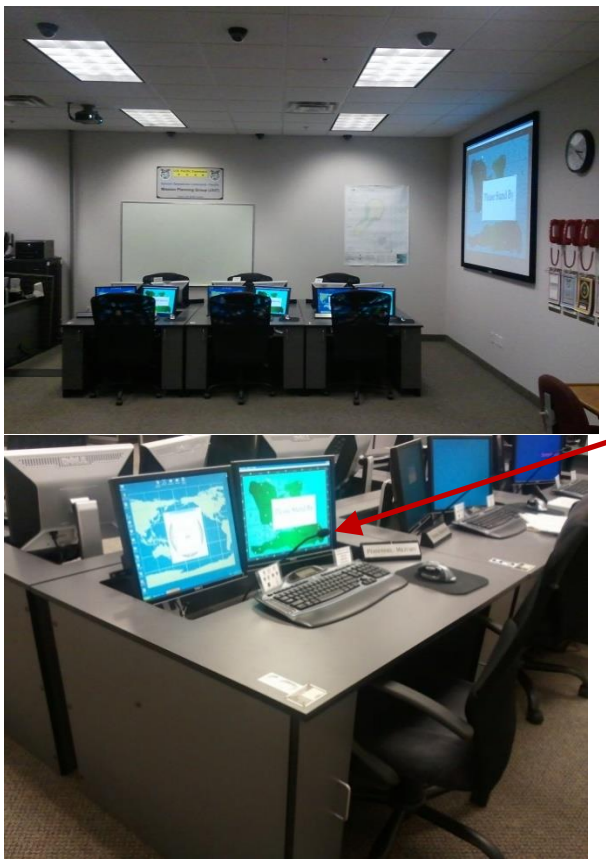
Before participants arrived the experimenters prepared the necessary paperwork, participants' workstations, and experimenters' computers.

1. It was assured that the arriving participants matched the participants listed on SignUpGenius for that day (and time slot).

2. Equipment set-up:
    a. The overhead projector was turned on using the remote.
    b. Standard desk top computers with duals screens, keyboard, and mouse were used in this experiment:
        i. The "Experimenter Right" computer was used to project a timer onto the big projection screen. This ensured that all participants were able to see the time remaining for each task. This computer was also used to project the quiz review power point presentation onto the projection screen for the "quiz review" portion of the experiment.
        ii. The "Experimenter Left" computer was used for audio recording the communication between group and team members. The program "Audacity" was used to capture all communication data. The audio files were saved after the experiment according to sessions (practice session, prediction 1, 2, and 3) and named according to team number and date.
    c. Workstations **1 through 3** (work station were permanently marked). For the team condition workstations were used in the following way:
        i. Station **1** was always "**Weekend Market Share**".
        ii. Station **2** was always "**Sentiment Analysis**".
        iii. Station **3** was always "**Movie Similarity**".
    d. Each workstation was provided with pen and paper. At the end of each experiment notes were collected (if participants had taken notes).

Experimenter room set-up





3. The training material was pre-loaded on each participants' computer according to condition (Individual/Group/Team) the following way:
    a. The training material consisted of a power point presentation illustrating the movie interface and data availability using the Disney movie *Frozen*, 2013.
    b. The folder "Movie Study" was accessed.

a. For **Individual and Group** condition the file "All Training_FINAL_.ppx" was selected.
b. For **Team Condition** files were selected according to station.
    i. For **Station 1** the following file was opened: "Team Training_Weekend Market Share_.ppx".
    ii. For **Station 2** the following file was opened: "Team Training_Sentiment Analysis_.ppx".
    iii. For **Station 3** the following file was opened: "Team Training_Movie Similarity_.ppx".

4. All workstation monitors were turned off once the correct training material was loaded.

5. Experimenters running the experiment ensured that the necessary paperwork was prepared before the participants arrived. Each packet contained the following paperwork:
    a. Informed Consent form (2 pages – staple so they don't become separated. TWO copies per participants, they are to be offered a copy to keep, 6 copies total)
    b. Movie Interface Quiz
    c. Team Process Rating sheets
    d. Team Communication Process Rating sheets
    e. Prediction Sheet
    f. Demographic Questionnaire
    g. TLX (2 copies for each participant, 6 copies total)
    h. Participant Certification Form (form for payment of participants)
    i. Debriefing Sheets
    j. Time Sheet

6. Additionally, an experimenter prepared the "Team Folder" – a manila folder that was marked with the following information: team #, date, and condition.
7. The team number and condition was assigned randomly following the "Team Log" sheet.
    a. Team ID's was assigned through the log:

Team #	Date	Condition	Experimenters	Notes
1		Individual		
2		Group		
3		Team		
4		Individual		

Experiment

When the participants arrived the experimenter checked whether the correct participant were present by using the sign-up sheet information from "SignUpGenius".

1. Next, the experimenter randomly assigned participants to workstation 1, 2, and 3 and directed them to their individual work station.
2. Once seated, the consent forms (2 each) were handed out and the following dialogue was read:

"Hello and welcome to our Movie Study. Thank you for coming in today. At your desk you have a consent form, which gives you information about the study, but most importantly, it is to let you know you are here of your own free will and if at any time you decide not to continue to participate, you can stop without any penalty. After you have read it, please sign it and we will pick it up. The second copy is for you to keep. If you do not wish to keep this copy, please leave it at the desk and we will collect it. We also would like to remind you that the audio of the experiment will be recorded today. The only people who will hear this information are the experimenters involved with this project.

3. The consent forms were collected by the experimenters once signed. Next, the flowing introduction dialogue was read to the participants:

"Before we begin, I just want to emphasize that as this is an experiment, we may say things formally, some things may seem redundant, or we may not always be able to help. We are just trying to keep things consistent for the experiment.

During the first part of the experiment you will review a Power Point Training Presentation. This presentation will familiarize you with the purpose of the study – which is – to predict how much money the presented movies made opening weekend, using data from Twitter and computational models.

Please complete the training in this presentation. Please, do not talk during this portion. If you have any questions, please raise your hand and we will assist you. You will have 30 minutes to complete the training. After the training you will be given a short quiz.

Also, we would like to point out that you will be able to refer to the presented information throughout the experiment, except – during the quiz portion, but other than that throughout. So, please focus on understanding the presented information and not on memorizing it.

Okay. Are there any questions before we move on? Please turn to your <u>left</u> monitor and begin the training."

4. The experimenters ensured that all participants accessed the training presentation. The timer was set for 30 minutes and projected onto the

projection screen. During the practice session the experimenters answered questions that the participants may have had about the interface/training content.

a. The quiz was handed out after the training session (after the 30 minutes). The timer was set for 5 minutes and projected onto the projection screen. The quiz was collect once the 5 minutes were over.
b. On Experimenter "Right Computer" the file titled "Quiz Review_.ppx" was accessed and projected onto the projection screen.
c. The experimenter reviewed the quiz with the participants by reading each question out load and then giving the correct answers to the questions. Additionally, the experimenter answered any questions the participants may have had about the quiz content or the interface.
d. Once the quiz was reviewed the participants received a 10 minute break. The following dialogue was read to the participants:

"Please take a 10 minute break. Please do not talk about the experiment during the break. Thank you."

5. While the participants were on break the three computer work stations were set up.
    a. The movie interface was accessed via Firefox (book marked) on each computer station (1, 2, and 3).

6. The teams were registered using admin login. Next, the interface was pulled up accordingly to condition (Individual/Group/Team).

7. For <u>Individual and Group</u> condition – use username and password for "All Tasks".
    i. Username: (intentionally left blank).
    ii. Password: (intentionally left blank).

8. For Team condition:
    i. Username: (intentionally left blank).
    ii. Password: (intentionally left blank).
    iii. Station 1 was always "Weekend Market Share" specialist (password, intentionally left blank).
    iv. Station 2 was always "Sentiment Analysis" specialist (password, intentionally left blank).
    v. Station 3 was always "Movie Similarity" specialist (password, intentionally left blank).

9. The experimenters ensured that each station was correctly loaded per assigned role (for Team condition only).

10. Movie Interface:
    a. The data for the first movie, Jack Ryan: Shadow Recruit, automatically loaded once the correct username and password were typed in by the experimenter.
    b. The data for the second movie, The Hobbit: The Desolation of Smaug, 2013 automatically loaded once the participant submitted his/her prediction of the previous movie (Jack Ryan).
    c. The data for the third movie, About Last Night, 2014 automatically loaded once the prediction of the previous movie (The Hobbit: The Desolation of Smaug) was submitted by the participant.
    d. The data for the forth movie, Robocop, 2014 automatically loaded once the prediction of the previous movie (About Last Night) was submitted by the participant.
    e. Once the last prediction was submitted a message appeared on the computer screen that the participant had completed the last task.
    f. All movies were presented in the same order throughout all the experiments.

11. Next, the experimenters ensured that each participants' computers screen were loaded properly with the following information displayed:
    a. Participants' Left computer screen displayed the movie interface.
    b. Participants' Right computer screen displayed the "Movie Training" power point presentation (in case they need to refer back to the information).
    c. The "calculator" was also opened on participants 'computer and display on participants' Right computer screen.

12. Once all participants returned from their break and sat down at their assigned work station the following script was read:

"Please direct your attention to your Left Computer screen. Please "click" on the Movie "Tutorial" bottom. Please take a few moments to review the tutorial. You will be able to access this tutorial thought all sessions as well as your Training PP and a calculator."

13. The experimenters waited until all participants had located the "Movie Tutorial" and reviewed it. Then the following script was read out load:

"Okay, we are now ready to begin with the practice session. Do you have any questions before we continue?"

14. FOR INDIVIDUAL CONDITION – the following script was read out loud:

"You will have 15 minutes to review the data. At end of the practice session you will be given 5 minutes to finalize your findings before submitting your prediction. Please do

not talk to the other participants during the remainder of the experiment. Also, please <u>STOP</u> after you have submitted your "Practice Prediction". Again, please do not proceed past your 1st Prediction.
I will set the timer for 15 minutes, you may begin now.

15. <u>FOR GROUP CONDITION</u> – the following script was read out loud:

"You will have 15 minutes to review the data. At end of the practice session you will be given 5 minutes to finalize your findings with your group members before submitting your prediction. You will submit one estimate as a group, so please make sure you agree upon an amount. Also, please <u>STOP</u> after you have submitted your "Practice Prediction". Again, please do not proceed past your 1st Prediction. I will set the timer for 15 minutes, you may begin now."

16. <u>FOR TEAM CONDITION</u> – the following script was read out loud:

"You will have 15 minutes to review the data. At end of the practice session you will be given 5 minutes to finalize your findings with your team members before submitting your prediction. You will submit one estimate as a team, so please make sure you agree upon an amount. Also, please <u>STOP</u> after you have submitted your "Practice Prediction". Again, please do not proceed past your 1st Prediction. I will set the timer for 15 minutes, you may begin now."

17. For all conditions:
    a. The timer was set to 15 minutes and projected onto the projection screen. During the practice session the experimenters monitored the participants and answered questions the participants may have had about the interface.
    b. Once 15 minutes were over they were prompt participants to finalize their findings.

18. Please note that for all conditions participants were promoted to write down their estimated on a provided sheet of paper first before entering it online. This was done for two reasons. One, not to lose any data points in case participants selected "submit" before they had entered their estimate. Second, in case of a possible system failure (data could have been lost).

Prediction Sheet

Please Write Down Your Prediction Before Submitting it Online

Practice Prediction	$
Prediction 1	$
Prediction 2	$
Prediction 3	$

Prediction Sheet given to each participant to fill out.

19. FOR INDIVIDUAL CONDITION – the following script was read out loud:

"You will now have 5 minutes to finalize your findings. Please write down your estimate on the provided sheet before submitting it online. Also, please do not advance past your practice prediction.
If participants needed additional time to wrap their data analysis up, then they were provided with it. They were given a maximum of 5 additional minutes .

20. FOR GROUP CONDITION – the following script was read out loud:

"You will now have 5 minutes to finalize your findings with your group. Please write down your group estimate on the provided sheet before submitting it online. Also, please do not advance past your practice prediction."
If participants needed additional time to wrap their data analysis up, then they were provided with it. They were given a maximum of 5 additional minutes .

21. FOR TEAM CONDITION – the following script was read out loud:

"You will now have 5 minutes to finalize your findings with your team. Please write down your team estimate on the provided sheet before submitting it online. Also, please do not advance past your practice prediction."
If participants needed additional time to wrap their data analysis up, then they were provided with it. They were given a maximum of 5 additional minutes .

22. For all Conditions:
    a. The timer was set for 5 minutes and projected onto the projection screen. If participants needed the additional time, then the timer was re-set for another 5 minutes.

23. The participants received automatically feedback once their practice prediction was submitted. The feedback consisted of the following information 1) displaying their own predication estimate 2) displaying the

"Actual Weekend Gross" for Jack Ryan. The feedback was only given during the practice prediction .

24. The printed NASA TLX form was handed out to the participants and collected once filled out.
    a. The individual workstation number was written on the TLX forms (station 1, 2, 3) as it was being collected. Additionally, the experimenter collecting the TLX forms wrote "#1" on each TLX form indicating that this was the first NASA TLX.

25. Next, the researcher read the following script out load:

"Okay, now you are ready to make your actual predictions. You will not receive any further feedback on how close you are to the actual opening weekend earnings amount. The remainder of the sessions will consist of 3 additional sessions. During each session you will be given 15 minutes to review the movie data and another 5 minutes to finalize your findings. Are there any questions before we proceed?"

Running Session 1, 2, and 3:

1. Each session followed the same procedure.
2. During each session the communication process rating sheets for Group and Team conditions were filled out by the experimenters.
    a. The "Team Process Rating" sheet was filled out during participants' discussion.
    b. The "Team Communication Process Ratings" sheet was filled out after the participants' discussion.

FOR INDIVIDUAL CONDITION – the following script was read out loud:
"Again, you will have 15 minutes to review the data. At end of the session you will be given 5 minutes to finalize your findings before submitting your prediction. Please do not talk to the other participants.
*****Please write down your prediction on the provided paper before entering it online. Also, if you end up needing a few extra minutes, please let me know."***

FOR GROUP CONDITION – the following script was read out loud:
"Again, you will have 15 minutes to review the data. At end of the session you will be given 5 minutes to finalize your findings with your group members before submitting your prediction. Remember that you will submit one estimate as a group, so please make sure you agree upon an amount.
*****Please write down your prediction on the provided paper before entering it online. Also, if you end up needing a few extra minutes, please let me know."***

FOR TEAM CONDITION – the following script was read out loud:
"Again, you will have 15 minutes to review the data. At end of the session you will be given 5 minutes to finalize your findings with your team members before submitting

your prediction. You will submit one estimate as a team, so please make sure you agree upon an amount. **Please write down your prediction on the provided paper before entering it online.
Also, if you end up needing a few extra minutes, please let me know."

3. **For ALL Conditions**: The timer was set to 15 minutes and projected onto the projection screen.
4. Once 15 minutes were over the participants were prompt to finalize their findings.

FOR INDIVIDUAL CONDITION – the following script was read out loud:
"You will now have 5 minutes to finalize your findings. Please write down your estimate on the provided sheet before submitting it online."
If participants needed additional time to wrap their data analysis up, then they were provided with it. They were given a maximum of 5 additional minutes .

FOR GROUP CONDITION - Read the following script:
You will now have 5 minutes to finalize your findings with your group. Please write down your group estimate on the provided sheet before submitting it online. Remember you need to work with your group to come up with one prediction.
If participants needed additional time to wrap their data analysis up, then they were provided with it. They were given a maximum of 5 additional minutes .

FOR TEAM CONDITION - Read the following script:
You will now have 5 minutes to finalize your findings with your team. Please write down your team estimate on the provided sheet before submitting it online. Remember you need to work with your group to come up with one prediction.
If participants needed additional time to wrap their data analysis up, then they were provided with it. They were given a maximum of 5 additional minutes .

5. **ALL Conditions:**
    a. The timer was set for 5 minutes and projected onto the projection screen. If participants needed the additional time, then the timer was re-set for another 5 minutes.

6. The remainder of the sessions were run the same way, following the same time line, and the same script for each session.

7. **AFTER ALL 3 SESSIOS:**

    a. Once the participants submitted their last prediction the second NASA TLX form was given to each participant.

b. The NASA TLX form was collected. Again, the experimenter collecting the TLX forms wrote down the participants' station number (station 1, 2, or 3) and "#2", indicating that that was the second TLX form (for that particular participant) collected.
c. Next, the experimenters handed out the "Demographic Questionnaire".
    i. The experimenters also ensured that each participant received the correct questionnaire marked with their station number (station 1, 2, or 3). The questionnaires were collected once filled out.

8. The following was stated once the questionnaires were collected:

"You are almost done with the experiment. The only part that is remaining is the short interview.'

Participants Interview:

1. The participants were taken one-at-a-time into the adjacent room and ask the interview question.
2. Printed screen shots of the movie interface were laid out in front of the participants. This was done in order to make it easier for the participants to recall the features of the interface, and, the data available to them.
3. The interview question was asked according to condition.
    i. For Individual: ***"What data did you use to make your predictions and why?"***
    ii. For Group: ***"What data did your group use to make predictions and why?"***
    iii. For Team: ***"What data did your team use to make predictions and why?"***

4. The interviews were recorded using a small hand-held recorder.

Concluding the Experiment:

a. The participants were paid for their time.
b. The "Debriefing" form was handed to each participant.
c. Any questions that the participants may have had about the experiment were answered by the researchers.
d. Lastly, the participants were thanked for their time.

APPENDIX Q

EQUIPMENT

Equipment

Server
The server used in this experiment has a quad-core with 2.7GHz. It has 16GB memory. It runs a Linux operating system.

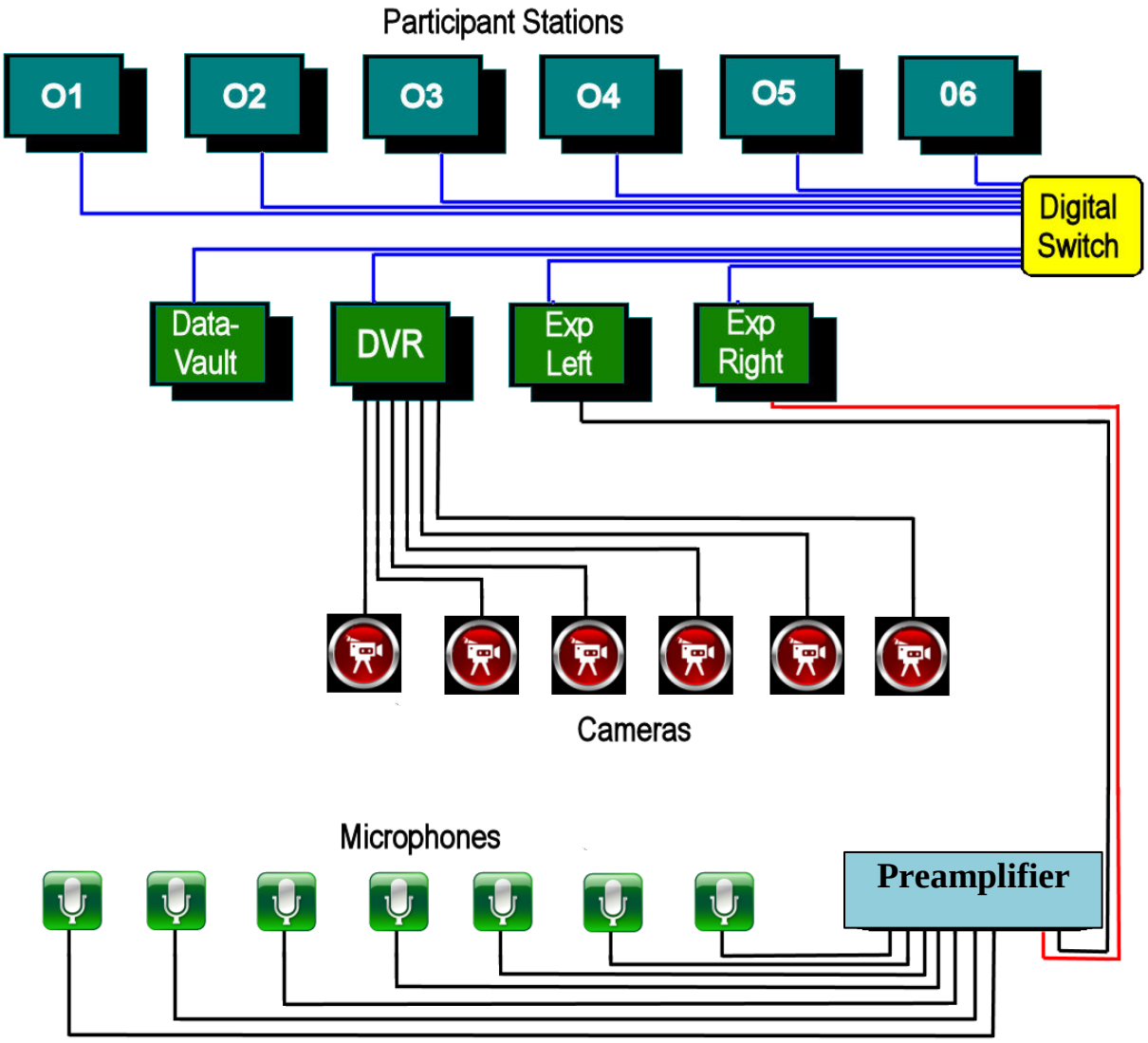
Movie Interface
The movie interface used in this experiment is a web-deployable program running on tomcat 7. The backend is developed in Java, and the frontend is developed in JavaScript. Ajax Call is used to communicate between the frontend and the backend. The data is stored in MySql Database and Mongodb. The model is developed using R, and the results are pre-calculated and stored in a csv file.

Projector
The test bed uses a single overhead projector and a 109 inch wall mounted screen. The projector specifications are: Dell 2400MP: ANSI lumens: 3000; Contrast (full on/off): 2100:1; Light Engine: 1024x768, native 4:3, single-chip DLP with a 2x speed 4-segment color wheel, 260W P-VIP lamp; Video Compatibility: HDTV 1080i, 720p, 576p, 576i, 480p. NTSC/PAL/SECAM. Data Compatibility: Computer resolutions up to UXGA (1600x1200). Connection Panel: one 15-pin VGA-in, one VGA-out, one USB port, one S-Video in, one composite port, one 1/8" audio in, one 1/8" audio out, a Kensington lock point, and an RS-232 port. Lens and Throw Distance: 1.20:1 manual zoom/focus lens. Throws a 100" diagonal 4:3 image from 13.5' to 16.2' Lamp Life: 2,000 hours (2,500 hours in eco mode).

Microphones
The microphones used in this experiment are: Shure Model. They are dynamic microphones. A PreSonus preamplifier provides phantom power to the microphones and also converts the signals to a line-level output. The Preamplifier digitizes the audio and sends it Audacity recoding software via an IEEE 1384 (Fire wire) connection to the Experimenter_Right computer.

Computers
Each console contains one computer, two monitors, a wireless mouse and a wireless keyboard. The specifications for these systems are:
Computer: Dell Model Optiplex 740
Monitors: Dell Model 1908FP
Keyboard/Mouse: Logitech MX5000 Bluetooth Desk set

Testbed Layout (please note that the video cameras were not used in this experiment)


Participant Stations
O1
O2
O3
O4
O5
06
Digital Switch
Data-Vault
DVR
Exp Left
Exp Right
Cameras
Microphones
Preamplifier

APPENDIX R

PARTICIPANT DEMOGRAPHICS

Participants

Type	Female	Male
Group	9	21
Individual	8	22
Team	6	24

Average Age

Type	Female	Male
Group	21.8	26.1
Individual	22.0	24.1
Team	24.8	23.7

Participants Employment

Job	Group	Individual	Team
Activity assistant	1	0	0
Beauty Advisor	1	0	0
Change agent	0	0	1
Coffee shop	1	0	0
Cook	0	0	1
Coordinator, Engr. Student Success	0	0	1
Customer Service Representative	0	1	0
Desk Assistant	0	1	0
Editor	1	0	0
EGR102 grader	0	1	0
Lube Technician	0	1	0
Naval Nucular Engineer	1	0	0
Not Given	1	1	1
Office Assistant	0	0	1
Research Assistant	2	3	5
Student worker	0	0	1
Teaching Assistant	2	0	0
Technology Analyist	1	0	0
Unemployed	18	21	19
Waitress	1	0	0
Youth Minister	0	1	0

Participants Education

Education Level	Group	Individual	Team	Overall
Some College	7	6	4	17
2 Year Degree	1	1	0	2
4 Year Degree	1	0	2	3
Profesional	1	0	0	1
Masters	19	22	23	64
Doctorial	1	1	1	3

Participants College Major

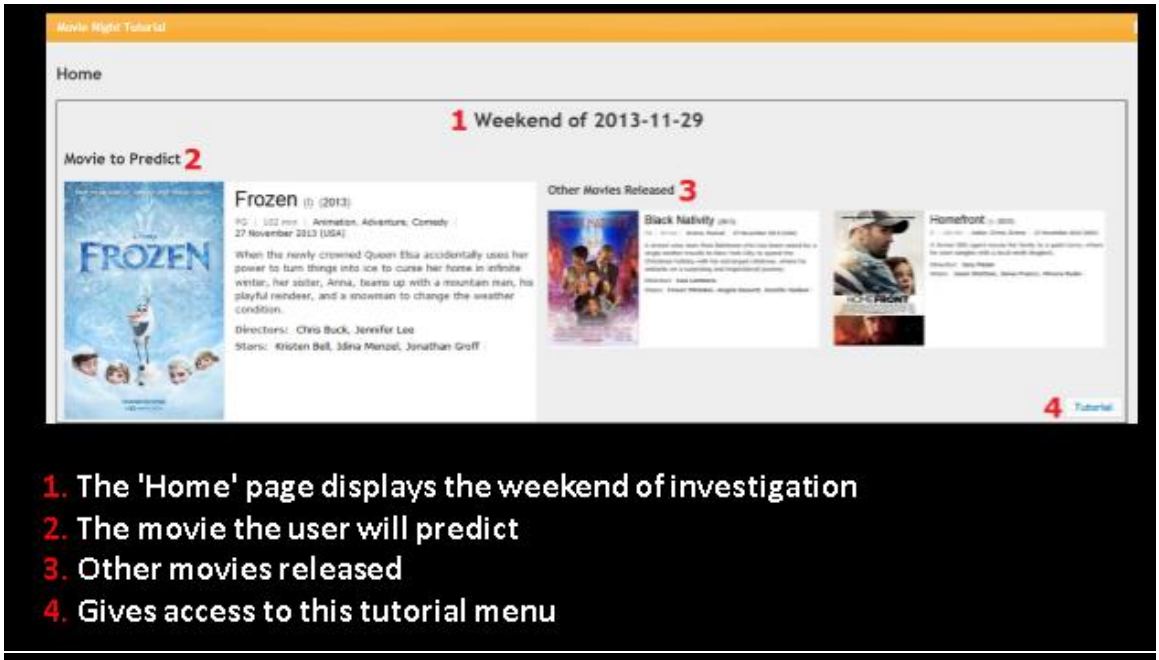
Major	Group	Individual	Team	Overall
Accounting	1	0	0	1
Air Transportation Management	0	0	1	1
Applied Biological Sciences	0	1	0	1
Areonautical Management Technology	0	1	0	1
Areospace Engineering	0	1	0	1
ASU	0	1	1	2
Bachelor of Technology, Electrical, and Industrial Engineeri	0	0	1	1
Chemical Engineering	3	0	1	4
Civil Engineering	0	0	1	1
Communication	0	0	1	1
Communications/ P.R.	1	0	0	1
Computer Engineering	2	0	1	3
Computer Information Systems	1	0	0	1
Computer Science	2	3	0	5
Economics and Marketing	1	0	0	1
Electrical Energy System Engineering	0	2	0	2
Electrical Engineering	2	3	4	9
Electrical Power	0	0	1	1
Electronics	1	0	0	1
Electronics / Energy	0	1	0	1
Elementary Education	1	0	0	1
Energy Systems	0	0	1	1
Engineering	0	1	1	2
Graphic Information Technology	0	1	0	1
Human Systems Engineering	1	0	2	3
Industrial Engineering	0	0	3	3
Information Technology	1	1	0	2
Materials Science	1	0	0	1
Mechanical Engineering	1	9	6	16
Not Given	1	2	1	4
Nuclear Sciences	1	0	0	1
Nursing	0	1	1	2
Power Engineer	1	0	0	1
Psychology	0	0	2	2
SMACS	2	0	0	2
Software Engineering	3	2	0	5

APPENDIX S

TRAINING MATERIAL

Welcome!

This Training tutorial will familiarize you with the visualization tools being used to analyze box office data. We want you to predict how much money a movie will make in its opening weekend using information from Twitter and computational models.
Our training tutorial movie is going to be
Disney's *Frozen*.

1. The 'Home' page displays the weekend of investigation
2. The movie the user will predict
3. Other movies released
4. Gives access to this tutorial menu

1. The release weekend for the movie to predict.

2. Movie to Predict: Description of the movie to predict.

3. Other Movies Released: Description(s) of other movies released on the same weekend.
Note: If there is no other movie released that weekend nothing will be displayed here.

4. Tutorial Button: Clicking the 'Tutorial' button will open this tutorial menu.

The 'Weekend Prediction' page displays movies released within the last three weeks and their "Opening Weekend Gross" details (Actual, Predicted, and Prediction Range).

1.Movies released within the last three weeks are displayed here.

a. By hovering over a movie title the data line corresponding to it will be highlighted.

b. If a movie's title is abbreviated, hovering over the title will also cause the full name to be displayed.

2. The x-axis scale displays the Opening Weekend Gross (OWG) for a movie in Millions of Dollars.

3. The 'Reference Line' gives an approximation of the dollar amount.
a. By hovering over a green circle or black square the exact values will be displayed in a tooltip.
b. Note: Actual gross is <u>not</u> shown for movies released the week of the movie to predict.

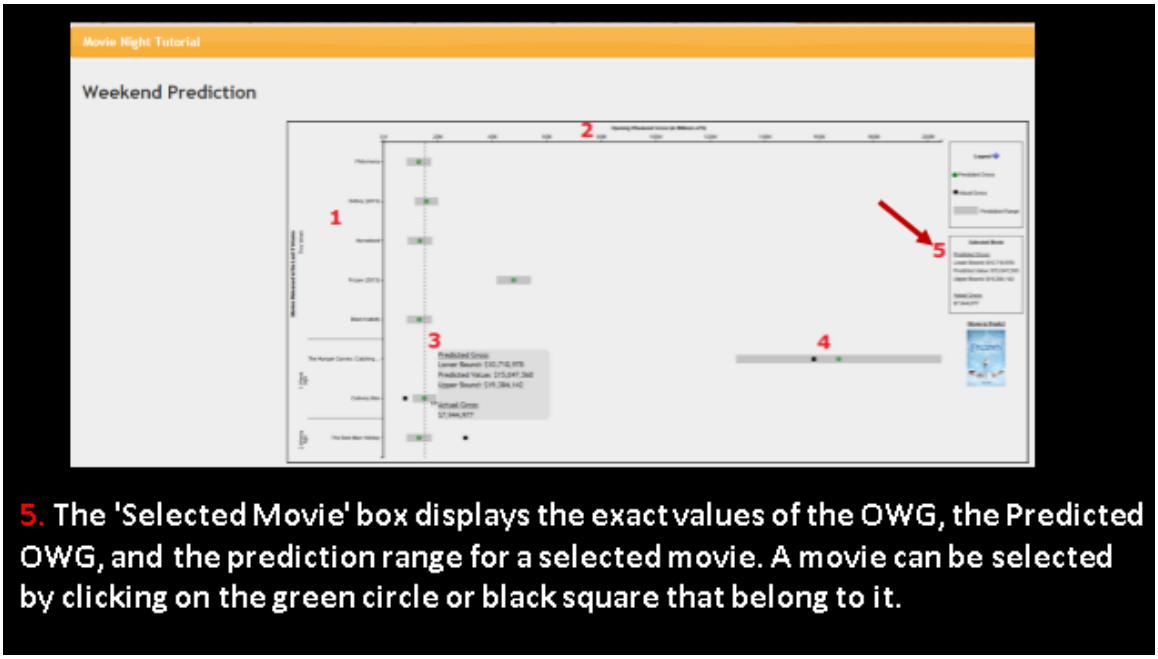
Actual – Actual amount (in $) the selected movie made opening weekend (access by clicking black square)

Predicted – The predicted amount of how much a movie will make opening weekend (access amount by clicking on green circle). This prediction is derived from a computational model.

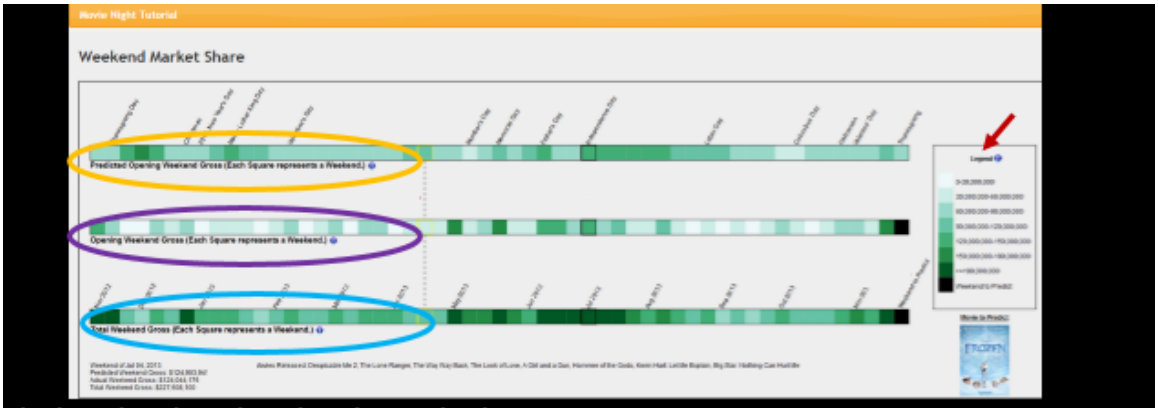
4. The green circles represent the OWG predicted by the model. The gray rectangles represent the prediction range of the prediction model. The black squares represent the actual gross reported from IMDB.

a. Click on a green circle or black square to see the exact values they represent under 'Selected Movie' (5).
b. Note: Actual gross is not shown for movies released the week of the movie to predict.

Prediction Range: from low to high (amount in $ Millions). This range is derived from the prediction model.



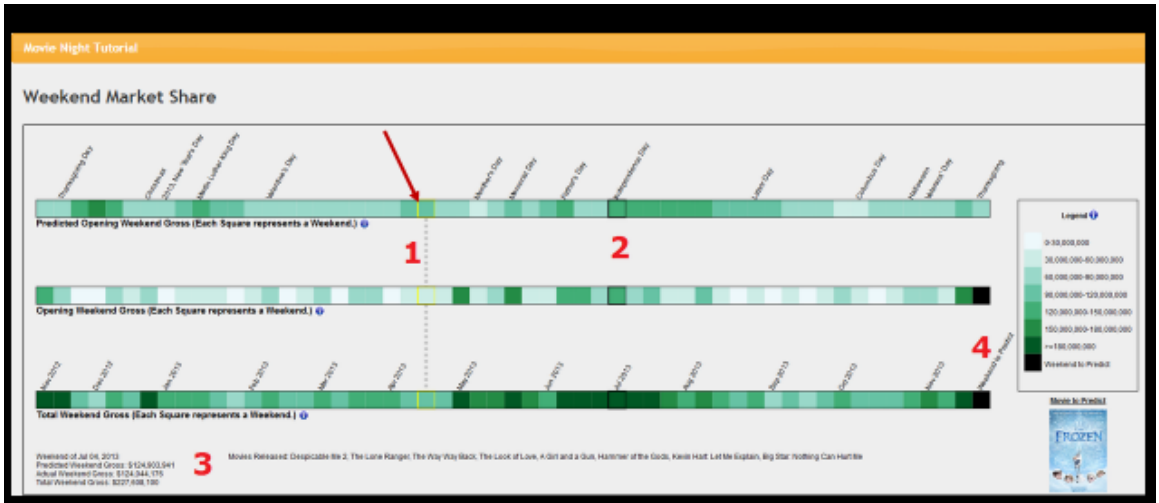
5. The 'Selected Movie' box displays the exact values of the OWG, the Predicted OWG, and the prediction range for a selected movie. A movie can be selected by clicking on the green circle or black square that belong to it.



The 'Weekend Market Share' page displays:

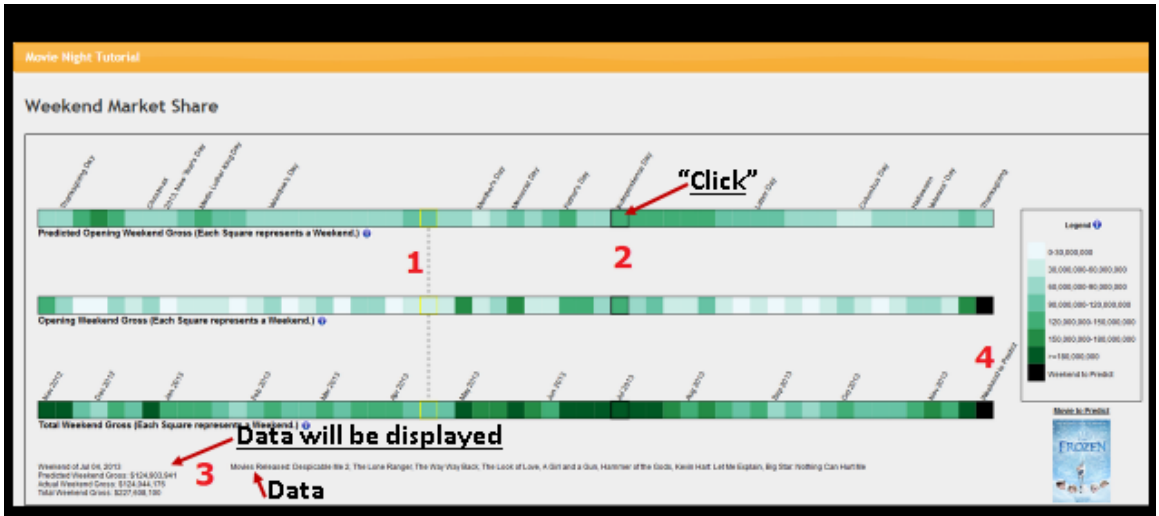
- Predicted Opening Weekend Gross (POWG)
- Actual Opening Weekend Gross (AOWG)
- Total Weekend Gross (TWG) for fifty-five weekends leading up to the weekend to predict.

Each square represents a different weekend and is colored based on its 'Weekend Gross Value' (refer to the legend for scaling).



1. When hovering over a square it will be highlighted in yellow and a reference line will be shown.

2. Clicking on a square will highlight it in black showing the information related to it (3).

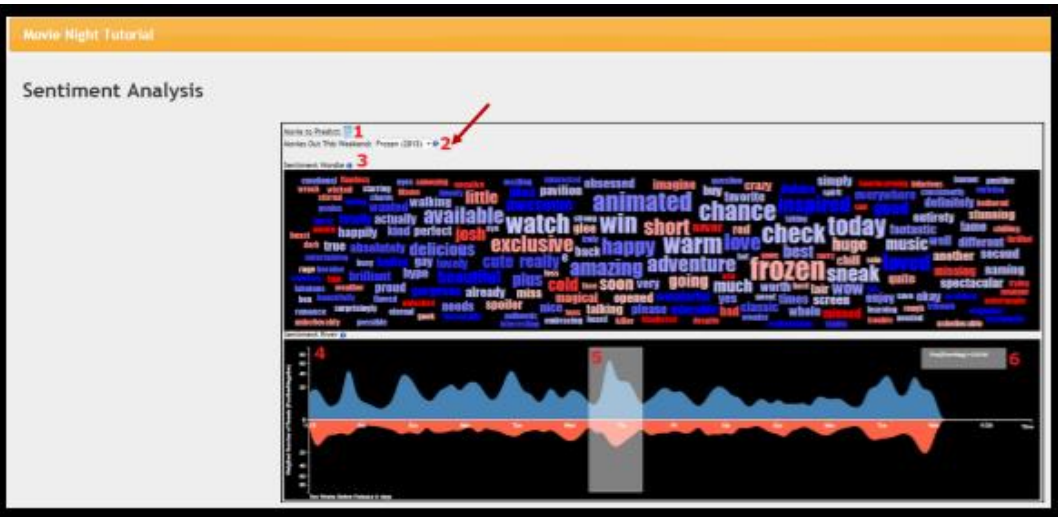


3. The date of the weekend clicked, PWG, AOWG, TWG, and the movies released that weekend are displayed here.

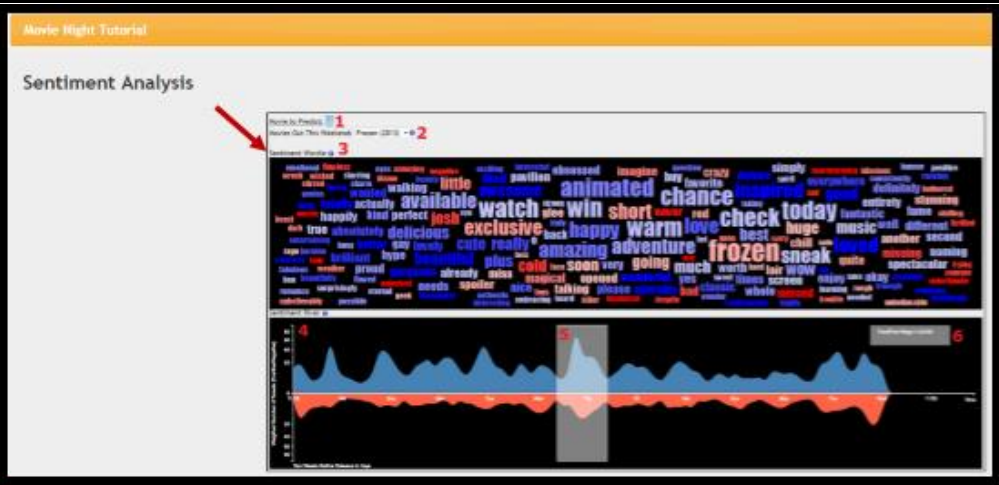
4. The "Weekend to Predict" squares for AOWG and TWG are colored black because those values are unavailable.

The 'Sentiment Analysis' displays Twitter tweets about the selected movie – in this case *Frozen*.

1. The 'Movie to Predict' image displays the movie to be predicted. Hover over the image to see an enlarged version of it.

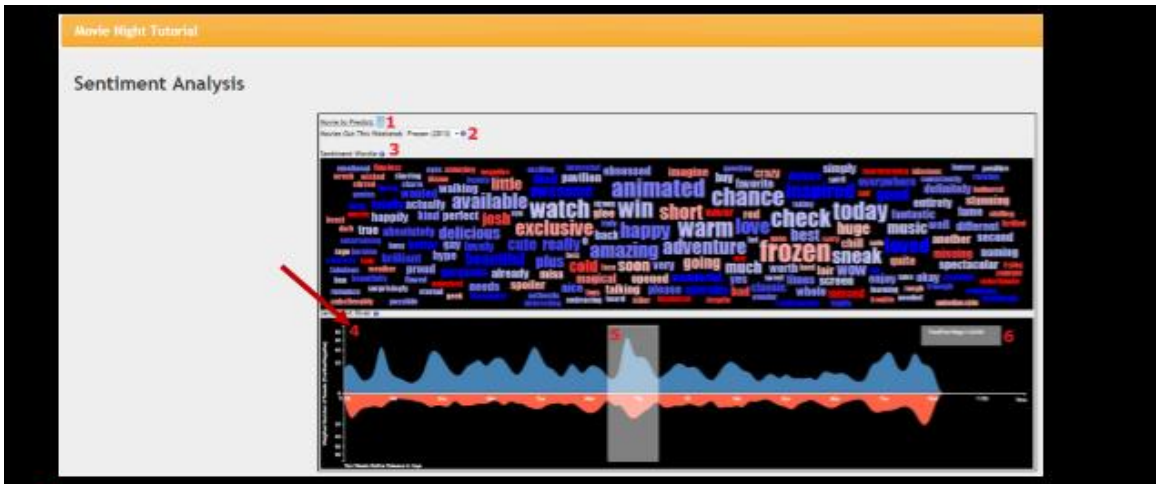


2. The 'Movies Out this Weekend' selector allows the user to switch between any of the movies released on the weekend of inspection.



3. The 'Sentiment Wordle' shows the most frequently used words in Twitter posts containing the selected movie.

- The larger the word is the higher the number of occurrences of the word.
- Positive words are in Blue and Negative words are in Red, the darker the hue the greater the sentiment.

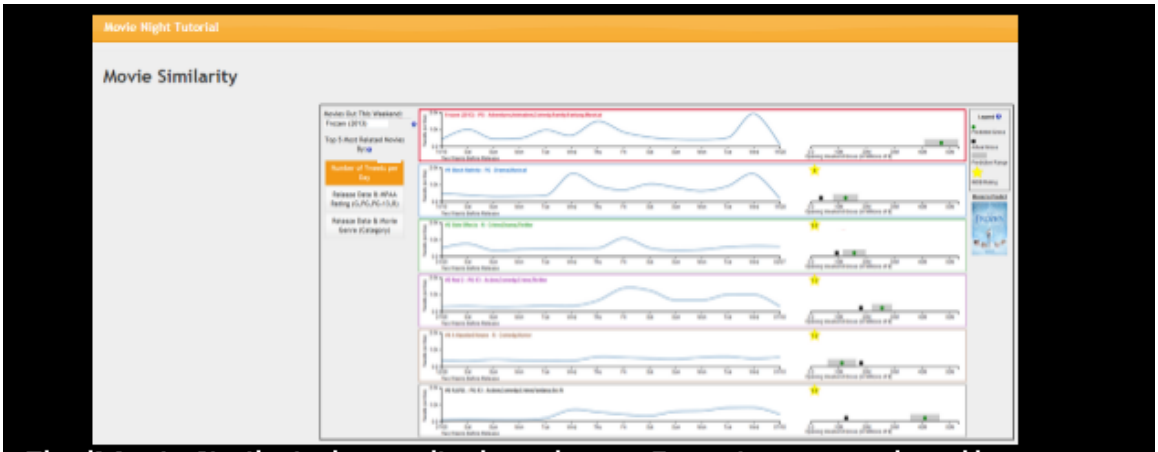


4. The Sentiment River shows the total sentiment score for both positive and negative at a given point in time for a two week period.

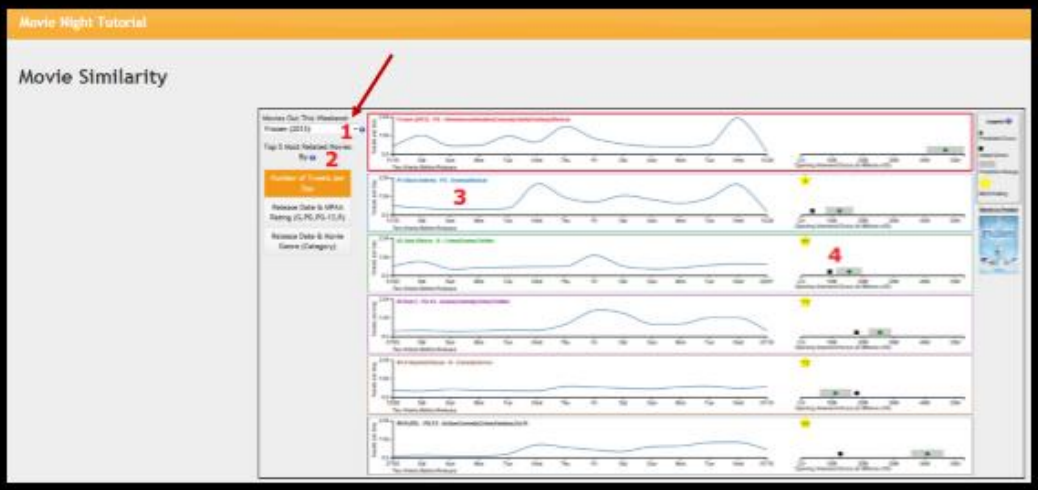
- Positive Comments are in Blue above the axis and Negative Comments are in Red below the axis.
- Highlight a region to see the positive to negative sentiment ratio.

5. Click and drag on the chart to highlight a region to see the positive to negative sentiment ratio.

6. Sentiment ratio, this shows how positive the overall sentiment is for a highlighted region.



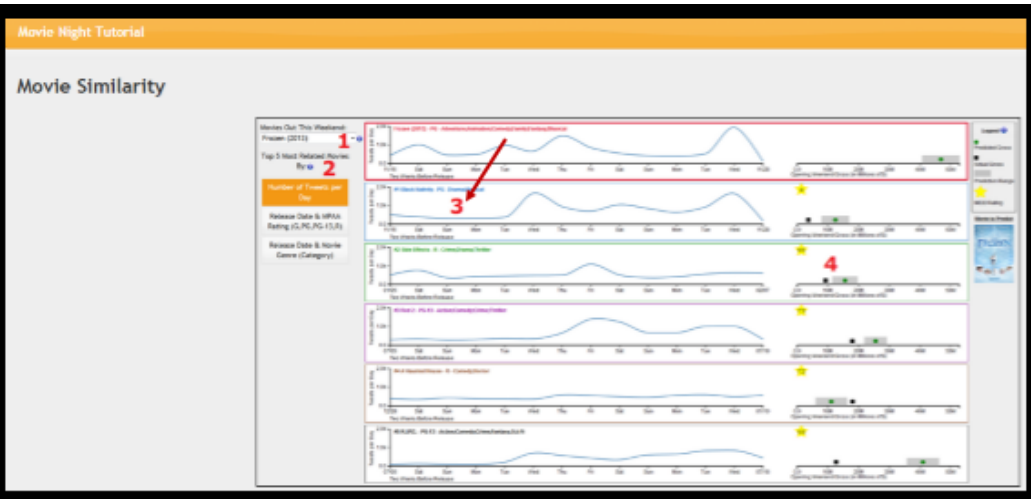
The 'Movie Similarity' page displays the top 5 movies most related by:
-the number of tweets per day
-the release date
-MPAA rating (G, PG, PG-13, R)
-the release data and movie genre (Category) for the movie selected.



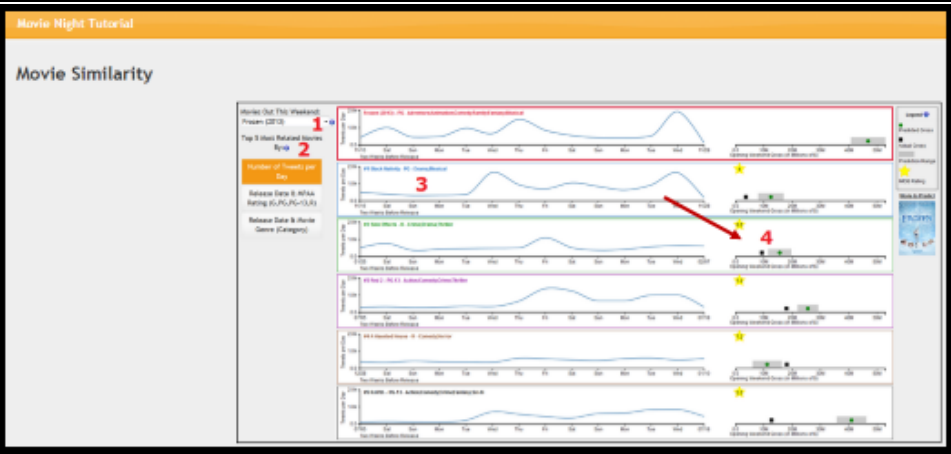
1. The 'Movies Out this Weekend' selector allows the user to switch between any of the movies released on the weekend of inspection.

2. The 'Top 5 Most Related Movies By' selector allows the user to choose how movies will be found in relation to the selected movie.

- Number of Tweets per Day: Movies with the most similar # of tweets per day will be found.
- Release Date & MPAA Rating: Movies with the most similar ratings in order of release will be found.
- Release Date & Movie Genre: Movies that are most similar in Genre/Category in order of release will be found.

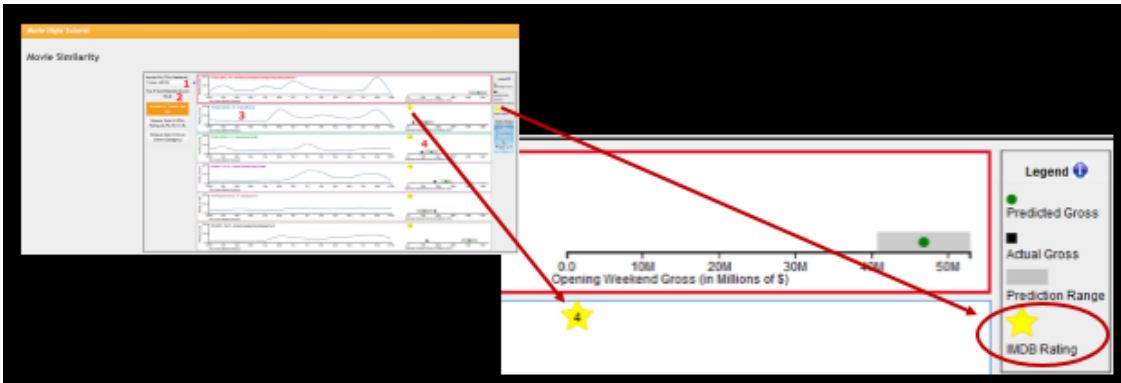


3. The line chart displays the movies' name and related information such as MPAA Rating and Genre/Category. The line chart also displays the number of tweets per day, in thousands, for two weeks before the movies release.

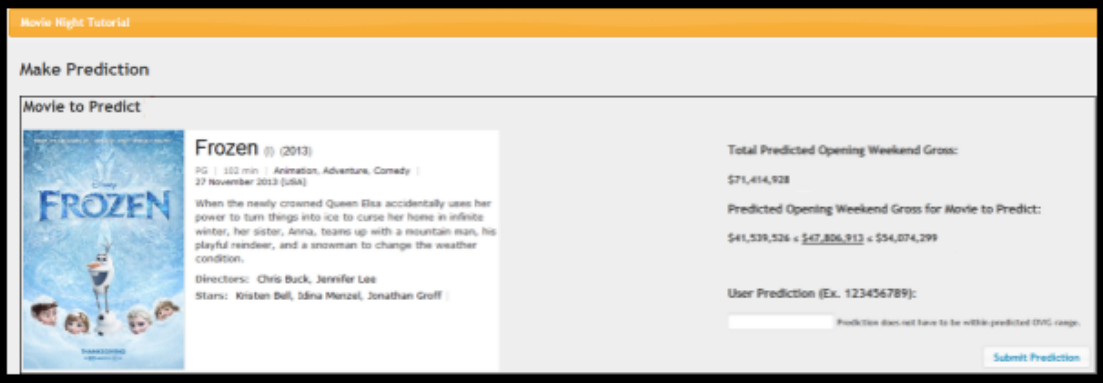


4. The bar chart displays the movie's rating from IMDB, the Actual Opening Weekend Gross (AOWG) as a black square, Predicted Opening Weekend Gross (POWG) as a green circle, and the Prediction Range as a gray rectangle.

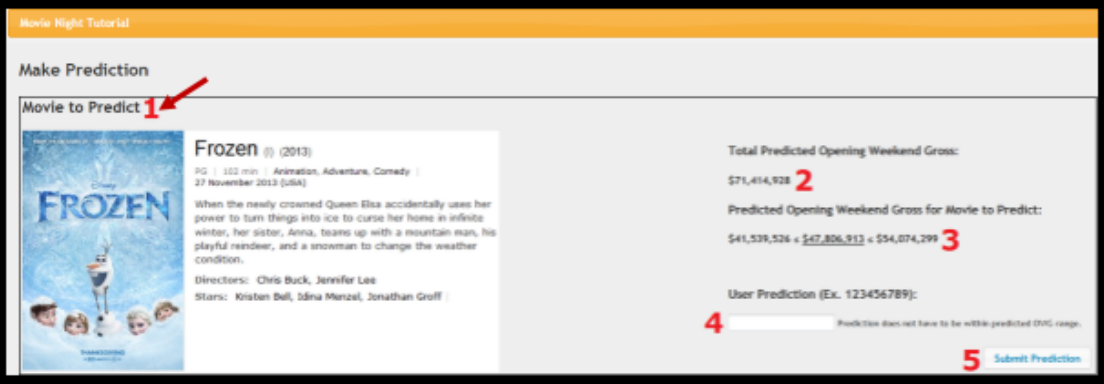
- Hovering over a green circle or black square will cause their exact values to appear in a tooltip.
- Note: The IMDB rating and the AOWG are not shown for movies released on the weekend to predict.



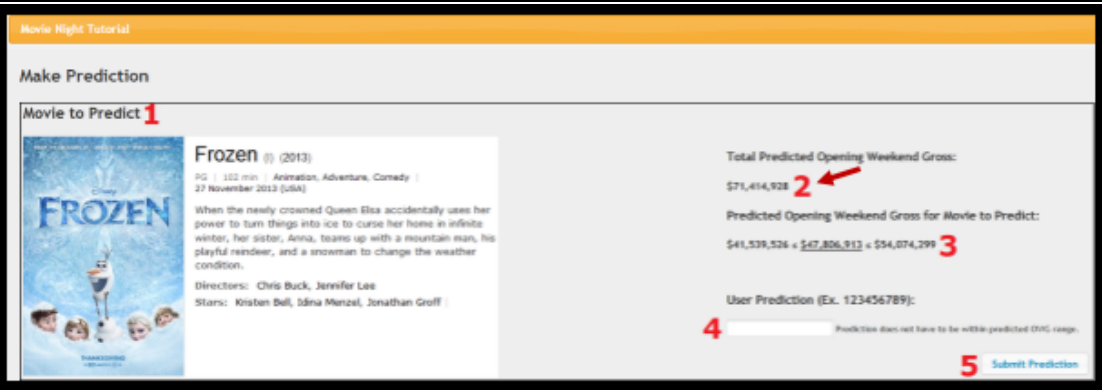
The Internet Movie Database – IMDB, is an online database for movies, television, and video games.
IMBD Rating Scale - an online rating scale from 1 to 10 (the higher the number the better the movie review).



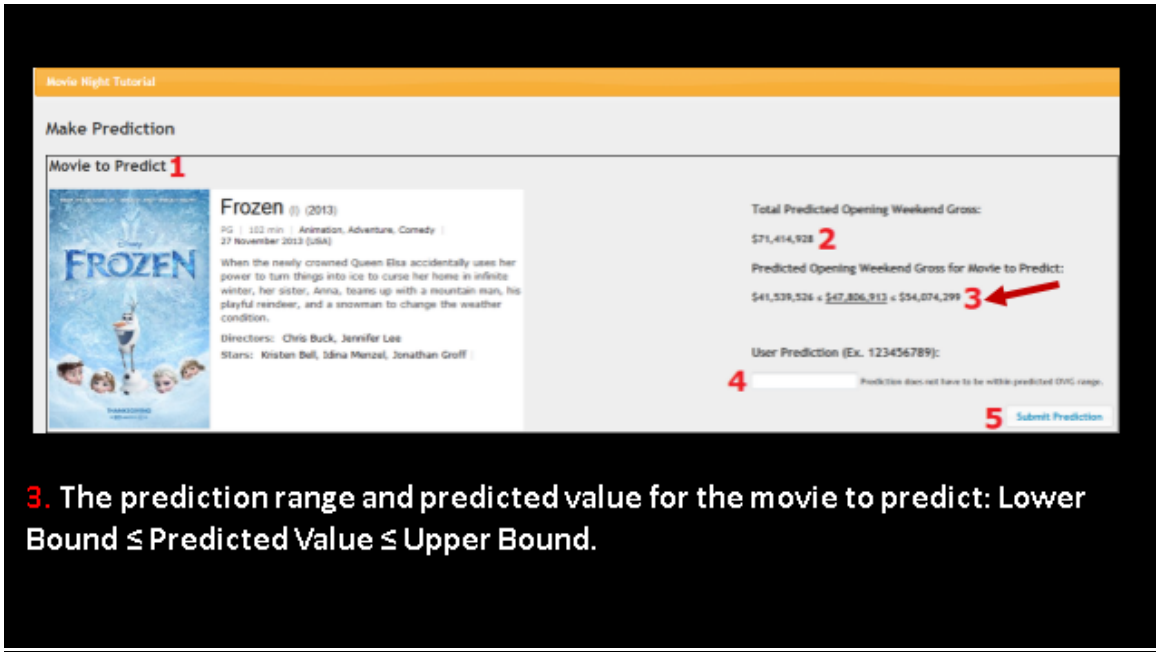
The 'Make Prediction' page displays:
- The movie the user will predict
- Opening Weekend Gross Values
- Allows the user to input and submit their prediction



1. Movie to Predict: Movie description about the movie to predict.



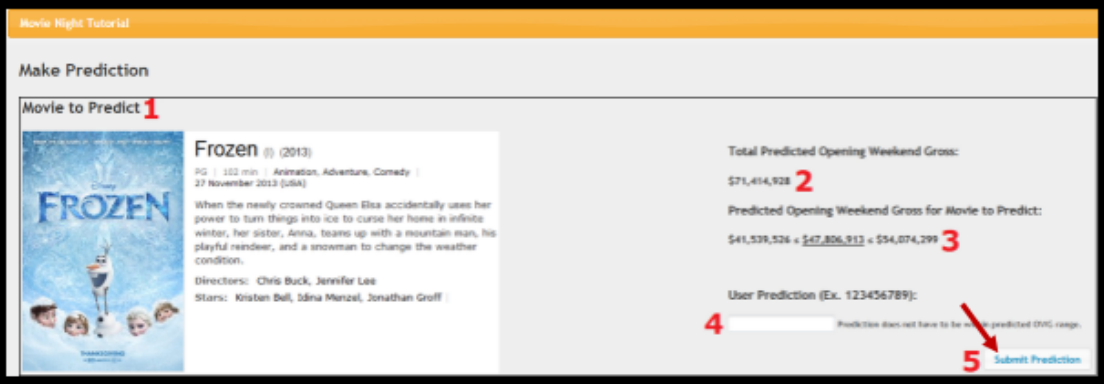
2. The 'Total Opening Weekend Gross' is the amount of money available for All movies to be released that weekend.
This amount is a prediction only. For example, if 3 movies are released the weekend in question, then it is predicted that movie A will make $35,345,457, movie B $20,956,103 and movie C will earn the remainder.



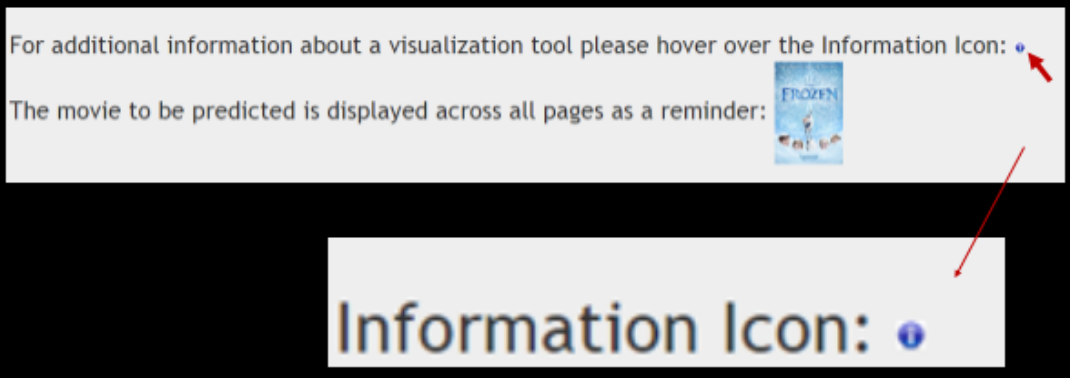
3. The prediction range and predicted value for the movie to predict: Lower Bound ≤ Predicted Value ≤ Upper Bound.

4. Input field for user to enter their prediction for the movie's Opening Weekend Gross. Please follow the example input: Do NOT put spaces, symbols ($, #, etc.), or commas (',') in your answer!

- The user prediction does not have to be within the Predicted Opening Weekend range.
- After the first movie prediction is submitted the 'Actual Opening Weekend Gross' will be displayed along with the user's prediction. This is only done for the first prediction submission.



5. The 'Submit Prediction' button is used to submit the user's prediction for each movie one at a time.



Congratulations!

You have completed the training module!

Next Steps

1. Take a short quiz.
2. Review the "Tutorial" using the actual movie interface.
3. Make your first prediction (practice prediction).