Title

Essays in Game Theory and Experimental Economics Relating to Social Choice, Criminal Justice, and Education

Author

Bloxsom, Mark Justice

Publication Date

2022

UNIVERSITY OF CALIFORNIA,
IRVINE

Essays in Game Theory and Experimental Economics Relating to Social Choice, Criminal Justice, and Education

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Economics

by

Mark Justice Bloxsom

Dissertation Committee:
Professor Michael McBride, Chair
Professor Stergios Skaperdas
Professor John Duffy

2022

DEDICATION

To

my family and friends

in recognition of their support

my kids

in recognition of their love

the universe

in recognition of the many trials and tribulations that I have endured in the pursuit of this degree and how strong this process had made me. I'm a man in need of a beer, a nap, and a shower, but I'm much wiser now as well.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express the deepest appreciation to my committee chair, Michael McBride. Without your insight, support, and positive attitude I could not have possibly completed this journey. I've learned more from you about economics, research, and persistence than I ever imagined. I owe you an immense debt of gratitude that I probably can never repay.

I want to thank Stergios and John for all their help and input. Their perspectives on my research questions as well as my approach to these papers have provided immense clarity as well as a variety of other thought-provoking conversations and questions. I'm incredibly grateful to have worked with you both, including more than a decade of collaboration with Stergios.

Chapter two of this dissertation was a collaborative effort between Mike McBride, Ron Breiger and Brint Milward of The University of Arizona, and me. Without the help of Mike, Ron, and Brint it would not have been possible to develop such a rigorous study of dark networks. It was a pleasure to work with them and it is my single favorite chapter of this dissertation as a result of our work together. I also want to thank all of the participants at the many workshops where this work was presented as well as the financial support we received from the Air Force Office of Scientific Research Award No. FA9550-10-1-0569 and the Army Research Office Award No. W911NF-11-1-0332.

Chapter one and two of this dissertation were completed using the facilities and resources of the Experimental Social Science Laboratory at the University of California, Irvine. I want to thank all of the administrators and colleagues who helped to make my experimental sessions possible.

Finally, a thank you to the incredible faculty and graduate students at The University of California, Irvine that I've had the chance to work alongside. I can't possibly express my gratitude for the lessons I've learned and the person that I've become as a result of my time at UCI. I'll always remember meeting my cohort in math camp, our early study groups, our celebration after passing comps and advancing to candidacy, my game theory and public choice field courses, the camaraderie of the Experimental Social Science Laboratory, and my time as a pedagogical fellow in the UCI Teaching, Learning and Technology Center.

VITA

Mark Justice Bloxsom

2005	B.S. in Computer Studies, University of Maryland University College
2011	B.S. in Economics, University of North Carolina, Charlotte
2011-13	Teaching Assistant, School of Social Sciences, University of California, Irvine
2013	M.A. in Economics, University of California, Irvine
2013-15	Graduate Student Researcher, Economics Department, University of California, Irvine
2015-22	Assistant Professor, Business Department, Alfred State SUNY College of Technology
2022	Ph.D. in Economics, University of California, Irvine

FIELD OF STUDY

Game Theory, Public Choice, and Experimental Economics

PUBLICATIONS

Demler, Erica & Nardozzi, Stephanie & Bloxsom, Mark & Moore-Russo, Deborah. (2018). Instructional Design and Facilitation as Contributors to Social Presence in Asynchronous Online Discussion Forums, BRC Journal of Advances in Education.

ABSTRACT OF THE DISSERTATION

Essays in Game Theory and Experimental Economics Relating to Social Choice, Criminal Justice, and Education

by

Mark Justice Bloxsom

Doctor of Philosophy in Economics

University of California, Irvine, 2020

Professor Michael McBride, Chair

This dissertation answers three questions in social choice, criminal justice, and education. Chapter 1 considers the possibility that changing the voting system used in a simple electoral environment might have a significant impact on the choice to participate when participation is costly. I find that the participation does significantly increase in an evaluative electoral system compared to a plurality voting system. Chapter 2 investigates the trade-off between centralization and productive complementarities in the creation of a dark network which is susceptible to disruption by outside elements. We find that networks become more decentralized as more of the network is visible to disruptors and they become more centralized as the productive incentives from centralization become stronger. Chapter 3 theorizes about the influence that different grading systems have on the effort choices of the students impacted by the grading system relative to their abilities. I find that eliciting high effort from all students is much easier in a more homogenous student population, but in a more heterogenous environment the absolute grading system offers more opportunity for high effort outcomes.

Introduction

Mechanisms are all around us. Some are obvious like the physical machines that brew coffee or ferry us to and from work each day, while others are less obvious like the ballot that we use to cast our votes in elections or the way law enforcement disrupts criminal social networks. Often times we take these less obvious mechanisms for granted as "the best" way to operate or simply just "the way we've always done it." Very few individuals devote a great deal of thought to how the everyday mechanisms we rely on can actually be influencing or even distorting the choices we make and the outcomes we experience. This dissertation is intended to serve as a brief study of three important mechanisms as well as their impact on the choices of those engaging with these mechanisms.

1 An Experiment in Costly Participation Under Different Voting Systems

1.1 Introduction

Many researchers have suggested that the voting system itself may play a significant role in voter turnout. Several researchers have suggested that moving from a plurality voting system to an approval or ranked choice voting system could contribute to significantly higher voter participation.[1] The predominant assertion to justify this belief is that when voters are given a more complete means to express their preferences, they will be more likely to vote. This claim clearly encompasses a combination of psychological as well as economic motivations which are difficult to differentiate in a real-world or field environment. This paper tests the assertion that an alternative voting system, specifically an evaluative voting system which offers a richer set of possible ballot expressions, will lead to a greater participation rate than a traditional plurality voting system in a laboratory en-

---

[1] See Brams and Fishburn (2007), Balinski and Laraki (2011), Forsythe et al. (1996), Brams and Herschbach (2001), Laslier (1997), Laslier and Sanver (2010), Felsenthal (1989), Alcantud and Laruelle (2013)

vironment where economic motivations can be more easily understood and controlled. The results demonstrate significant support for this claim, although the political environment as well as the candidates themselves would have an undeniable effect on the external validity of these findings.

A variety of different voting systems have been suggested as possible alternatives to plurality voting with variations on approval voting, ranked voting, and range voting dominating the conversation in the United States. Economists, mathematicians, and political scientists have studied and theorized about the benefits and drawbackss of each. It has been echoed by nearly all alternatives to plurality voting that alternative voting systems offer a more comprehensive ballot which would result in greater voter participation. Balinski and Laraki (2010) mention increased voter turnout as a feature of several alternative voting systems. Brams and Fishburn (2005) explicitly argue that using the approval voting system would increase voter engagement and by extension voter turnout. Alcantud and Laruelle (2013) argue that the evaluative voting system used in this experiment would also result in an enriched set of options for the voter and therefore increased voter participation. These arguments tend to draw on one of two mechanisms, the expressive value of a more comprehensive ballot or the utility value of the set of ballot options relative to the choice to abstain.

The expressive value of a more comprehensive ballot comes from the voters ability to more accurately express their preferences for each candidate. Expressive value is derived from the action of voting and not from the outcome of the election itself therefore pivotality is less influential and the completeness of ones expression is more influential. Approval and ranked voting allow for greater expressivness across all candidates while range and evaluative voting additionally offer greater expressiveness within the evaluations for each candidate. The set of possible ballots in a plurality system are typically contained within the sets of possible ballots for more expressive systems. The purpose of conducting this experiment in a controlled laboratory environment with apolitical mechanisms is in part to remove this expressive value and allow the focus to be narrowed to the utility value of the evaluative ballot.

The utility value of the set of ballot options relative to the choice of abstention is of great importance since it does not require any political influence or consideration in order to demonstrate the value of this alternative voting system. The argument is simply that there is greater utility in having a broader set of possible ballots for voters to choose from regardless of the candidates or electorate themselves. There could also be increased value in this alternative system when certain electoral conditions exist, such as the presence of a spoiler candidate, where the more comprehensive ballot could allow for more honest voting and representation of preferences for all candidates (Brams and Fishburn, 2007). In all voting systems, absent a government mandate to vote, the option of abstention is possible and costless. This suggests that the voter will only participate if there exists a ballot in the set of possible ballots for that voting system whose expected utility for the voter exceeds the expected utility of abstention. Naturally, a larger set of possible ballots would increase the likelihood that such a ballot exists.

This paper reduces the process of voting to a simple economic and probabilistic analysis of costs and benefits in order to identify the impact that the voting system may have on these calculations in a multicandidate environment. The primary hypothesis is that introducing a more complete set of ballot expressions allows the voter to increase their expected benefits of voting and should therefore result in greater participation overall in addition to participation at a higher cost threshold. The ability to express preferences for multiple candidates should result in a form of electoral insurance where risk, as well as strategic voting, is reduced and participation is increased. In addition, in an electoral scenario where a spoiler candidate is present, the value of the 2nd best option is increased and an increased participation effect under the evaluative voting system should be observed. Finally, the order with which the treatments are applied should not significantly influence the voter behavior as the treatments themselves are independent. Although this claim is not ultimately supported by the data, it is most likely a function of the finite nature of the experiment and the influence of the free-rider effect that arises in all public goods experiments. Regardless of treatment order, the amount of participation declines over time as subjects observe that it is possible to simply free ride on the participation of the other subjects.

Section 1.2 of this paper will outline the evaluative voting system used for this experiment.

Section 1.3 will define the key hypotheses which have guided this research. Section 1.4 will explain the experimental environment, the stages of the game itself, and the descriptive statics of the subjects. Section 1.5 will discuss the results of the experiment as well as the anomalies of interest contained within the data set. It will also discuss some of the comments made by the subjects in the end of session questionnaire and how they correspond to the common popular arguments for and against alternative voting systems. Section 1.6 will provide some general conclusions and suggestions for future research along these lines.

1.2 The Evaluative Voting System

This paper uses the evaluative voting system labeled by Alcantud and Laruelle as Dis&Approval voting. This system was chosen for several reasons, but to justify those reasons it is important to first clarify the complexity of voting in general. The process of voting entails the construction of a preference relation among the candidates as well as a preference expression on the ballot. Due to the existence of strategic voting and the inherent weaknesses in all voting systems, these two endeavors will be correlated, but not identical (Baugard and Irgersheim, 2011). Some voting systems, such as a traditional plurality vote, require the expression to be reduced to a single action on a single candidate while others, such as approval or rank choice voting, allow for a more comprehensive expression across multiple candidates. According to Michel Balinski and Rida Laraki (2011), a more effective way to think of voting is as an independent evaluation of each candidate and not simply a comparison of options.

The form of evaluative voting used in this experiment incorporates three possible evaluations on a discrete scale of 1, 0, or -1 and can be interpreted as approval, indifference, and disapproval. The reason for using this system is because it offers a richer set of ballot expressions in two key dimensions; evaluative voting uses a more comprehensive expression similar to approval voting in that every candidate is independently evaluated as opposed to a single evaluation on a single candidate, it also captures a more robust form of expression in that it allows the voter to explicitly convey approval as well as disapproval and indifference. Several recent papers have documented the psychological influence of disapproval or negative values when expressing preferences (Baujard et al., 2020; Harfst et.al, 2021). The aim in using this voting system is to contrast, in the most important or desirable aspects of ballot expression, the plurality ballot with a reasonable alternative. In addition, the evaluative voting system is a practical alternative, easy to understand/interpret, and used in many other capacities including popular culture and European elections (Baujard et al., 2018).

The theoretical literature on voter participation flows from the seminal works of Downs (1957), Tullock (1967), and Riker and Ordeshook (1968) through the work of Ledyard (1981, 1984), and Palfrey and Rosenthal (1983, 1985). Since the work of Palfrey and Rosenthal, there have been many theories offered to address Down's "paradox of not voting" and several criticisms leveled towards the rational choice theory of voting in general. Several papers have done a thorough job of summarizing the most influential research in voter participation since the 1980s.[2] Although this paper is not specifically intended to contribute to this theoretical literature, it draws from the models and experiments of Palfrey and Rosenthal as well as a general model of elections presented by Myerson and Weber (1993) which compares plurality, approval, and borda count voting systems in multicandidate elections. The participation game used in this experiment is also similar to those used by Levine and Palfrey (2007), Schram and Sonnemans (1996), Duffy and Tavits (2008), and many more. The unique contributions of this paper are that the subjets have three options, randomly assigned unique individual costs, and the use of the evaluative voting system with negative values.

It is difficult to measure the impact that a specific voting system may have on voter behavior in a real world electoral event or in a field experiment due to the confounding institutional, political, and socioeconomic factors that contribute to voter behavior (Levine and Palfrey, 2007). Several avenues of voting research have theorized about the expressive nature of voting, the motivational nature of political parties or political power, and the structural nature of elections themselves (Francois and Gergaud, 2019). This laboratory experiment significantly reduces many of these confounding factors by controlling the economic incentives and eliminating many social or political incentives. A wide variety of papers in the recent economics and political science liter-

---

[2] See Medina Sierra (2018), Blais (2000, 2006), Blais et al. (2016), Dhillon and Peralta (2002), Feddersen (2004), Feddersen and Sandroni (2006), and Geys (2006)

ature have used field experiments (Van der Straeten et al, 2010) or in situ experiments in French or other European elections [3] to analyze alternative voting systems, but this paper is the first to be able to combine the use of an evaluative voting system with a costly participation decision in a laboratory setting. This experiment was conducted online so there was an additional level of anonymity afforded to subjects in that they never observed any physical, economic, social or political characteristics of the other subjects.

1.3 Hypotheses

Based on the research mentioned above, the experiment was designed with three specific hypotheses in mind. The purpose of this experiment was to test for and measure any differences in participation that resulted from a change in voting system. The intention was to keep the environment and the task as similar as possible and only vary the way that subjects expressed their preferences. The existing literature, as suggested above, indicated that a more complete ballot would result in greater participation so this is the first hypothesis.

Hypothesis 1. *The evaluative voting system will result in greater participation than the plurality voting system.*

An alternative way to look at this realization is to consider the fact that the greater expected utility gain from the evaluative ballot suggests that voters will also be willing to participate at higher individual costs in that voting system. This perspective on the differences between voting systems motivates the second hypothesis of this paper.

Hypothesis 2. *On average, the individual cost of voters who have chosen to participate will be greater in the evaluative system than in the plurality system.*

Another important difference that exists as a result of the contraints placed on the plurality ballot is that the voter has no means by which to express support for an alternative or 2nd best candidate. This inability to demonstrate positive support for more than one candidate incentivizes

---

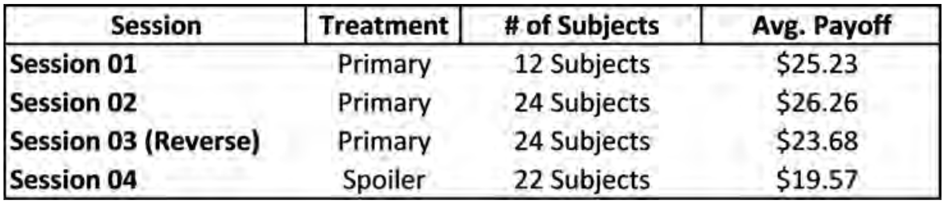
[3] See Baugard and Igersheim (2011), Baugard et al. (2014), Laslier and Van Der Straeten (2007), Darmann et al. (2017)

more strategic voting and results in an inferior expected utility gain when a second candidate, referred to as a spoiler candidate, exists which might cause the electorate to split their votes across the two candidates resulting in a third, strictly less preferred candidate, winning the election. The comprehensive nature of the evaluative ballot instead allows voters to place positive support on more than one candidate which increases the likelihood of a preferred candidate winning as well as reducing the likelihood of the least preferred candidate being victorious. In other words, the expected utility of ballots which capture this expression would be greater than is normally expected in an election that does not include a spoiler candidate. This suggests that when a spoiler candidate is present in a three candidate election, the evaluative voting system will result in an even greater participation rate relative to the plurality system. This observation motivates the third and final hypothesis of this paper.

Hypothesis 3. *The introduction of a spoiler candidate into the preference structure will magnify the effect of the evaluative voting system on voter participation.*

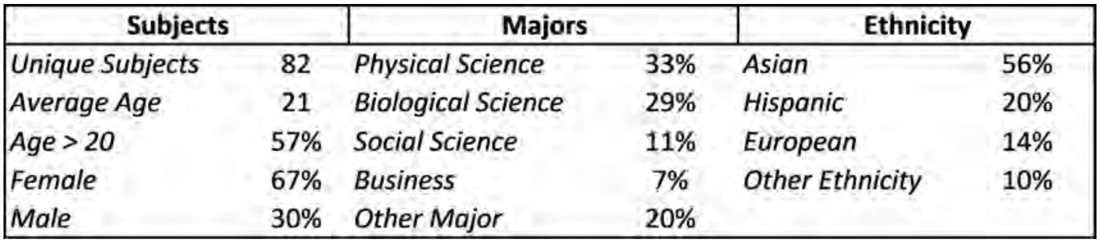
1.4 Experimental Design

The experiment was conducted in multiple sessions online using the oTree open-source framework (Chen et. al, 2016). Subjects were recruited using the subject pool of the Experimental Social Science Laboratory located at the University of California, Irvine (UCI). The experimental sessions were advertised to interested participants through the online portal as well as through email notifications. Interested subjects registered to participate, were given a unique subject ID, and were provided with the link to the experiment portal after they had submitted their payment information. Subjects were only allowed to participate in one session. Payments were made through the Venmo and Zelle apps following the completion of the session. The first session consisted of one group of 12 subjects while all subsequent sessions consisted of two groups of 12 subjects. The only exception was the final session which incorporated a different treatment referred to as, "the spoiler treatment." The spoiler treatment had a different distribution of types which necessitated two groups of 11 instead of 12. Subjects received experimental units referred to as "points" and the conversion of points to dollars ($) was 100 points for $5. Subjects were informed of this conversion rate prior to starting the experiment. Each session lasted between 65 and 85 minutes. The average payoff earned was $18.45 plus a $7 show-up payment with a minimum experimental earning of $2 and maximum of $30. Table 1 and Table 2 below provide basic information about the sessions and subjects.



Session	Treatment	# of Subjects	Avg. Payoff
Session 01	Primary	12 Subjects	$25.23
Session 02	Primary	24 Subjects	$26.26
Session 03 (Reverse)	Primary	24 Subjects	$23.68
Session 04	Spoiler	22 Subjects	$19.57

Table 1: Session Information



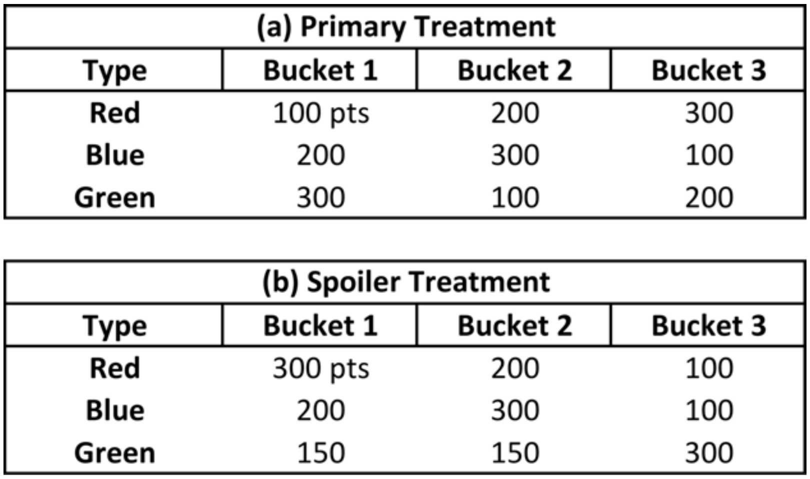
Subjects		Majors		Ethnicity	
Unique Subjects	82	*Physical Science*	33%	*Asian*	56%
Average Age	21	*Biological Science*	29%	*Hispanic*	20%
Age > 20	57%	*Social Science*	11%	*European*	14%
Female	67%	*Business*	7%	*Other Ethnicity*	10%
Male	30%	*Other Major*	20%		

Table 2: Demographic Information

1.4.1 A single round

Each round consisted of a costly choice to participate or abstain followed by the opportunity to place tokens in buckets where the bucket with the most tokens at the end of the round was declared the winner and payoffs were distributed accordingly. Screenshots of the participation choice and game screens are included in Appendix A. All subjects were randomly assigned a type denoted by a color (Red, Blue, or Green) which they retained across all rounds and all voting systems in order to reduce the need for learning a new payoff table during the experiment. Each subject was also randomly assigned a cost drawn from a discrete uniform distribution, U(0,100) at the beginning of each round. The subjects were shown their cost, their type, and the payoff matrix each round prior to making their participation decision. They were also reminded at the beginning of each round of the distribution of types within the group that they were engaging with. Each subject participated in two 10 round phases preceded by three practice rounds each. Both phases used the same payoff table, but with different voting systems. The payoff table is shown below as Table 3(a). The payoff table was designed to provide each participant with a clearly defined ordinal ranking of the buckets. The definition of a clear first and second choice was essential to understand the competitive and strategic motivations of each participant.

One phase is modeled on the traditional plurality vote where the participant is allowed to place a single token in one of three buckets of their choosing. The other phase is modeled on an evaluative voting system where the participant is allowed to add a token, remove a token, or ignore each of the three buckets. Three sessions were conducted with the plurality phase first followed by the evaluative phase and two sessions was conducted with the evaluative phase first followed by the plurality phase in order to identify any order effects that may exist. Each session consisted of three practice rounds and 10 scoring rounds for each phase followed by a short questionnaire for demographic and qualitative information. In all sessions, subjects were regrouped every round, but retained their original type and each group retained its type distribution at all times. The winnings for each round were determined by the winning bucket and each subject's color type. If two or more buckets were tied then the payoff to each subject was the average of the two or more respective buckets. This equates to the standard in the voter participation literature of flipping a coin in the case of a tie. The final payoff for the round consisted of the subject's winnings less their participation cost if they chose to participate. At the end of each phase, one of the 10 scoring rounds was chosen at random and distributed to the subject as payment for that phase. The subject was shown a history table of their choices in previous rounds as well as the results of each round before advancing to the next round.



(a) Primary Treatment			
Type	**Bucket 1**	**Bucket 2**	**Bucket 3**
Red	100 pts	200	300
Blue	200	300	100
Green	300	100	200

(b) Spoiler Treatment			
Type	**Bucket 1**	**Bucket 2**	**Bucket 3**
Red	300 pts	200	100
Blue	200	300	100
Green	150	150	300

Table 3: Payoff Table

Although the mechanics of the primary treatment and the spoiler treatment were identical, two different payoff tables were used and the type distribution within each group was slightly different. The primary treatment was a Condorcet cycle with equal distribution of types, four subjects in each type, so there was no clear advantage for any type. The spoiler treatment had a type distribution of five (5) green subjects and three (3) subjects of each of the blue and red type. The spoiler treatment used the payoff table shown as Table 3(b) and was designed to increase the value of the second best option for both the blue and red types which should result in a greater treatment effect for the evaluative voting system. If the red and blue types were able to coordinate their efforts then they could dominate the preferences of the green types, but without this coordination the green types would benefit from the divided interests and would be victorious.

This design accomplished a few important tasks. It reduced learning within each treatment by providing practice rounds and constant types as well as payoff tables. It placed each subject in both voting systems under identical conditions in order to remove external factors. It also created variability through the random assignment of participation costs so each subject's ability to accurately predict the level of participation within the other types was reduced.

1.5 Results and Discussion

The primary question that this experiment was designed to address was whether or not changing from a plurality voting system to an evaluative voting system would have a significant impact on participation when participation is costly. This suggests that the first comparison is a simple test of participation rate across the two voting systems. Figure 1 below shows the relative value of each along with a 95% confidence interval. The evaluative system clearly has a higher participation rate, but there is significant overlap between the confidence intervals so more granular statistical analysis is necessary.

Each subject participated in 10 rounds of one voting system followed by 10 rounds of the other system and when comparing the first 10 rounds to the second 10 rounds regardless of voting system, there is a clear reduction in participation for the second 10 round period. This is displayed in Figure 2 below. This figure also suggests that the reduction in participation may be taking place within each 10 round session so Figure 3(a) shows the rate of participation by round by voting system in the first 10 rounds and Figure 3(b) shows the same for the second 10 rounds. It is visually obvious that participation is significantly changing based on the round number with later rounds showing less participation regardless of voting system. This is not surprising when considering that the costly participation would encourage a great deal of free-riding, especially when combined with the balanced nature of the distribution of types. These findings suggest that

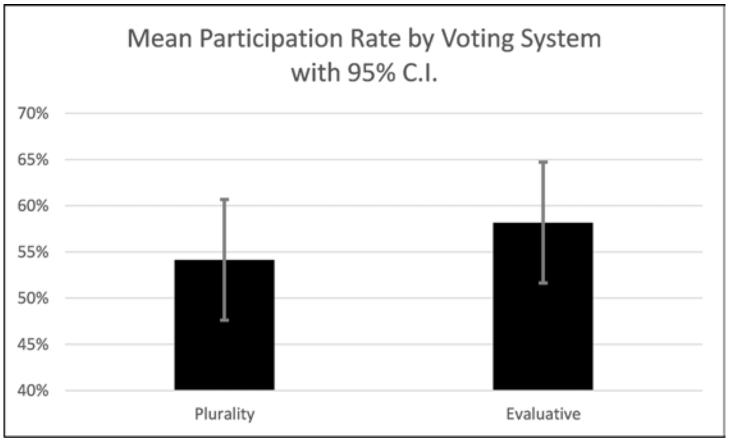


Figure 1: Participation by Voting System

there needs to be a mechanism to control for the specific round so each round was numbered from 1 to 26 which included the three practice rounds prior to each 10 round session.

Several other factors were taken into consideration in regards to the participation choice made by each subject in each round. Specifically, the outcome of the previous round seemed to have a significant impact on the participation decision in each subsequent round. There were two dimensions that stood out as obvious considerations, whether the subject's preferred candidate was victorious in the previous round and the subjects realized payoff in that round. The assumption was that these would both be positively correlated with participation as the subject enjoyed the "winning effect" and the dopamine that comes with winning. It is important to note that winning could also exacerbate the free-rider effect suggesting that a subject could continue to enjoy high earnings without incuring the cost of participation in the next round. In either case, a dummy variable labeled "First Choice" was created to capture whether a subject's preferred candidate was victorious in the previous round, and the player's payoff from the previous round was retained for the subsequent round in a variable called "Previous Payoff."

1.5.1 Hypothesis 1

Hypothesis 1 states that participation should be higher in the evaluative voting system. Previous work suggests that the participation decision made by the subject is a comparison between the

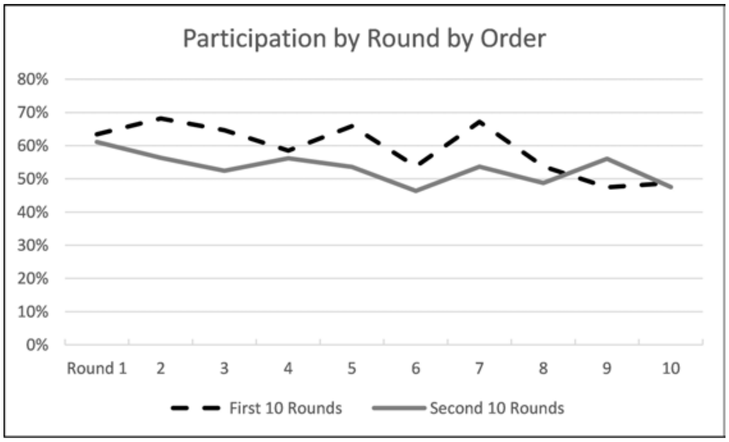


Figure 2: Participation by Order

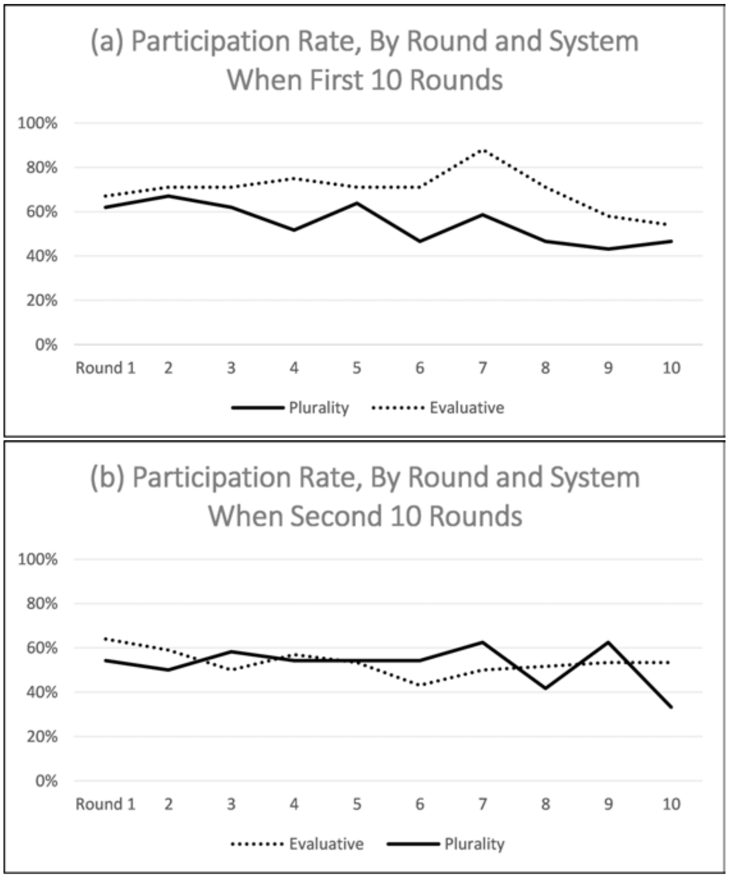
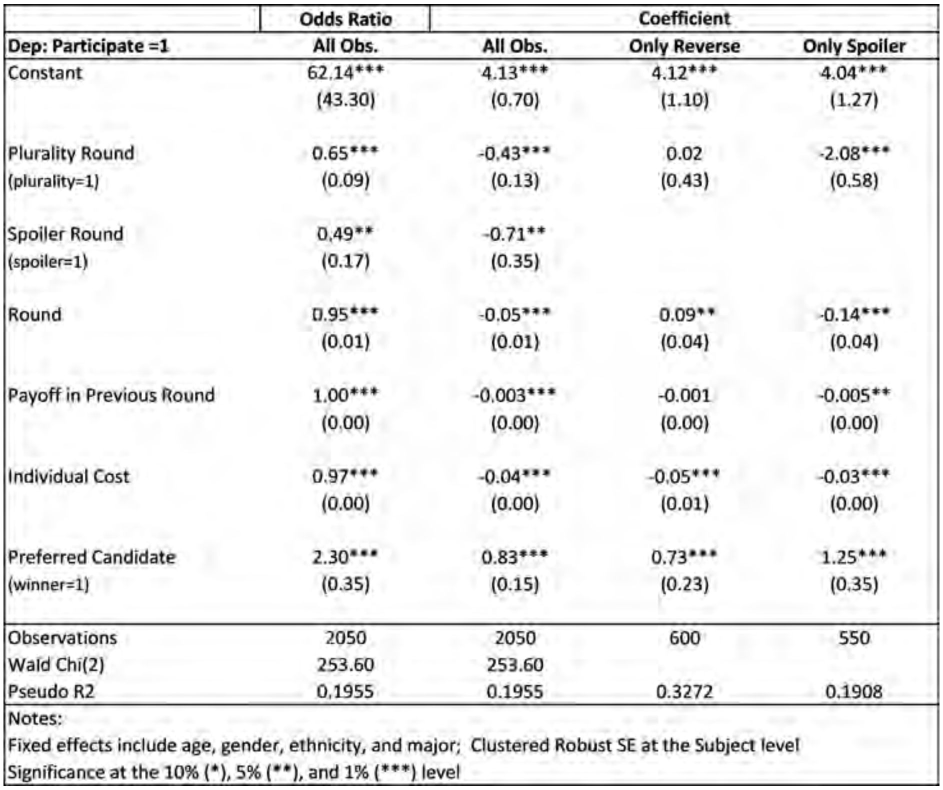


Figure 3: Participation by Round and System

expected payoff from participation and the realized individual cost for that round. Given all of the additional behavioral factors that may have influenced this decision, it is appropriate to use regression analysis to control for these factors as well as typical fixed effects such as age, gender, ethnicity, and college major. A logistic regression was used due to the binary and probabilistic nature of the dependent variable in question. The results of these regressions are shown in Table 4 below. As evidenced in the first two columns, we observe that the likelihood of participation falls when the subject is using the plurality voting system. In fact, across all observations in this experiment we see that the odds of participation are about 35% lower in the plurality voting system when controlling for variables such as whether that round was a practice round or a scoring round, the size of the payoff in the previous round, the individual cost to that subject in that round, and whether that subject's preferred choice of candidate won in the previous round. We also observe that the coefficients on many of these variables are statistically significant and their correlation to participation is in a predictable manner. Subjects are slightly more likely to participate in the early practice rounds as they learn the system and get a feel for the other subjects in their voting group. Participation declines as each round progresses and is inversely related to the individual's private cost, but we also observe that participation falls as the payoff in the previous round increases which supports the free-rider mentality and the desire to avoid the participation cost if possible without sacrificing too much of their payoff. We also observe that the "winner effect" seems to be in play resulting in an increase in participation when the subject's preferred candidate was victorious in the previous round. This evidence clearly supports hypothesis 1.



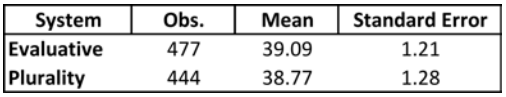
	Odds Ratio	Coefficient		
Dep: Participate =1	**All Obs.**	**All Obs.**	**Only Reverse**	**Only Spoiler**
Constant	62.14***	4.13***	4.12***	4.04***
	(43.30)	(0.70)	(1.10)	(1.27)
Plurality Round	0.65***	-0.43***	0.02	-2.08***
(plurality=1)	(0.09)	(0.13)	(0.43)	(0.58)
Spoiler Round	0.49**	-0.71**		
(spoiler=1)	(0.17)	(0.35)		
Round	0.95***	-0.05***	0.09**	-0.14***
	(0.01)	(0.01)	(0.04)	(0.04)
Payoff in Previous Round	1.00***	-0.003***	-0.001	-0.005**
	(0.00)	(0.00)	(0.00)	(0.00)
Individual Cost	0.97***	-0.04***	-0.05***	-0.03***
	(0.00)	(0.00)	(0.01)	(0.00)
Preferred Candidate	2.30***	0.83***	0.73***	1.25***
(winner=1)	(0.35)	(0.15)	(0.23)	(0.35)
Observations	2050	2050	600	550
Wald Chi(2)	253.60	253.60		
Pseudo R2	0.1955	0.1955	0.3272	0.1908

Notes:
Fixed effects include age, gender, ethnicity, and major; Clustered Robust SE at the Subject level
Significance at the 10% (*), 5% (**), and 1% (***) level

Table 4: Logistic Coefficient Estimates

1.5.2 Hypothesis 2

Hypothesis 2 states that the evaluative system can result in a greater expected payoff which should result in a willingness to participate at higher private individual costs. If we compare the mean individual costs in rounds where subjects chose to participate across voting systems then we should



System	Obs.	Mean	Standard Error
Evaluative	477	39.09	1.21
Plurality	444	38.77	1.28

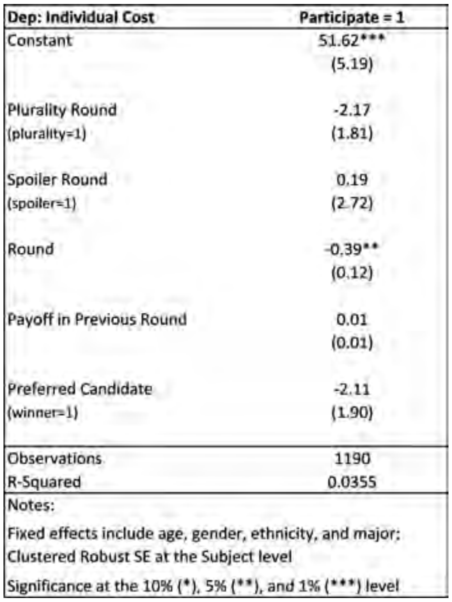
Table 5: Mean Individual Cost in Participation Rounds

see a significantly higher individual cost in the evaluative rounds. In their questionnaire responses, several subjects mentioned that the threshold for their participation decision was higher in the evaluative system, but it was not a large enough change or a common enough difference to be reflected in the overall means.

Table 5 shows the means and standard errors for these two groups. Although the evaluative mean is higher, a simple t-test of means confirms that the difference is not significantly different from zero. Despite the clear result in Table 5, a linear regression was appropriate to attempt to control for other factors in this relationship. Table 6 outlines the results of the linear regression conducted on rounds where subjects chose to participate. The regression clearly reinforces the findings of the comparison of means in the fact that the sign on the plurality round is negative, but not significant. The only significant variable in these regressions was the variable that indicated what round the choice was occuring in. The negative and significant value of this coefficient supports the fact that subjects became more hestitant to participate in later rounds and therefore their average individual cost in later rounds was lower. Hypothesis 2 is not supported by these results.

1.5.3 Hypothesis 3

One of the greatest advantages to the evaluative voting system is that it allows the voter to place positive influence on their 2nd best choice as well as their preferred choice. In light of this advantage, the existence of a spoiler candidate which might force voters to behave more strategically and select a lesser candidate or split the vote across two largely preferred candidates would seem to increase the value to the subject of using an evaluative voting system. This increased value should theoretically increase the participation rate of the evaluative system relative to the plurality system. Hypothesis 3 above attempts to capture this effect.



Dep: Individual Cost	Participate = 1
Constant	51.62***
	(5.19)
Plurality Round	-2.17
(plurality=1)	(1.81)
Spoiler Round	0.19
(spoiler=1)	(2.72)
Round	-0.39**
	(0.12)
Payoff in Previous Round	0.01
	(0.01)
Preferred Candidate	-2.11
(winner=1)	(1.90)
Observations	1190
R-Squared	0.0355
Notes:	
Fixed effects include age, gender, ethnicity, and major; Clustered Robust SE at the Subject level	
Significance at the 10% (*), 5% (**), and 1% (***) level	

Table 6: Linear Coefficient Estimates

Looking back to Table 4 above, we see that the likelihood of participation is reduced in the spoiler rounds, but when the regression is run on only the spoiler rounds we see a dramatic and statistically significant increase in the participation rate in the evaluative rounds. The size of the evaluative effect in the spoiler treatment is nearly five times as great as the effect across all treatments. The findings from this experiment mirror two important observations regarding the existence of spoiler candidates. The first being that voters experience frustration with their inability to coordinate around a single preferred candidate as demonstrated by the more dramatic decline in participation in later rounds of the spoiler treatment. The second is that the majority of voters would prefer a means by which they could ensure that the least preferred candidate is not victorious as a result of a failure to coordinate through strategic voting practices. This can be seen in the experimental data as the increased willingness to participate is much greater in the evaluative system where subjects are able to reduce the likelihood of the least preferred candidate wining through negative votes. The evaluative system also facilitates increased coordination around the majority preferred candidates by allowing voters to cast positive votes for multiple candidates. This result brings the discussion back to one of the most dominant arguments used in support of election reform in the United States, under normal conditions an alternative voting system may marginally increase voter participation even if it does not fundamentally change the outcome of the election. Where these alternative voting systems really offer a significant improvment over the existing system is in the more complicated races which contain spoiler candidates and result in both increased voter frustration and less preferred candidates being elected.

1.5.4 Qualitative Questionnaire Responses

A short questionnaire was conducted after the experiment concluded to ascertain the thoughts and feelings of the subjects after experiencing both voting systems. The subjects were asked a series of common questions including their demographic characteristics as well as their impression of the experiment itself. An examination of the subject's responses reveals a few key insights about their experiences as well as the influence of the voting systems themselves.

Key Lesson 1: The subjects frequently adopted cutpoint strategies when comparing their individual costs to the expected utility of participating. When asked whether they used/adopted a strategy in their approach to the experiment, 70% of subjects indicated that they did. When given the opportunity to share their strategy, 28% of subjects explicitly mentioned their cutpoint strategy. The following two responses demonstrate the general feedback from subjects. "If the cost was higher than 50 then I decided not to participate because the deducation from the total pay would have been less than if I had not participated." and "If the cost was less than 60 then I paid the price and added a token to Bucket 1 in Version 1. In Version 2, I added a token to Bucket 1 and removed tokens from Bucket 2and 3. In Version 1, the efficiency was mediocre but it really paid off in Version 2." both demonstrate the cutpoint strategy. The most commonly mentioned cost threshold was 50 points which respresented 50% of the maximum possible cost of 100 points. Interestingly, this suggests that the threshold may have been defined based on the distribution of costs and not necessarily on the expected return from voting.

Key Lesson 2: The subjects clearly perceived and acknowledged the differences between the plurality and evaluative treatments. The second comment shared above demonstrates this acknowledgement. Another subject commented, “I was green, so I always put in bucket 1 in the first half of the experiment. Took out of bucket 2, then abstained from bucket 3. In the 2nd half of the experiment, I always went in on bucket 1 unless my cost was over 60.” The feedback made it clear that subjects were recognizing the different voting environments, the changes in the other players strategies, and adapting their own strategy accordingly. They also clearly indicated that the introduction of a spoiler candidate was obvious and had a significant impact on their voting strategy as well as their appreciation for the more complete ballot expression in the evaluative system. One subject clearly explained the spoiler dilemma, “In variation 2, if every member had perfect information, red and blue team are set up to win over green. In variation 1, with perfect information, red and blue team are at a conflict of interest and the odds of unity against green team are impossible without thorough communication and understanding.”

Key Lesson 3: The feedback from the subjects closely aligned with the literature on alternative voting systems. Several subjects mentioned that the plurality system was simpler or easier to understand while others mentioned that the evaluative system gave them more control over their expression and by extension more control over the outcome of the round. A few subjects mentioned that the complexity of the evaluative voting system made the outcome more interesting and unpredictable. Several subjects explicitly mentioned the ability to remove tokens as an exhancement of the process and providing them the feeling of having more control over the ultimate outcome of that round. Subjects commented, “I was able to influence the other two buckets and was able to remove tokens from them so my bucket of preference had more of a likelihood of winning.” and “It [version 1] was easier and less complicated than version 2.” and “I felt like I had more control over what the outcome would be as there were choices, whereas version 1 felt very limiting.”

The results of the questionnaire provide even more support for the hypotheses above as well as credibility for the experimental design. The many similarities between the subject’s feedback and the theorized impacts of an alternative voting system of this type demonstrate the importance of studying these systems and devleoping as complete an understanding of their impact on voter behavior and satisfaction as possible. As mentioned above, many alternatives to plurality voting exist and each represents a unique combination of benefits and costs to voters.

1.6 Conclusion

Voting systems and the electoral process inevitably garner more attention immediately preceding or following a highly contested election. One of the key messages from the existing voting literature can be simplified to a realization that the existing electoral system in the United States is inherently flawed, but the alternative systems also contain their own unique problems, in many cases the devil you know is preferred to the devil you do not know. One of the key components of a representative democracy is voter participation and the engagement of all citizens with the electoral process. The literature above suggest that voter participation would be greater under an evaluative voting system relative to the existing plurality voting system. The results from this laboratory experiment support this hypothesis. The unique contribution of this paper is the fact that these findings were derived from a laboratory experiment where the social and political influences could be removed and the focus could be narrowed simply to the economic influences on a subject's decision to participate or abstain.

Although the findings of this paper align with the existing literature on voter participation, this paper arrives at these findings through a uniquely controlled environment which removes some of the possible confounding factors present in field experiments intended to test the same hypothesis. This paper also demonstrates the increased value of an alternative voting system, such as the evaluative system, when addressing elections which contain spoiler candidates. The evaluative system results is far greater participation among subjects in the presence of a spoiler candidate as well as additional voter satisfaction as evidenced by the responses to the qualititative questionnaire conducted after the experiment.

This paper supports the existing literature and the findings of field experiments, but it leaves some questions unanswered and ripe for further investigation. The evaluative system is one of many alternative voting systems and therefore the findings of this paper could naturally be expanded to include other alternative systems such as approval voting, Borda counts, ranked choice voting, and evaluative voting with different scoring alternatives. The scope of this paper is also relatively small so value would be added by conducting this experiment on larger groups which would reduce the pivotality of each individual voter and might result in different behaviors. There is a great deal to be explored, but this paper and the experiment contained within begin to look at these questions from a unique and controlled environment that could continue to bring new insights to this discussion.

2 An Experimental Study of Dark Network Formation and Disruption

2.1 Introduction

Scholars have long recognized that a group's desire to conceal its activities leads it to adopt a decentralized organizational structure (Simmel 1906; Erickson 1980). Such is particularly true of covert and illegal groups called dark networks (Raab and Milward 2003; Everton 2012; Gerdes, 2015) that include terrorist organizations, drug cartels, arms and human traffickers, and criminal gangs. Because these groups face the constant threat of infiltration and disruption by law enforcement and the military, a decentralized structure is believed to improve the group's security by reducing the impact of a disruption. In effect, the group trades off the productivity advantages of centralization for the security advantages of decentralization. Studies of dark network organizations increasingly support this view. For example, Al Qaeda's organizational structure was somewhat vertical before September 11, 2001, but shifted to a decentralized structure after the American invasion of Afghanistan (Raab and Milward 2003), and Hamas moved from a centralized to a compartmentalized structure after being outlawed by the Israeli government (Gambil 2002).[4]

---

[4] For other examples, see Baker and Faulkner (1993), Arquilla and Ronfeldt (2001), Bakker, Raab, and Milward (2012), Krebs (2002), Milward and Raab (2006), and Shapiro (2013).

Despite this compelling evidence, it is difficult to prove a definitive causal link between the threat of disruption and the organizational structure for many reasons. Because these groups value secrecy, accurate information about the organizational structure of a dark network is virtually impossible to obtain (Krebs 2002). Those parts of the organization that are observed may be systematically different from the unobserved parts in unknown ways, thus raising questions about the viability of making accurate inferences from the available data. Another challenge is that changes in an organization's structure that coincide with a shift in law enforcement or military strategy may actually be caused by other unobserved factors. For example, a change in an organization's goals or tactics, such as an increased use of violence, may cause both a shift in organizational structure and heightened scrutiny from the government. An example of this is when Al Qaeda in Iraq transformed itself into ISIS. In this case, the transformation included a shift from a decentralized network to a hierarchical organization as ISIS proclaimed itself to be a caliphate that controlled territory and people and even claimed to have a capital city. Separating out the effect of the change in strategy from the change in government policy may not be possible since they are linked. Because organizations are more powerful than networks, if the government weakens, the dark network can become an organization as ISIS did. Finally, there are obvious practical and ethical reasons that prevent a systematic field experiment that randomly assigns both organizational goals and law enforcement strategy.

This paper presents empirical evidence from a controlled, laboratory experiment on dark network formation and disruption that avoids the difficulties inherent in field studies. In each experimental trial, one subject takes the role of network designer and is paired with another subject in the role of network disruptor. The designer, who acts as leader of a dark network, has 18 members of their group and must decide how to allocate 18 collaborative links between group members. These collaborative links are tied directly to group production; the more links, the more production, and in some treatment conditions, the group production benefits from having a more centralized structure. After the network designer constructs their network, the network disruptor is given information about the network and decides which member (node) to remove from the organization.

Removing that member involves the complete removal of that person and all their links from the network's production. The more output produced by the final, post-disruption network, the larger the designer's payoff, and the smaller the disruptor's payoff. The order of moves and payoff structure is known to both actors. A key design feature of our experiment is that we exogenously vary both the incentive to centralize and the disruptor's ability to monitor, thus allowing us to identify the causal effect of the threat of disruption on organizational structure.

Our main result is that designers of dark networks do intentionally decentralize in the face of strong interference by law enforcement: holding productive capacity fixed, designers construct more decentralized networks as the disruptor's monitoring of the network increases. We also find that the trade-off between maximizing productive output and maximizing security is a salient one: holding the monitoring ability fixed, designer's networks are more centralized as the production advantage to centralization increases. The evidence further indicates a novel finding particular to our experiment, namely, that the order in which subjects are exposed to the monitoring levels has an effect on the extent of centralization. Designers construct networks that are more centralized after first being exposed to monitoring and then having that monitoring removed, however, they still construct networks more decentralized than optimal. The implication is that subjects are learning about the optimal network design, but are still deterred from full centralization even when constructing a fully centralized network is optimal and all monitoring is absent. Overall, our experimental results provide compelling evidence that the structure of a dark network reflects the twin organizational concerns of productivity and security.

Our paper is closely related to two strands of existing literature. The first is a theoretical literature on the strategic interplay between members of a dark network and a network disruptor. See, for example, Baccara and Bar-Isaac (2008), Enders and Jindapon (2009), Hoyer and De Jaegher (2010), Hong (2011), and Goyal and Vigier (2014). The game-theoretic decision setting for our experiment is most similar to that of Enders and Su (2007), the key difference being the network production function. The production function we use was specifically selected with our experiment in mind as it provides a concrete way to vary an incentive to centralize that is easily understood by experimental subjects.[5] The second is the experimental literature on dark networks reviewed by Arce et al (2011).[6] We here specifically mention McBride and Hewitt (2013), McBride and Caldara (2013), and Candelo et al (2013). The first two of these studies focus on the disruptor's decision when shown a randomly selected dark network, while the latter considers the formation of a dark network among decentralized actors in the face of a random (not strategic) disruption. The empirical literature on dark networks, summarized in Bakker, Raab, and Milward (2012), suffers from the same problem as the experimental literature, it is essentially one sided in that it looks at dark networks as they respond to shocks and attacks.[7]

This paper makes a unique contribution to the experimental and empirical literatures as it is a two-sided model while also using a simpler design in which a single actor (the designer) chooses the network structure. This simpler design avoids the coordination problem inherent in decentralized network formation, thus allowing for sharper findings on changes to the network structure.

2.2 The Designer-Disruptor Setting

2.2.1 The Game

The Designer-Disruptor Game is a two-player game in which one player constructs a network and the other chooses a single node to remove from the network. The game proceeds in two stages.

In the first stage, the first mover, called the *designer*, allocates 18 collaborative links among 18 nodes. The nodes represent members of the dark network, and the links represent the collaborative

---

[5] There are also various other related directions. In Colonel Blotto games, an attacker and defender fight over multiple battlefields; these settings are akin to network-based conflict albeit with simple network architectures (see Roberson 2006). Arce et al (2012) consider conflict between a terrorist with multiple attack technologies and a defending government actor. Franke and Öztürk (2015) study how network structure determines conflict intensity. Still other approaches have been used, such as computational models (e.g., Albert et al 2000; Callaway et al 2000; Cohen et al 2001) or order theory (Farley 2010). Feinstein and Kaplan (2010) work to bridge the gap between the computational and game-theoretic approaches. Arin et al (2011) examine spending on terrorism and counter-terrorism without examining network structure. Still other work examines the key player problem of identifying which set of nodes to remove to have the largest impact on the networks (Borgatti 2003; Ballester et al 2006).

[6] Many of the related experiments study Colonel Blotto games, e.g., Chowdhury et al (2013). See Dechenaux et al (forthcoming) for a review of the experimental literature on contests, all-pay auctions, and tournaments. For an earlier survey of network-related experiments, see Kosfeld (2006).

[7] Again, there is extensive literature that reflects a diversity of methodological approaches. For example, Everton (2012) presents a comprehensive social network analytic approach, and Miffen et al (2004) use random graphs to infer underlying terrorist networks.

efforts between dark network members. For simplicity, each node is numbered $i = 1, 2, ..., 18$. Each link can be used to create a tie between two different nodes. These links can be allocated between any pair of nodes, and not all links need to be used. Let g denote a network, where $g_{ij} = 1$ denotes a link between nodes i and j in network g and $g_{ij} = 0$ denotes that there is no link between i and j in g. Assume $g_{ii} = 1$. The set of possible networks from which the designer can choose is the set of all g such that

$$\sum_{ij, i \neq j} g_{ij} \leq 36.$$

The 36 captures the constraint that the designer uses at most 18 links.

Once the designer has selected g, the game enters the second stage. In this stage, z of the 18 nodes are randomly selected and the disruptor observers the identities and direct ties of those z nodes (and of these alone). As discussed in detail below, we will consider three observation levels: $z = 0, 7,$and 18. After this observation, the disruptor selects exactly one node to remove (i.e. arrest) from the network. That node and all of that node's direct links are removed, with the rest of the network remaining intact. Let $g - i$ denote the network that results from the designer's choice of network g in stage one and the disruptor's choice to remove node i. Network $g - i$ is thus exactly equal to g except that all direct ties that involve i are removed. After the disruptor's choice is made, both actors receive their payoffs, and the game ends.

Network $g - i$ is appropriately called the final network because this is the network from which the players derive utility. Specifically, each node produces

$$p_i(g - i) = a + b_1 \# L_i(1, g - i) + b_2 \# L_i(2, g - i),$$

where $a, b_1,$and b_2 are non-negative constants with $b_1 > b_2, \# L_i(1, g - i)$ is the number of nodes of geodesic distance 1 from i in $g - i$, and $\# L_i(2, g - i)$ is the number of nodes of geodesic distance 2 from i in $g - i$. In words, each node produces a alone, produces an additional b_1 for each direct tie, and produces an additional b_2 for each second-order tie (and produces zero for ties of order three or higher). The designer acts to maximize the sum of the individual nodes' production:

$$P(g-i)=\sum_{i=1}^{18}p_i(g-i).$$

The disruptor acts to maximize their payoff function:

$$X-P(g-i),$$

where X is a constant. In other words, the disruptor acts to minimize $P(g-i)$.

We will consider three production settings that differ in the incentives to centralize. Under the "None" condition $(a=15,b_1=10,b_2=0)$,there is no incentive to centralize because $b_2=0$ implies that there are no returns for second-order links. Any network that uses all 18 links produces the same total pre-removal production. As b_2 increases (but remaining less than b_1), there is an increasing productive gain by having a highly centralized node. Under the "Weak" condition $(a=15,b_1=15,b_2=5)$, there is an incentive to centralize, but the second-order links provide a relatively small benefit. Under the "Strong" condition $(a=15,b_1=30,b_2=20)$, the benefits from second-order links are much higher.

The designer and disruptor both commonly know this set-up. The only informational difference between the two is that the disruptor will sometimes (i.e. when $z<18$) not know the full initial network g when making her disruption choice.

We now discuss the designer's incentive to centralize under various observation levels and production parameters. Note that we have not obtained a complete characterization of the set of all sub-game perfect equilibria because of the exceedingly large number of network configurations (even non-isomorphic ones) and observational realizations that are possible and must be checked. Our discussion is thus restricted to the key logic and intuition behind the motivating conjecture that optimal centralization should decrease as observation increases, but should increase as the productivity gains from centralization increase.

2.2.2 Hidden Observation

When the disruptor observes none $(z = 0)$ of the nodes and links, then she is indifferent between any selection and essentially chooses a node at random. This implies that the designer's optimal strategy depends on the value of the constants a, b_1, b_2. When first-order links are the only source of production, i.e. $b_2 = 0$, any configuration of the 18 links has the same expected production ex ante since there is a fixed number of first order links that can be created and each node has the same probability of being removed. With $b_2 > 0$, then there is a compounding effect of connecting a single node to multiple other nodes. Although the maximum number of first-order connections is 36, there are only 18 links to allocate, each of those first-order connections can become as many as 16 second-order connections if a connected star configuration is used.

For example, if node 2 is centrally connected to all other nodes then the connection between node 1 and node 2 would result in b_1 additional units of production from node 1 and node 2. It also results in b_2 additional units of production from each of the other 16 nodes connected to node 2, who are now connected to node 1 as well. This compounding effect implies that the network with the greatest amount of centralization will result in the greatest expected production assuming that the disruptor cannot accurately detect the hub node. Therefore, the optimal configuration should be a connected star with one additional, arbitrary-placed link. This network is highly centralized.

2.2.3 Partial Observation

When the disruptor observes only a part of the network, the removal choice depends on the expected total impact of each node. We define total impact as the difference in total production between the original network, g, and the network absent that node, $g - i$. It is important to note that the total impact of the removal of node i can be greater than just the production directly associated with that node. If $b_2 > 0$, then node i's first-order links can act as a pathway to connect other nodes at the second-order. In that case, there is an indirect production impact, which we will denote as $\#\tilde{L}(1, g)$, that represents the number of second-order links maintained by the other 17 nodes solely through a first-order link that belongs to node i. The exact value of $\#\tilde{L}(1, g)$ depends on the distribution of redundant links found within g, and takes on a value between 0 and $[\#L_i(1,g)^2 - \#L_i(1,g)]$, where 0 equates to a perfectly redundant network and $[\#L_i(1,g)^2 - \#L_i(1,g)]$ equates to absolutely no redundancy.

When $b_2 = 0$, then the optimal removal is either the observed node with the highest degree centrality or the unobserved node with the highest expected degree centrality, as shown by McBride and Hewitt (2013). When $b_2 > 0$,then additional calculations are required. The direct impact of the removal of node i is equal to $a + 2b_1\#L_i(1,g) + 2b_2\#L_i(2,g)$ because the production is lost to both nodes involved in the link. The indirect impact of the removal is equal to $b_2\#\tilde{L_i}(1,g)$. Therefore the total impact of the removal of node i is equal to

$$TI_i = a + 2b_1\#L_i(1,g) + 2b_2\#L_i(2,g) + b_2\#\tilde{L_i}(1,g).$$

As this expression indicates, the node with the largest number of first-order links will be a candidate to consider for removal, yet a node with more second-order links than first-order links may instead be the best node to remove.

The designer's optimal strategy in this environment also depends on the production parameters. If $b_2 = 0$, then the designer simply wants to minimize the number of first-order links lost because there is no productive advantage to creating a more centralized network. This is accomplished by utilizing any regular network with degree two, because any sort of clustering will, with high probability, be revealed through the partial observation of the network. If $b_2 > 0$, then there is an incentive to create clusters in order to maximize the number of second-order links in the network. This results in a typical trade-off between clustering for increased production and decentralizing for increased security. When b_2 is small relative to b_1, then the set of optimal networks contains regular networks with degree two to maximize security, but as second-order production becomes relatively more valuable the optimal strategy becomes a small number of clusters. The loss upon detection of one cluster is outweighed by the increased production of the remaining ones.

We chose the values for b_1 and b_2 in the weak and strong incentive structures to emphasize the friction between increased production and increased security. The large number of possible equilibrium structures prevents us from obtaining a formal characterization of equilibrium structures, yet we make a few observations. Under the “strong” incentive, $(b_1 = 30, b_2 = 20)$, an effective design has three equal clusters of six nodes configured as one hub and five stems. This allows the designer to achieve high production from centralization while creating enough clusters to minimize the total impact of the removal of one of the hub nodes. Under the “weak” incentive, $(b_1 = 15, b_2 = 5)$, the coefficients were chosen so that the designer faces an approximately equal expected payoff between a regular network of degree two and the clustering scenario described above. This results in the decision to centralize or decentralize being entirely one of preference.

2.2.4 Full Observation

When the disruptor observes the entire network, then the optimal removal choice becomes the maximization of the total impact function shown above regardless of the production function. Despite the large number of possible network configurations that comprise the set of subgames, we can still make some general statements about possible equilibria. One observation is that the best node to remove should again typically have high degree centrality. Moreover, the designer’s optimal strategy in this environment is similar to that of the partial observation. The difference is that under partial observation, sub-optimal strategies such as two-cluster and four-cluster networks still offer a relatively high expected payoff due to the increased uncertainty of the true nature of the network. Under full observation there no longer remains any temptation to use either of these strategies as they are strictly dominated by a significant amount by the optimal strategies documented below.

When only first-order production occurs $(b_2 = 0)$, the regular network of degree two is optimal because it still minimizes the expected loss of links. When second-order production also occurs $(b_2 > 0)$, the set of optimal networks depends on the production incentives. Under the strong incentive, the three equal cluster design remains viable as it is able to maximize the gains from centralization while minimizing the total impact of the removal of a hub node. Under the weak incentive, the expected production is again approximately equal between a regular network of degree two and the three equal cluster design.

Summarizing this discussion, centralization should be much greater under Hidden observation than under Partial and Full observation, and higher under Weak and Strong production incentives than under the “None” production incentive. There may be differences between Partial and Full observation and between Weak and Strong production incentives, though these are likely to be smaller. We will state the predictions in our experiment more explicitly below.

2.3 Experiment Design

2.3.1 Basics

We conducted our experimental sessions at the Experimental Social Science Laboratory (ESSL) located at the University of California, Irvine. The subject pool consists of both undergraduate and graduate students recruited using classroom advertisements and an online registration portal. Subjects in the subject pool receive an email invitation to sign-up for experiment sessions. Interested subjects then respond by clicking an electronic ticket thus reserving a spot in the session. The confirmed respondents received a reminder e-mail the day before the experiment. Subjects were only allowed to participate in at most one session. For experiment management and data collection we employed the Multistage software package.[8]

Subjects arrived at the lab at the allotted time, signed in, and then were randomly assigned to computers in the lab. At the start of each experiment session, the subjects received instructions about the decision scenario and then began the decision-making phase. After the decision-making phase, each subject completed a short questionnaire that asked for the subject to report their sex, disciplinary major, and other characteristics. Each subject received a fixed $7 show-up payment for participating, with additional salient earnings based on the decisions made and experiment session payoff parameters (described below). The average total take-home payment was approximately $26 for about 75 minutes of participation. Table 7 provides basic information about the sessions and subjects.

---

[8] Information about the Multistage software platform can be found at http://multistage.ssel.caltech.edu:8000/multistage/.

Session	Order-Incentive	Subjects	Percent Male[a]	Percent with 1+ Stats course[a]	Percent with 1+ Econ course[a]	Percent of Subjects by Major[a]					Avg. Payoff[b]
						Business or Econ.	Social Science	Phys. Science or Engine.	Bio., Life Science, or Pub. Health	Other	
1	Increasing-None	48	38%	50%	42%	13%	33%	21%	29%	4%	\$25.76
3	Increasing-Weak	40	33%	60%	60%	23%	13%	33%	30%	3%	\$25.46
5	Increasing-Strong	44	45%	61%	39%	7%	32%	27%	30%	5%	\$25.67
2	Decreasing-None	44	39%	55%	48%	20%	16%	25%	36%	2%	\$26.15
4	Decreasing-Weak	54	50%	59%	50%	26%	19%	26%	19%	11%	\$25.67
6	Decreasing-Strong	40	43%	68%	50%	30%	23%	20%	13%	15%	\$26.16

[a] Information collected from questionnaire. [b] Includes \$7 show-up payment.

Table 7: Session Information

2.3.2 A Single Round

Each round consists of a single, complete designer-disruptor setting: the designer constructs a network, the computer selects z nodes randomly, the disruptor is then shown the direct links of those nodes, the disruptor makes the arrest decision, and then the outcomes of this interaction are revealed to both the designer and disruptor.

In the first round of each treatment condition, the designer starts with an empty network and has 120 seconds to assign links. In each subsequent round, the designer starts with the network constructed in the prior round and has 30 seconds to make changes to it. Figure 4(a) contains a screenshot of what the designer sees during this network construction stage. The ties are created in the matrix on the left side by changing cell entries from 0 to 1 to create a link or from 1 to 0 to remove a link. The diagram on the right side responds in real-time to show the designer a spatial representation of the network being constructed in the left-side table. Also shown to the designer is the time remaining for construction and the unused links (collaborations) that can be assigned. During this stage, the disruptor's screen instructs the disruptor to wait while the designer completes network construction.

Upon the completion of the network or the expiration of the allotted time, the network is then submitted to the computer to start the network disruption stage. The computer randomly selects z nodes according to the assigned treatment (0, 7, or 18). All nodes are shown to the disruptor on the disruptor's screen, with the z nodes colored green to signify that they were monitored and the

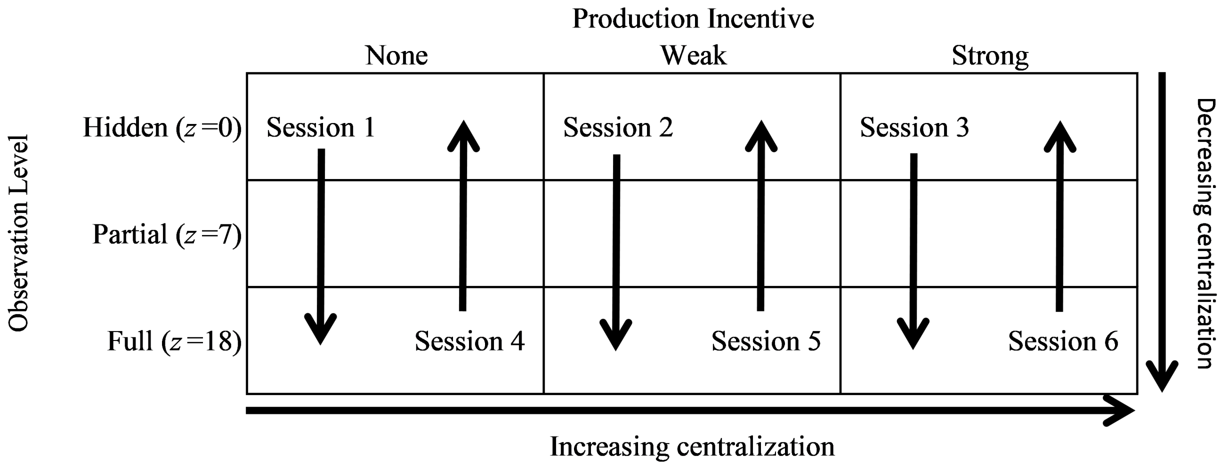


Table 8: Summary of Predictions

other nodes colored blue. All of the direct ties of the monitored nodes are also shown. The nodes are aligned in a circle, and the order of the nodes along that circle is randomly permuted each round. See Figure 4(b). The disruptor has no explicit time limit in which to make her decision. The disruptor then makes an arrest by clicking with the mouse on the desired node to remove.

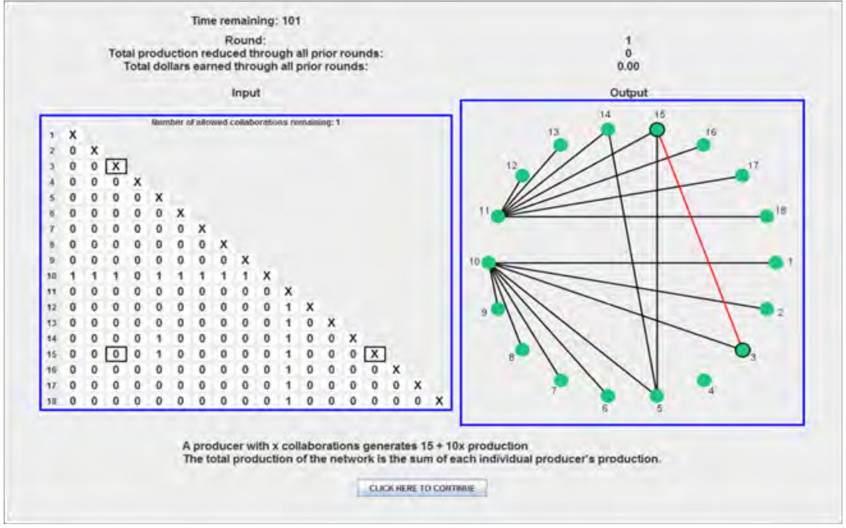
After the disruptor's arrest decision, both the designer and disruptor are shown the results screen. As seen in Figure 4(c), this screen reveals the entire network to both subjects and highlights the arrested node and its links. The disruptor is still shown the shuffled version of the network as before. The total pre-removal production, total post-removal production, production prevented, and percent production reduced are also reported to both subjects. After a short time to review the outcomes, the experiment advances to the next round.

2.3.3 Sessions

With three observation levels and three production incentives, we have the nine different treatment conditions as shown in Table 8**.** We implement these different conditions using a hybrid, 3x3 factorial design, in twelve different experiment sessions. Each session consisted of a single production incentive (None, Weak, or Strong) with 10 rounds of each observation level given to the subjects in either increasing or decreasing order.

In the "Increasing-None" Session 1, the subjects were randomly paired to interact for 30 total

(a) Designer Network Construction Screen



Time remaining: 101

Round: 1
Total production reduced through all prior rounds: 0
Total dollars earned through all prior rounds: 0.00

Input

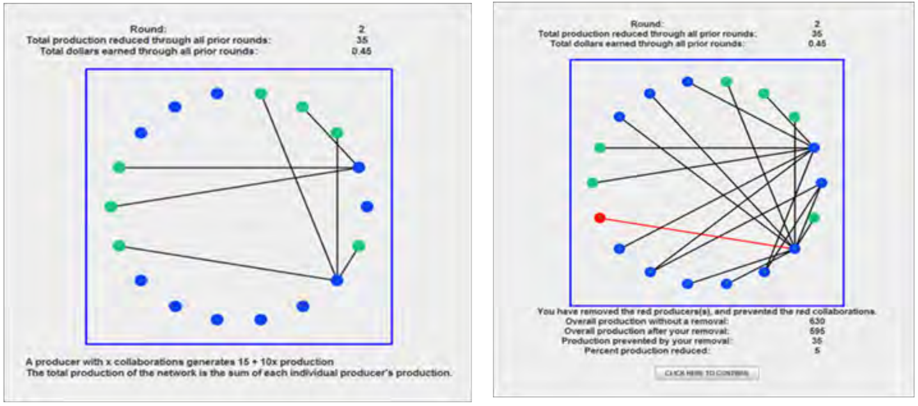
Number of allowed collaborations remaining: 1

1	X																	
2	0	X																
3	0	0	X															
4	0	0	0	X														
5	0	0	0	0	X													
6	0	0	0	0	0	X												
7	0	0	0	0	0	0	X											
8	0	0	0	0	0	0	0	X										
9	0	0	0	0	0	0	0	0	X									
10	1	1	1	0	1	1	1	1	1	X								
11	0	0	0	0	0	0	0	0	0	0	X							
12	0	0	0	0	0	0	0	0	0	0	1	X						
13	0	0	0	0	0	0	0	0	0	0	1	0	X					
14	0	0	0	0	1	0	0	0	0	0	1	0	0	X				
15	0	0	0	0	1	0	0	0	0	0	1	0	0	0	X			
16	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	X		
17	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	X	
18	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	X

A producer with x collaborations generates 15 + 10x production
The total production of the network is the sum of each individual producer's production.

(b) Disruptor Arrest Screen with Partial Observation (z=7)

(c) Results Screen



Round: 2
Total production reduced through all prior rounds: 35
Total dollars earned through all prior rounds: 0.45

A producer with x collaborations generates 15 + 10x production
The total production of the network is the sum of each individual producer's production.

Round: 2
Total production reduced through all prior rounds: 35
Total dollars earned through all prior rounds: 0.45

You have removed the red producers(s), and prevented the red collaborations.
Overall production without a removal: 630
Overall production after your removal: 595
Production prevented by your removal: 35
Percent production reduced: 5

Figure 4: Experiment Screenshots

rounds with the same partner. The first 10 rounds were of the Hidden observation ($z = 0$), then for 10 rounds of the Partial observation, and then for 10 rounds of the Full observation, again with the same partner for all 30 rounds. The "Increasing-Weak" and "Increasing-Strong" sessions were identical except for the production incentive. A "Decreasing-None" session was similar to the "Increasing-Weak" session except the subjects did 10 rounds of Full observation, then 10 rounds of Partial observation, then 10 rounds of Hidden observation. Two sessions were conducted for each Order-Incentive, i.e., two sessions of "Increasing-None," two sessions of "Increasing-Weak," and so on, for a total of twelve experiment sessions. Data from identical sessions are pooled in the analysis below, though we will test for differences between increasing and decreasing sessions.

2.3.4 Hypotheses

We identify five hypotheses. The first two concern basic decision-making acumen by the subjects to determine whether the subjects took their roles seriously in attempting to maximize payoffs.

Hypothesis 1. *The designer uses all 18 links.*

Each additional link can only improve network production, irrespective of the disruption strategy. Therefore, it is always optimal to use all links. We test this hypothesis to verify whether or not the subjects in the designer role took the decision-making seriously.

As discussed earlier, the optimal removal by the disruptor typically corresponds to removing the node with the highest degree, but the ability to do so is constrained by the observation level. Our second hypothesis follows.

Hypothesis 2. *Under partial and full observation, the disruptors are more likely to remove a node with a high degree.*

The next two hypotheses follow from our discussion in Section 2.2. As the disruptor is shown more of the network, the designer realizes that the network is more susceptible to precision removal. This motivates the designer to create a less-centralized network which will better absorb the loss of a single node.

Hypothesis 3. *Holding production fixed, centralization is decreasing in observation, i.e., networks will be more centralized under Hidden observation than under Partial observation, and more centralized under Partial observation than under Full observation.*

As the centralization incentive increases the production of second-order links, the designer has a stronger incentive to create more centralized networks. Second-order production is maximized by creating a single star-shaped component, so we are able to create a significant friction between the benefits of centralization and the desire to maximize security. At a given level of observation, we expect that a greater production incentive will result in a more centralized network.

Hypothesis 4. *Holding observation fixed, centralization is increasing in the production incentive, i.e., networks will be more centralized under the Strong incentive than under the Weak incentive, and more centralized under the Weak incentive than under the None incentive.*

Finally, decentralization should, in principle, depend only on production parameters and observation level. Of course, additional factors may be present that create an ordering effect, a possibility we test with our fifth hypothesis.

Hypothesis 5. *There should be no effect of ordering on the level of decentralization.*

2.4 Results

Result 1: *Essentially all designers use all 18 links, except in early rounds.*

The median number of links used in each round is 18, thus implying that most subjects do indeed use all links. However, some subjects do not use all links; the mean links used for the first five rounds climbs each round from 13 to 17. There are two possible explanations for these patterns. The first is that the subjects are still learning and familiarizing themselves with the game structure. The second is that subjects are making a large number of choices in a short period of time and are not able to implement their entire strategy before time runs out. By round six the mean number of links used is greater than 17, and this holds true for the remaining 24 rounds as well. These findings are strong evidence in support of Hypothesis 1.

Result 2: *Designers display a strong tendency to remove a node with the most observed links.*

Figure 5 shows the distribution of removals by treatment. The removal is essentially random with no observation ($z = 0$) as expected. With partial observation, disruptors remove the max monitored or the max non-monitored over 75% of the time, and with full observation the max degree node is removed over 78% of the time. The removals are not perfect, even in full observation, but some of this variation can be reasonably attributed to experimental behavior or a failure of understanding on the disruptor's part. However, despite this deviation from optimal removals, most choices are consistent with Hypothesis 2. These first two initial results function as a check on consistency, subject understanding, and subject motivation. On the whole, it appears that most

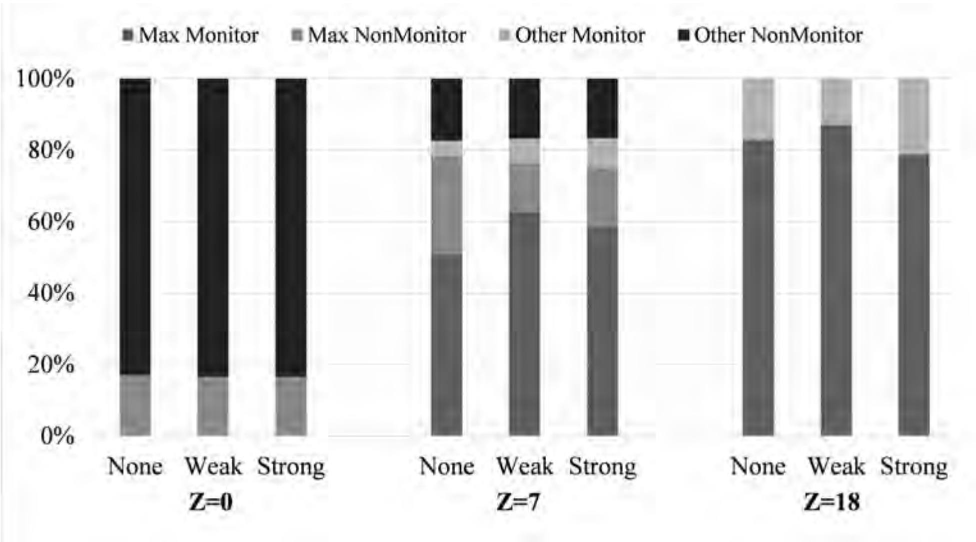


Figure 5: Distribution of Removals by Treatment

subjects take the experimental task seriously and act to achieve high payoffs albeit with some error.

The rest of our results examine the structural properties of designed networks. The structural characteristic of interest is the degree of centralization, and our preferred metric of centralization is based on the commonly used and referenced measure of Linton Freeman (Freeman, 1979). Group centrality is measured by

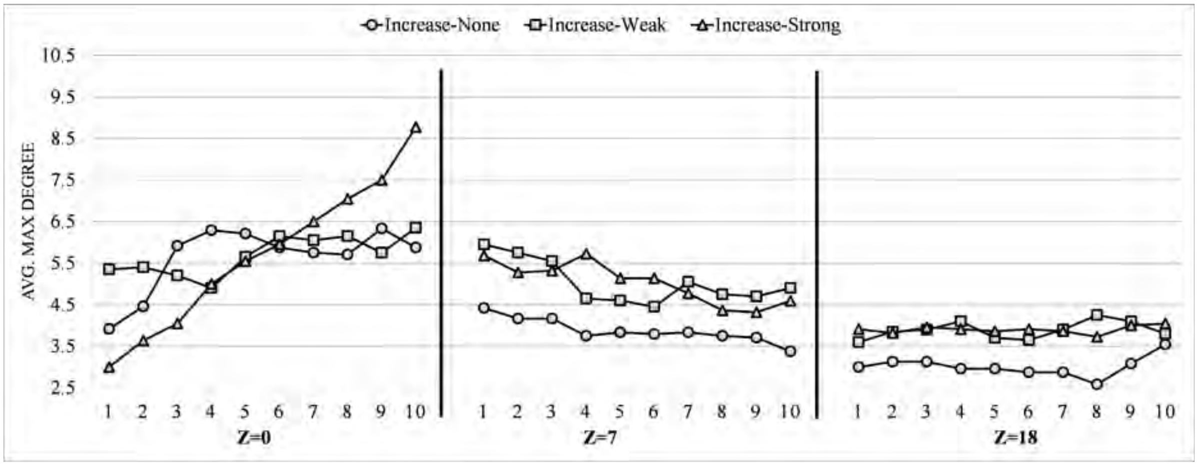
$$C_D = \frac{\sum_{i=1}^{n}[C_D(\dot{p}) - C_D(p_i)]}{max \sum_{i=1}^{n}[C_D(\dot{p}) - C_D(p_i)]}$$

where $C_D(p_i)$ is the degree centrality of node i and $C_D(\dot{p})$ is the degree centrality of the maximum degree node. This metric is essentially measuring the distance between the max degree node and all other nodes and then standardizing this measure to fall between 0 and 1. On this scale we can think of 0 equating to a completely decentralized structure such as a connected ring while 1 equates to a completely centralized structure such as a star network.

This measure of centrality is useful for many reasons: it captures the idea of a hub or highly central actor, it also captures the extent of centralization around that hub, and as Freeman, Roeder, and Mulholland (1980) find, his measure closely reflects the efficiency of a network to coordinate and solve problems. Both of these features are directly relevant to the operation and coordination of dark networks that we hope to shed light upon with this research.

Figure 6 charts the average group centrality in each round for each treatment. We observe some evidence of convergence in this measure as well as the patterns of behavior outlined in Hypotheses 3-5. These figures provide visual confirmation of our hypotheses, but we will also conduct formal statistical analysis presented below.

(A) Avg. Centrality By Treatment and Rd. (Increasing)



(B) Avg. Centrality By Treatment and Rd. (Decreasing)

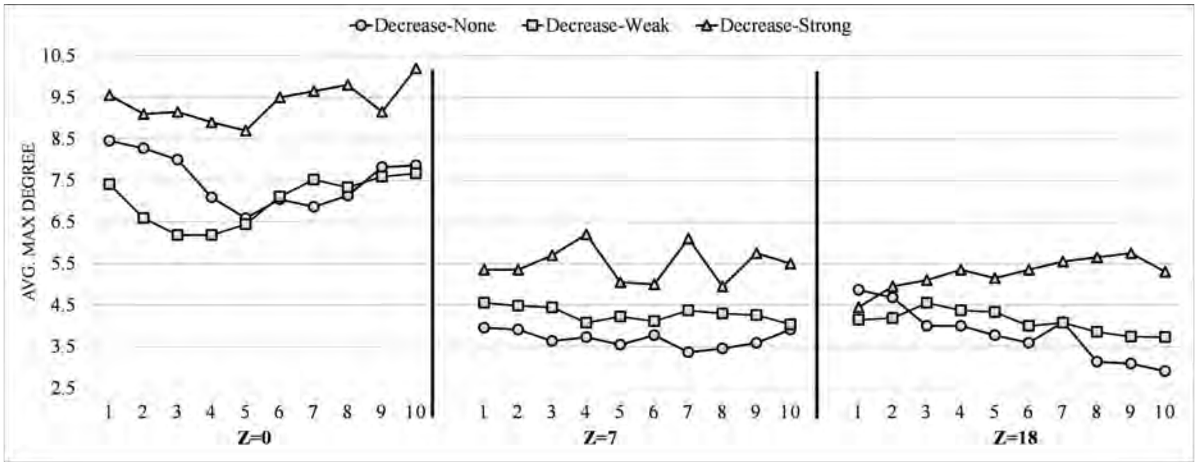


Figure 6: Average Centrality by Treatment and Round

Result 3: *Designers construct networks with lower average centrality as observation increases.*

As shown in Figure 7, after controlling for the production incentive, there is a distinct decrease in the group centrality as observation increases. This pattern holds for all production incentives,

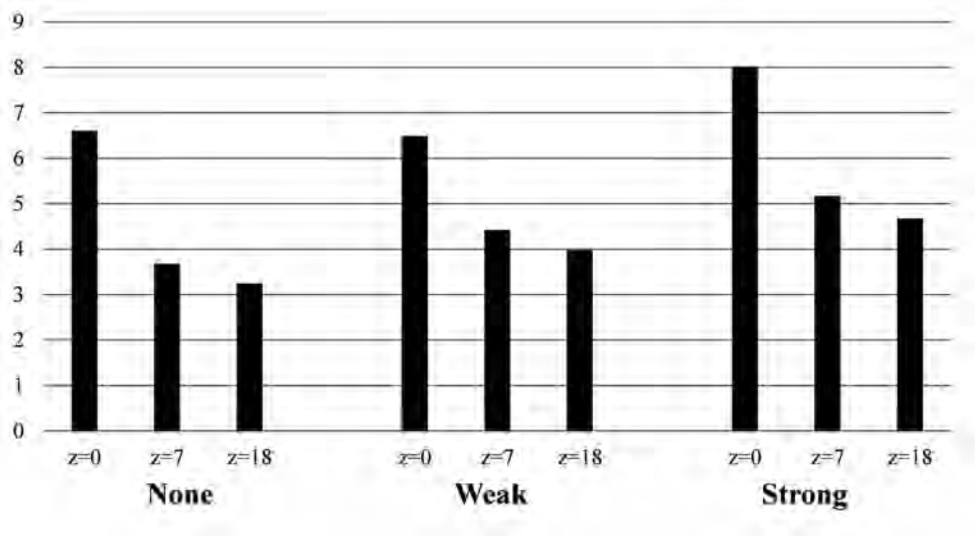


Figure 7: Average Centrality Rounds 4-10 (Incentive Fixed)

and we confirm the statistical significance of our results in regressions presented later.

Result 4: *Designers construct networks with higher average centrality as the production incentive increases.*

After controlling for observation level, we see a distinct increase in the group centrality as we increase the production incentive, as see in Figure 8. This pattern holds for both partial and full observation, but is unexpectedly weaker under no observation. Statistical significance of this effect is presented in the regression analysis below.

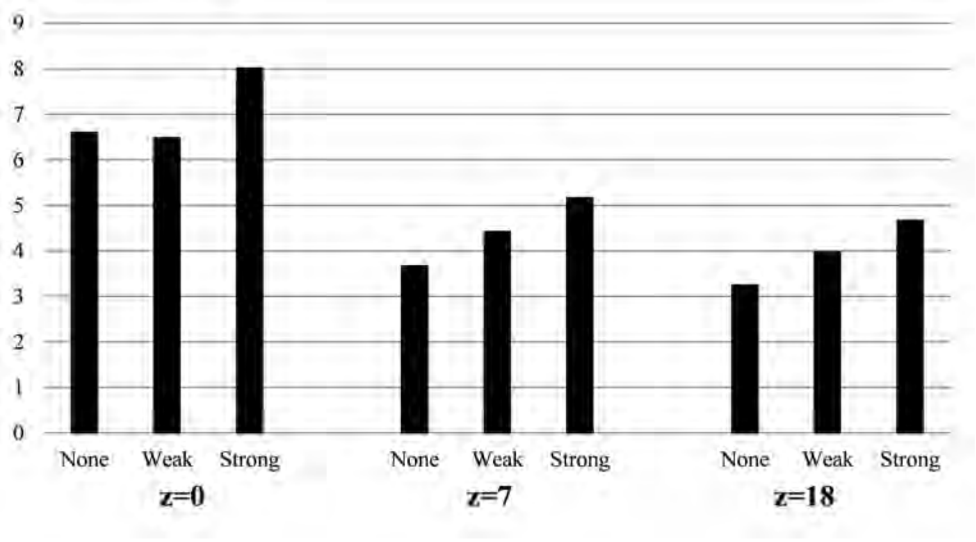


Figure 8: Average Centrality Rounds 4-10 (Observation Level Fixed)

To formally test Results 3-5, we ran a series of linear regressions which offer a prediction of the impact on group centrality for individual characteristics of each treatment. Results are given in Table 9. Regressions 1 and 2 were run on all of the observations while regressions 3 and 4 excluded the observations associated with weak production incentives and moderate monitoring $(z = 7)$. Each regression uses decreasing observation, the None incentive, and Hidden observation as the excluded categories. Regression 1 includes dummies for increasing observation, weak incentive, strong incentive, moderate observation $(z = 7)$, and full observation $(z = 18)$. We also include a variable for the round number and round squared in order to control for learning that took place during the 10 rounds of each observation level. Regression 2 contains the same variables and includes subject fixed effects.

	All observations		Excluding Weak and Partial	
	1	2	3	4
Constant	0.29***	0.29***	0.31***	0.31***
	(0.02)	(0.02)	(0.03)	(0.03)
Order (1=Increasing Obs.)	-0.06***	-0.06***	-0.12***	-0.12***
	(0.02)	(0.02)	(0.03)	(0.03)
Weak (1=Weak Prod. Incent.)	0.28	0.28	—	—
	(0.02)	(0.02)		
Strong (1=Strong Prod. Incent.)	0.08***	0.08***	0.07**	0.07**
	(0.02)	(0.02)	(0.03)	(0.03)
Partial (1=Partial Obs.)	-0.16***	-0.16***	—	—
	(0.02)	(0.02)		
Full (1= Full Obs.)	-0.19***	-0.19***	-0.20***	-0.20***
	(0.02)	(0.02)	(0.03)	(0.03)
Round Number	0.00	0.00	0.01**	0.01**
	(0.00)	(0.00)	(0.00)	(0.00)
Round Number Squared	0.00	0.00	0.00*	0.00*
	(0.00)	(0.00)	(0.00)	(0.00)
Fixed Effects	N	Y	N	Y
Clustering by Subject	Y	Y	Y	Y
Observations	4049	4049	1760	1760
Adj-R2	0.17	0.18	0.24	0.24
test: Strong-Weak=0	4.27**	4.28**	-	-
test: Full-Partial=0	8.04***	8.11***	-	-

Note: Dependent variable is the centrality of the network. Robust S.E. in parenthesis. *, **, and *** denote significance at 10%, 5%, and 1% levels.

Table 9: Regression Results, Pooled

The coefficients are virtually identical when subject fixed effects are included. There is a statistically significant decrease in the group centrality as the level of observation increases which provides formal statistical support for hypothesis 3. There is also a statistically significant increase in the group centrality as the centralization incentive increases which provides strong support for hypothesis 4.

We also ran multiple regressions on each individual session. These results can be seen in Table 10. These regression results provide strong support for hypothesis 3 within each individual treatment. Regression 10 shows that there is not a significant difference between partial and full observation in the decreasing-strong treatment, but we do see significant decentralization when comparing both partial and full observation with the no observation outcomes.

	Inc-None	Inc-Weak	Inc-Strong	Dec-None	Dec-Weak	Dec-Strong
	5	6	7	8	9	10
Constant	0.22***	0.25***	0.11***	0.58***	0.40***	0.45***
	(0.02)	(0.03)	(0.03)	(0.11)	(0.04)	(0.10)
Partial (1=Partial Obs.)	-0.15**	-0.06	-0.22***	-0.30***	-0.15***	-0.22***
	(0.06)	(0.04)	(0.06)	(0.09)	(0.05)	(0.07)
Full (1=Full Obs.)	-0.20***	-0.13**	-0.30***	-0.38***	-0.20***	-0.21**
	(0.07)	(0.05)	(0.07)	(0.11)	(0.06)	(0.10)
Round Number	0.01	0.00	0.03***	-0.01***	-0.01***	-0.00
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
Round Number Squared	0.00	0.00	0.00***	0.00	0.00**	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Fixed Effects	Y	Y	Y	Y	Y	Y
Clustering by Subject	Y	Y	Y	Y	Y	Y
Observations	720	600	660	660	809	600
Adj-R2	0.22	0.13	0.18	0.28	0.20	0.27
test: Full-Partial=0	7.37**	8.25***	14.62***	6.71**	6.36**	0.02

Note: Dependent variable is the centrality of the network. Robust S.E. in parenthesis. *, **, and *** denote significance at 10%, 5%, and 1% levels.

Table 10: Regression Results by Treatment

Result 5: *A strong order effect is found with no observation, and weakly under full observa-*

tion. Designers with decreasing observation create networks with much higher group centralization than designers with increasing observation. No difference is found for rounds with partial observation.

After controlling for the production incentive, designers facing a no observation environment should create networks with relatively similar levels of group centrality regardless of future or previous levels of observation. Our findings contradict this claim using both simple averages and regression analysis. We observe that designers who are initially exposed to high levels of monitoring tend to create more centralized networks on average for both full observation and no observation environments. A significant increase in centralization for only no observation environments might suggest the effect is primarily attributable to learning, but our results clearly show the same effect in the same direction for both full and no observation rounds. This clearly contradicts hypothesis 5 when there is both full and no observation. Under partial observation we see small yet statistically insignificant differences. This difference is demonstrated in Figure 9 as well as the regression results in Table 9. The coefficient on the increasing dummy is negative and significant in all four of our regressions. We have eliminated the first three rounds of each observation level in Figure 9 in order to reduce the possibility that these effects are learning related.

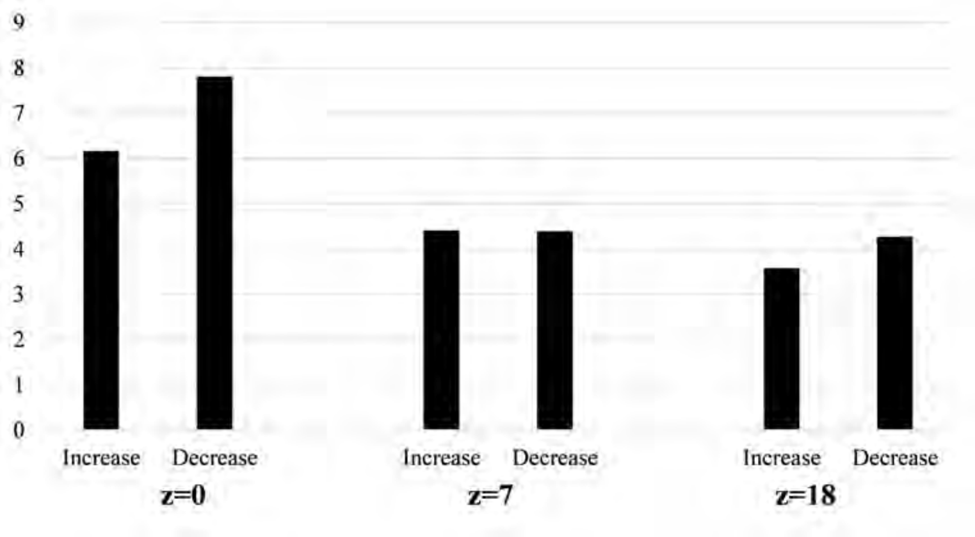


Figure 9: Average Centrality by Direction (Rounds 4-10)

2.5 Conclusion

The existing theoretical literature on dark network formation clearly outlines the logical argument that increased monitoring leads to a decentralized network structure and that decreased monitoring leads to a more centralized network structure. The field and empirical literature demonstrates a similar behavior among commonly studied dark networks, such as terrorist and criminal networks. A limitation of such work is that it is difficult, and unethical in most cases, to establish a direct causal relationship between monitoring and network structure. This paper attempts to resolve this challenge in a controlled environment where monitoring and production incentives can be varied exogenously, therefore allowing for a more legitimate demonstration of the implied direct causal relationship.

Using an experimental approach, we show that subjects tend to create more decentralized networks, on average, as the level of monitoring of their activities is increased. We also show that the friction between production and concealment is a salient one; i.e., the networks become more centralized as the production incentives for centralization increase. This experiment also provides evidence that subjects become more willing to risk exposure by centralizing in scenarios where early monitoring is extensive and then reduced over time as opposed to minimal monitoring that becomes more invasive over time.

It is instructive to consider some limitations of our study to understand possible directions for future work. In actual dark network disruption, there is an important distinction between killing a node and capturing a node. The tactics involved in each disruption may vary both in monetary cost and potential lives lost. For example, an aerial strike that puts only a pilot in harm's way may be sufficient to kill a prominent member of a dark network, while a team of operatives on the ground may be necessary to capture and kidnap. but the potential gains may also differ. Although both killed and captured nodes will need to be replaced, a captured node may be interrogated and potentially reveal more information that leads to a larger impact of an arrest, such as removal of that node's network neighbors. Because our experiment considers only the removal of a single node, it might better be thought of as killing rather than capture. Although Candelo et al (2014) vary the impact of an arrest in their experiment in a way that mimics capture and higher impact disruption, they do so under decentralized network formation and non-strategic disruption. Future work should consider the strategic interplay when different disruption impacts are possible. Included in this would be the possibility for the disruptor to decide on the level of impact of an arrest, presumably at a cost. The lessons learned from such work may better aid law enforcement in their decisions about not just whom to target when disrupting a dark network but also what type of disruption to pursue.

Another limitation of our study is that the network production occurs in a very stylized manner such that the highest degree node is the optimal node to remove. This simple formulation was intentional on our part for practical reasons associated with the experiment design. We wanted dark network production to be simple enough for our human subjects in both the designer and disruptor roles to understand and determine how best to design and disrupt. Future work can examine more complex network production functions that capture the richer features of a real world network. One possibility is that nodes can be of different types, e.g., planners, financiers, messengers, recruiters, foot soldiers, etc., and their locations in the network may reflect those roles. In such a setting, the best disruption might not be the node with highest degree but instead might be the node with highest betweenness. A systematic examination of richer network settings would be a step forward in this literature, though such work would need to overcome some of the practical challenges associated with placing human subjects into designer and disruptor roles. More complicated production settings are harder to convey to subjects and may lead to confusion that confounds the interpretation of experimental results. Recruiting subjects from the military or other law enforcement might be one way to overcome this hurdle.

3 A Game Theoretic Analysis of Absolute and Relative Grading Schemes

3.1 Introduction

According to Davis (2009), we generally observe two primary grading schemes in college level coursework, either standards-referenced grading or norm-referenced grading. Standards-referenced grading, also referred to as absolute performance evaluation, is the classical threshold style of grading where the grade is determined by the student's individual ability to achieve certain standards, such as answering more than 90% of test questions correctly. Norm-referenced grading, also referred to as relative performance evaluation, distributes grades relative to the performance of the class as a whole and is also commonly referred to as grading on a curve. These two schemes encompass a wide variety of specific grading methods and ultimately the finer details of the grading scheme lie with the particular faculty member (see Walvoord and Anderson (2011); Lang (2009)).

A great deal of controversy surrounds these two schemes in the academic community. Erickson, Peters, and Strommer (2009) suggests that decisions about grading schemes depend a great deal on an instructor's values and educational philosophy. Dominowski (2001) claims that standards-referenced grading is extremely inflexible and depends a great deal on the subjective "standards" of each faculty member, but Aviles (2001) suggests that the practitioners of norm-referenced grading are often using the relative scheme to mask teaching deficiencies. By definition we know that absolute standards more closely evaluate the volume of knowledge retained by students, while relative standards focus more on establishing a rank ordering of students by performance. One consideration could be that the scheme which best supports the purpose of higher education is the superior approach, but unfortunately disagreement between the human capital and signaling purpose of higher education only serves to further clutter the discussion. This controversy and the idiosyncratic nature of individual grading methods suggests a strong ideological influence on grading decisions. Unfortunately, there is also a shortage of research into the strategic influence that these two grading schemes have on student performance.

The purpose of this paper is to use a game theoretic approach to focus primarily on the influence that absolute and relative grading has on student's decisions to exert effort in the classroom. It will answer two primary questions. Is there a grading scheme that maximizes students' efforts? What are the factors which most significantly impact these behavioral choices? Where this paper differs from previous literature is that the purpose is not to study the utility maximizing quantity of effort for the student, but is instead to inform the faculty member about the choice of grading scheme. The student is simply presented with a decision of high or low effort based on their utility maximization and the strategic interdependence created by the grading scheme. The assumption that this paper makes in order to focus on strategic effort choices is that higher student effort is always the most desired outcome for the faculty member. The vast array of positive externalities attributed to education suggest that ultimately society benefits when the maximum number of students are exerting the maximum amount of effort (see Manda et al. (2002); Hall (2006)).

This paper shows that the effort maximizing grading scheme depends on the heterogeneity of the students. A comparison of the parameter values which allow for high levels of effort shows that both grading schemes become more effective among a more homogenous population, but the relative scheme is more significantly impacted by variation in ability levels. This paper also provides a clear explanation for the commonly observed occurrence of disadvantaged students in a mixed tournament exerting more effort than classical theory predicts (see Wu et. al (2007)). The model shows that there are several conditions under which an equilibrium can be reached in these mixed tournaments where high ability agents will choose low effort, but low ability agents will choose high effort. The implication is that high ability agents observe the disincentives of the relative scheme and choose low effort while the low ability agents observe the lack of effort by their high ability peers as an opportunity to close the distance and possibly realize the "surprise upset."

It is now important to mention the dynamics of the education process that this paper will refer to throughout. The process of education is composed of two primary components, the teaching and the learning. What the students learn is primarily a function of their ability and their effort towards learning. With regards to teaching, we assume there is a common influence across students which reflects the efficacy of the teaching, as well as a random component, unique to each student, which could be thought of as a result of different learning styles or simply as the measurement error of the assessment. The influence of each individual component is unobservable so educators must design a scheme by which they can determine the efficacy of their teaching and the ability-effort decisions of each student. The grading scheme is typically how this determination is made. These schemes are designed to reflect the desired information, but the influence that they have on the behavioral decisions of the students can be easily overlooked. The way in which we determine grades can reveal the information we desire, but can also influence the answer that we receive.

Section 3.2 will provide an overview of the existing literature pertaining to this subject. Section 3.3 will present the model and analysis in the absolute environment. Section 3.44 will present the model and analysis in the relative environment. Section 3.5 will conclude with a direct comparison of the two as well as a discussion of the implications of these findings.

3.2 Literature Review

There is little consensus with regards to the superiority of absolute and relative performance evaluation. There are prominent theoretical papers such as Lazear and Rosen (1981), Green and Stokey (1983), and Nalebuff and Stiglitz (1983) which indicate that relative performance evaluation is superior to absolute performance evaluation when agents are risk averse and there is a significant level of common shock across agents. This school of literature revolves primarily around the ability to provide a form of insurance to agents by removing the common shocks through relative evaluation. A recent empirical paper, Brownback (2018), suggests that as class size increases the increases in effort seen as a result of relative grading schemes also increase. In contrast, empirical papers such as Bandiera et al. (2005), Wu et al. (2005), and Marinakis (2018) have demonstrated that performance in the classroom as well as other forums is enhanced by using absolute standards. These results are attributed to disincentive effects in mixed tournaments, peer effects within groups that develop social connections, and the presence of negative externalities associated with success in a relative environment.

Other papers support the use of relative or rank-order performance standards such as Bull et al. (1987), Gibbons and Murphy (1990), and Cherry and Ellis (2005). Cherry and Ellis argue that they observe improved performance by higher ability students without a loss of performance from lower ability students due to the removal of a "peak effect." This effect is signified by high ability students exerting only enough effort to achieve top marks and then ceasing all effort for the remainder of the course. Bull et al. demonstrate in an experimental setting that the performance of test subjects in a laboratory converges to the superior theoretical levels of effort when using relative performance measures. The only exception to this conclusion is observed when there exists an advantage for some agents in which case this mixed tournament results in predicted performance levels for advantaged agents, but consistently higher than predicted performance levels for disadvantaged agents.

The education literature also contributes to the ambiguity of this issue (see Slavin (1977), Michaels (1977), Deutsch (1979), Natriello (1987), and Crooks (1988)). A common argument reflected in Slavin (1977), is that relative performance evaluation fosters a classroom environment which replaces collaboration with competition. This causes some students to increase their effort levels individually, but eliminates the student-to-student learning opportunities as individuals realize that they are now competing in a game filled with externalities. Bolocofsky and Mescher (1984) discuss the negative effect that this competition has on students who possess low self-esteem. Niederle and Vesterlund (2007) also show that women tend to self-select away from competition, which could mean that they are at a disadvantage in a relative grading environment regardless of ability. Crooks (1988) suggests that norm-referenced evaluation encourages students to attribute success and failure entirely to ability. This discourages both high and low ability students as the high ability recognize their advantage and so they discount the return on their effort while low ability students realize the futility of their effort in a grading scheme determined primarily by ability. In contrast though, Becker and Rosen (1992) show that competition can stimulate effort provided that the rewards for achievement are sufficient and the standards are not too high or too low.

The contribution of this paper is to present this comparison in a new light. The models that are used are vastly simplified from the true environment in the classroom, but are constructed to capture the essential relationships between grading scheme and behavioral choices. The results of this analysis offer support for classical tournament style contests, but also serve to enhance our understanding of the behavior of disadvantaged students in the classroom. While there exists disagreement over the purpose of grades within higher education, there is no disagreement with regards to the fact that greater effort always leads to an increase in learning. Therefore the objective for every instructor is to maximize the effort of all students through optimal grading scheme and course design. This paper serves to increase our understanding of what this optimal grading scheme is.

3.3 The Absolute Evaluation Standard Model

Assume that there are two types of agents, high ability (A_H) and low ability (A_L). The agents are aware of their individual type, as well as that of their peers. This assumption is based on the fact that by the time a student reaches a college classroom, they have undergone at least 12 years of education and during that time have been able to very clearly determine their own abilities and those of their peers, as demonstrated by Stipek (1981). Each agent will then choose one of two levels of effort, high effort (ε_H) or low effort (ε_L). This effort is costly to the student, but the costs are identical across abilities. The hour that a high ability student spends studying is not less costly, but is instead more effective than the hour that a low ability student spends studying. We will therefore assume that the cost of low effort is 0, while the cost of high effort is C regardless of ability. Assume also that the utility of passing the course is identical across abilities and therefore denote the utility of passing as $\dot{U}$. The utility of failing is equal to 0. One final piece of this model remains, the determination of an outcome as a function of the student's choice of effort. The determination will be done in a probabilistic manner. This interpretation will represent the stochastic components and measurement error that exists in most common forms of evaluation.

The probability of high performance will be denoted as $P_j(\varepsilon_i)$ where j denotes the agent's ability and i denotes the agent's chosen level of effort. Assume that receiving a passing grade requires high performance since anything less than that would provide a disincentive to the high ability agent. Therefore the expected payoff of each agent is

$$P_j(\varepsilon_i)\dot{U} - C(\varepsilon_i)$$

In both the absolute and the relative environments, it is impossible to distinguish a high ability student with low effort from a low ability student with high effort. This assumption is based on the standard assumption of the additive influence of ability and effort on performance (see Tsoulouhas (2018)). The observed performance is increased in a similar manner by either additional ability or effort. Given this assumption the probabilities for each of three performance levels are defined in table 11.

	High Ability, High Effort	**High Ability, Low Effort**	**Low Ability, High Effort**	**Low Ability, Low Effort**
High Performance	γ	α	α	0
Moderate Performance	$(1-\gamma)$	$(1-\alpha)$	$(1-\alpha)$	0
Low Performance	0	0	0	1

Table 11: Probability of Each Outcome as a Function of Ability and Effort

In the absolute environment the students do not consider the performance of their peers, but instead make their effort decision based solely on a cost-benefit analysis of high effort. The agent's choice of effort level is determined by comparing the expected payoff of high effort with the expected payoff of low effort. This leads to the following two conditions for high effort.

Condition 1: $\gamma > \alpha + \frac{C}{\dot{U}}$

Condition 2: $\alpha > \frac{C}{\dot{U}}$

High ability agents will choose high effort when condition 1 holds and low ability agents will choose high effort when condition 2 holds. In both models, assume that all agents choose low effort when they are indifferent between the payoffs. The utility of passing and the cost of effort are constant across agents therefore the parameters of interest are γ and α. These two conditions define four distinct regions in the parameter space for γ and α which result in four distinct outcomes. The first region refers to the values which satisfy both conditions and therefore will result in all students choosing ε_H. The second region refers to the values which don't satisfy either condition and will result in all students choosing ε_L. The third region is where condition one is satisfied, but condition two is not so only the high ability students will choose ε_H. The fourth region is where condition two is satisfied, but condition one is not so only the low ability students will choose ε_H. These regions can be seen more clearly in Figure 10 below. The discussion in section five will focus on the comparison of the first region in this model to the corresponding regions in a relative grading scheme.

This division illustrates two main points about the performance of students in an absolute environment. The first is that the cost of effort plays a significant part in establishing the region of the parameter space where high effort is optimal. This is not counterintuitive, but it does stress the fact that reducing the cost of effort for students, possibly through mobile or online technology, could have a significant effect on student performance. The second point is that merely making a course more difficult is counterproductive where effort is concerned. Instructors should instead be focusing on creating distance between the γ and α parameters. The practical meaning of this distance is merely that effective assessment should demand preparation more than pure ability. This could easily be achieved by asking questions from the textbook which were not explicitly discussed in lecture or by asking students to comment on current affairs which relate to the subject matter.

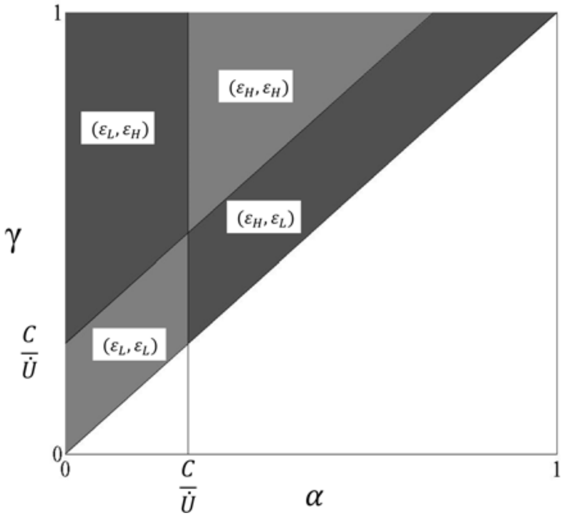


Figure 10: Parameter Space Defined Under Absolute Standards (Displayed as (A_L, A_H))

3.4 The Relative Evaluation Standard Model

The relative performance model is defined in exactly the same way as the absolute model, with the one exception being that there are now only two agents, one of which will pass and the other will fail. The outcome of this game is determined merely by observed performance. The student who performs at a higher level passes regardless of their outcome in absolute terms, and if the two perform equally then each will have a 50-50 chance of passing the course. This scheme creates three possible games that the agents could be playing. The first is a match between two high ability agents, the second is a match between two low ability agents, and the third is a match where the abilities are mixed with one being high and the other being low. Each of these matches can be represented by a 2x2 game matrix. Figure 11 outlines the three matches and associated payoffs. These payoffs are derived using the expected payoff, under the assumption that once both agents have chosen their strategies, there are nine possible outcomes each with a unique joint probability determined by the effort choices of the two agents. This paper will consider the three games as three different cases and address each individually.

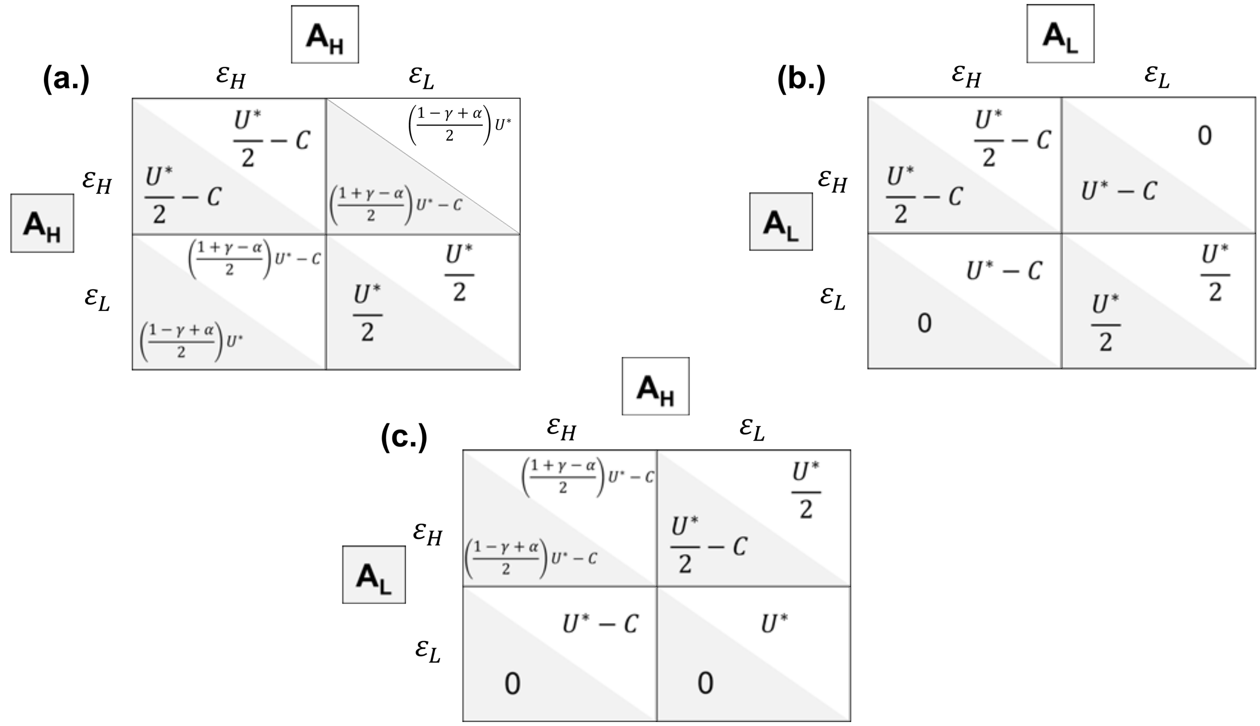


Figure 11: Payoff Matrices for Three Matches

3.4.1 Case One: Two High Ability Agents

Figure 11(a.) gives the payoff matrix for the game involving two high ability agents. The Nash Equilibrium of this game is also a dominant strategy equilibrium, but the strategies that compose this d.s.e. depend on the values of γ and α. For a distinct range of values there exists a dominant equilibrium where both agents choose high effort, but any values outside of that range result in a dominant equilibrium where both agents choose low effort. The condition which allows for the high effort equilibrium is $\gamma > \alpha + \frac{2C}{\dot{U}}$. It should be noted that this condition is very similar to condition 1 in the absolute model, but condition 2 from that model is not required. Figure 12 demonstrates this division of the parameter space.

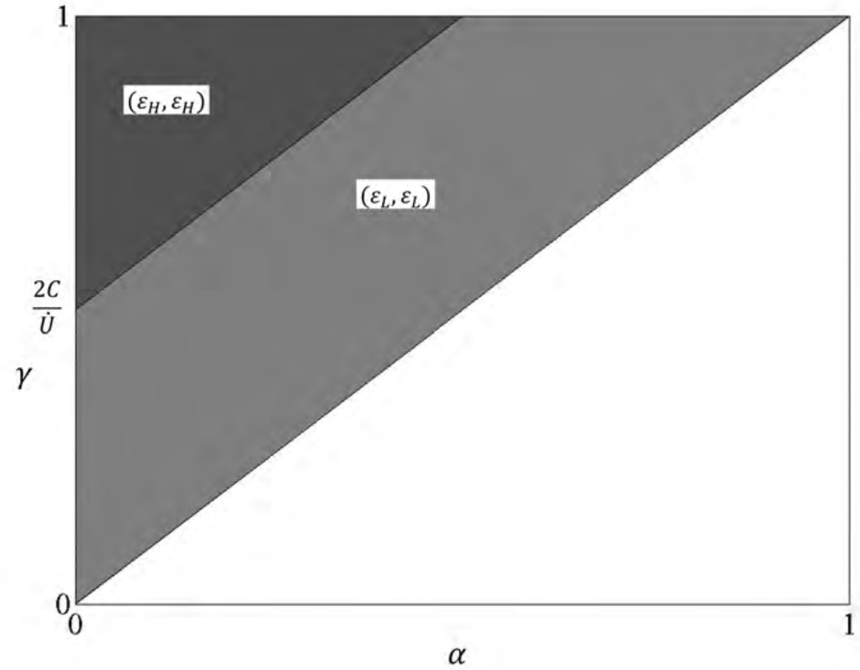


Figure 12: Parameter Space for Case One (Equilibrium Strategy Displayed as (A_H, A_H))

3.4.2 Case Two: Two Low Ability Agents

Figure 11(b.) gives the payoff matrix for the game involving two low ability agents. It should be clear with this game that the payoffs are now a function of the constant values determined for C and $\dot{U}$. As long as $1 > \frac{2C}{\dot{U}}$ then all values in the parameter space result in a dominant strategy equilibrium of high effort by both agents. If this initial condition is not met then all values in the parameter space result in a dominant strategy equilibrium of low effort by both agents. This clearly results in the largest parameter space motivating an equilibrium of high effort by all agents thus far. It again becomes obvious that the cost of effort in this environment is a key component to the size of the desired parameter space.

3.4.3 Case Three: Mixed Abilities (One High Ability Agent and One Low Ability Agent)

Figure 11(c.) gives the payoff matrix for the game involving mixed abilities. Since the determination of optimal strategies is more complicated for this game, we will first define the conditions under which certain strategies are optimal for each ability type. It will also become clear that the value of $\frac{2C}{U}$ plays a key role in the determination of equilibria so we will define the three possible ranges that this proportion can take and the equilibria associated with each.

Proposition 1. *High Ability and low ability students will choose high or low effort based on the values of* γ,α, *and* $\frac{2C}{U}$ *as defined below:*

(a) A_H *will always choose low effort if* A_L *chooses low effort*

(b) A_H *will choose high effort if* $\gamma > \alpha + \frac{2C}{U}$ *and* A_L *chooses high effort*

(c) A_L *will choose high effort if* $\gamma < \alpha + (1 - \frac{2C}{U})$ *and* A_H *chooses high effort*

(d) A_L *will choose high effort if* $\frac{2C}{U} < 1$ *and* A_H *chooses low effort*

Proposition 1b. - 1d. establish three possible ranges for the value of $\frac{2C}{U}$, each with its own unique Nash equilibria.

Range 1: $\frac{2C}{U} > 1$

When $\frac{2C}{U}$ is defined in this range the game fails to satisfy Proposition 1b. -1d. This implies that A_L will always choose low effort which means that the best response of A_H will always be to choose low effort. The resulting dominant strategy equilibrium is for both students to choose low effort regardless of the values for γ and α.

Range 2: $1 > \frac{2C}{U} > 0.5$

When $\frac{2C}{U}$ is defined in this range we have different equilibrium depending on the chosen values for γ and α. Proposition 1d. is clearly satisfied and the incorporation of Proposition 1b. and 1c. into the parameter space reveals three distinct regions. The first region is that which satisfies Proposition 1c., the second is the space which doesn't satisfy either Proposition 1b. or 1c., and the last region is the space which satisfies only Proposition 1b. We observe the same unique Nash Equilibrium in both region 1 and region 2 where A_H chooses low effort and A_L chooses high effort. Region 3 does not result in any Nash Equilibrium because for all outcomes at least one student has an incentive to deviate.

Range 3: $0.5 > \frac{2C}{U}$

When $\frac{2C}{\dot{U}}$ is defined in this range we again have different equilibrium depending on the chosen values for γ and α. Proposition 1d. is again satisfied and the incorporation of Proposition 1b. and 1c. into the parameter space again form three distinct regions. The first region is that which only satisfies Proposition 1c., the second is the space which satisfies both Proposition 1b. and 1c., and the last region is the space which satisfies only Proposition 1b. The only Nash Equilibrium in region 1 is where A_H chooses low effort and A_L chooses high effort. Region 3 again results in zero Nash Equilibrium as at least one agent always has an incentive to deviate regardless of the outcome. Fortunately, the overlap of Propositions 1b. and 1c. within region 2 results in a unique Nash Equilibrium where both agents choose high effort. This is the only occurrence under the mixed game where our optimal equilibrium of unanimous high effort is achieved. Figures 13 and 14 show the parameter space under Ranges 2 and 3 respectively.

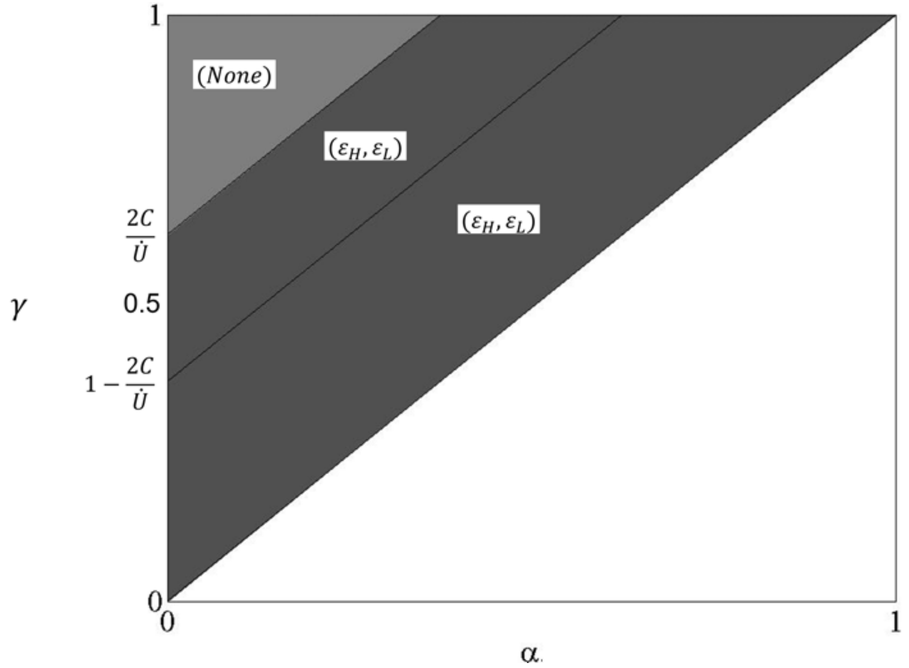


Figure 13: Parameter Space for Range 2 (Equilibrium Strategy Displayed as (A_L, A_H))

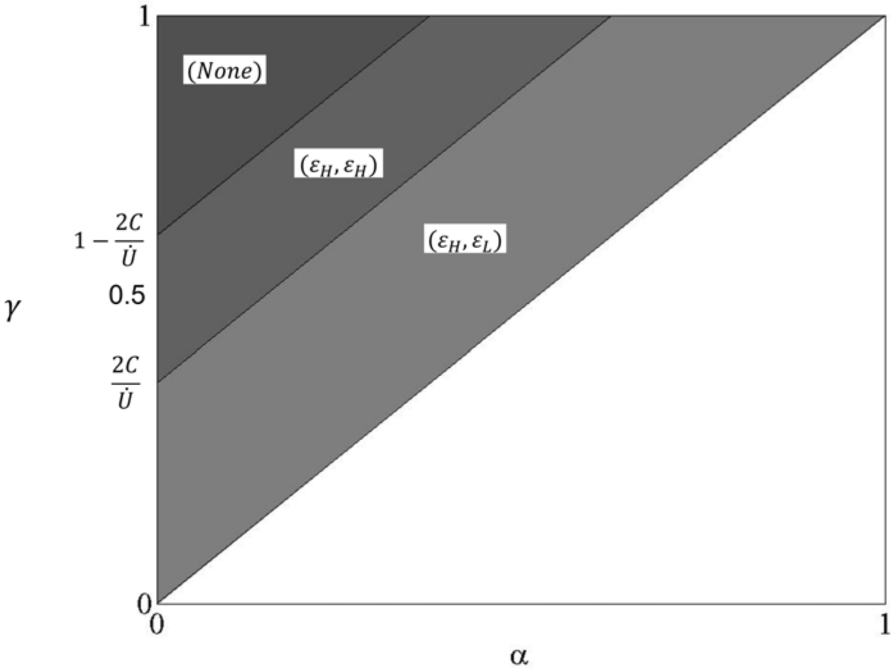


Figure 14: Parameter Space for Range 3 (Equilibrium Strategy Displayed as (A_L, A_H))

This third case is clearly distinct from the first two in that the opportunities for unanimous high effort equilibria are much more limited. It is a well-known fact that mixed tournaments result in disincentive effects which reduce the optimal effort choices and the results of this model are entirely in-line with our previous understanding of mixed tournaments. A significant contribution of this model is the vast array of parameter spaces which result in Nash Equilibrium of A_H choosing low effort and A_L choosing high effort. Several field and laboratory experiments have been conducted to identify the theorized disincentive effects of mixed tournaments, see Wu et. al (2007), Schotter and Weigelt (1992), and Bull, Schotter, and Weigelt (1987). Each of these papers identified an oversupply of effort by disadvantaged subjects, but none could clearly identify the reason for this behavior. The results above offer a simple explanation, the parameters fall within the many spaces resulting in our A_H choosing low effort and A_L choosing high effort Nash Equilibrium.

3.5 Conclusion

This paper defines all of the models over the same parameter space, allowing for a direct comparison across models. As stated earlier, the assumption that this paper rests its analysis on is that the policy goal for faculty should be the motivation of the maximum effort from all students. This assumption is made in consideration of the welfare of an educated society above that of the individual student. Several papers, including the two cited above, clearly demonstrate the importance of education and the positive externalities that society realizes from a highly educated workforce (see Higher Education (1998); Moretti (2004); Krueger and Lindahl (2000)).

The models used in this paper define the probability of success in a very targeted way, but in reality these parameters are far less controllable. The implication of this observation is that the optimal grading scheme will be the one with the largest portion of the parameter space resulting in simultaneous high effort outcomes from all ability levels. The most obvious conclusion of this comparison is that when considering students with homogeneous ability, both models maintain a large region of high effort outcomes. The results of case two also indicate that when addressing a group of low ability students, the relative scheme is far superior to that of the absolute. This is not always the case when addressing a group of high ability students, where the optimal scheme is impacted by the magnitude of the cost of effort. In this "gifted" type of environment, a very low cost means that the advantage goes to the relative scheme, but a significantly higher cost of effort can swing that advantage to the absolute scheme.

This raises an interesting question about the cost of effort relative to ability. The standard assumption in most theories related to information asymmetries and ability is that high ability agents have a lower cost of effort and are therefore willing to exert more effort towards education in order to distinguish themselves, see Spence (1973). This conclusion would seem to imply that the relative scheme is superior in the gifted environment, but what if opportunity costs are considered? Wouldn't the high ability agent also have a higher opportunity cost of their time since their high ability could be used in several other capacities? This would imply that high ability agents have a higher cost of effort which would result in the superiority of the absolute scheme. This is clearly a question which cannot be answered by theory alone, so an interesting extension of this paper would be to collect either field data about costs or laboratory data about effort decisions and to use that empirical information to provide clarity on this matter.

When a mixture of ability levels are studied the absolute model becomes far more restrictive and the relative model nearly eliminates all equilibria involving unanimous high effort decisions. When comparing these two schemes in a mixed environment it is obvious that the absolute scheme offers more opportunity for high effort outcomes. The relative environment only offers one specific region within one specific range of values where a high effort strategy is adopted by both ability levels. This conclusion is also one which could be further applied using empirical data. Since very little data has been collected in regards to the γ and α parameters defined above, some empirical research into the distribution of these parameters could add more clarity to the conclusions of this model.

These models also provide some insight into the behavior of the low ability student when paired against a high ability student. The literature on disincentives clearly shows that the disadvantaged student should forego all effort as a result of the inevitability of their inferior performance. The suggestion of Wu et al. (2007) is that the relative scheme removes a portion of the stochastic component of the observed performance, therefore ensuring that the disadvantaged player has less of a chance for the “surprise upset.” In field and laboratory experiments it was observed that this did not occur. Disadvantaged players were instead exerting a higher than anticipated level of effort. The models outlined above provide a simple and clear explanation for this phenomenon. There are several parameter values within the relative environment where an equilibrium is reached involving high effort for the low ability player. Additional empirical work, either in the field or experimental settings, would be needed to verify this equilibrium, but this is a strong avenue for further research on this topic.

This result has significant implications for the use of relative grading schemes in the college classroom, specifically when considering coursework at the graduate level. In general it is assumed that the low ability students are the group that requires greater incentives to elicit high effort levels, but if this is occurring at the graduate level then a vast majority of the highest ability students are underperforming due to a lack of incentives. The appropriate scientific approach moving forward is that new empirical work must be done to not only study effort decisions, but to also attempt to understand the strategy of students. It would also be useful to collect data on how accurately students can identify their abilities relative to their peers.

There are several immediate extensions of these models that could be undertaken to expand the insight into these behavior choices. The first is clearly to adapt the models to a field or laboratory environment and begin to collect empirical evidence of the theorized behaviors. Another interesting extension would be to assume a continuum of students whose abilities were drawn from a distribution and compare the resulting outcomes with those outlined in this paper. Additionally, the assumption of complete information could be relaxed and a Bayesian type analysis could be performed within the relative model to better represent the true classroom environment. As long as grades continue to be the means by which we measure education, there will continue to be a need to better understand the motivations that these grades provide to students.

References

[1] Albert, R., Jeong, H., Barabási, L., 2000. Error and Attack Tolerance of Complex Networks. Nature 406, 378-382.

[2] Alcantud, José and Annick Laruelle. 2014. "Dis&Approval Voting: A Characterization." Social Choice and Welfare, 43(1), 1-10.

[3] Arce, D. et al, 2011. Terrorism Experiments. Journal of Peace Research 48, 373-382.

[4] Arce, D., et al, 2012. Weakest-link Attacker-defender Games with Multiple Attack Technologies. Naval Research Logistics 59, 457-469.

[5] Arin, K.P. et al, 2011. Exploring the Dynamics between Terrorism and Anti-Terror Spending: Theory and UK-Evidence. Journal of Economic Behavior & Organization 77, 189-202.

[6] Arquilla, J. and Ronfeldt, D. 2001. Networks and Netwars: The Future of Terror, Crime, and Militancy. Santa Monica, CA: RAND Corporation, 2001.

[7] Arteaga, C. (2018). The effect of human capital on earnings: Evidence from a reform at Colombia's top university. Journal of Public Economics, 157, 212–225.

[8] Aviles, C. (2001). Grading with norm-referenced or criterion-referenced measurements: To curve or not to curve, that is the question. Social Work Education, 20(5), 603–608.

[9] Baccara, M., Bar-Isaac, H., 2008. How to Organize Crime. Review of Economic Studies 75, 1039-1067.

[10] Baker, W. and Faulkner, R. 1993. The Social Organization of Conspiracy. American Sociological Review. Vol. 58(6): 837-860.

[11] Bakker, R. M., Raab, J., Milward, H. B., 2012. A Preliminary Theory of Dark Network Resilience. Journal of Policy Analysis and Management 31, 33-62.

[12] Ballester, C. et al., 2006. Who's Who in Networks. Wanted: The Key Player. Econometrica 74, 1403-1417.

[13] Balinski, M. L. and Rida Laraki. 2010. Majority Judgment : Measuring, Ranking, and Electing / Michel Balinski and Rida Laraki. Cambridge, Mass: MIT Press.

[14] Balinski, M. and R. Laraki. 2011. "Election by Majority Judgment: Experimental Evidence," B. Dolez, B. Grofman and A. Laurent, In Situ and Laboratory Experiments on Electoral Law Reform: French Presidential Elections. New York: Springer, 13-54.

[15] Balinski, Michel and Rida Laraki. 2011. Approval Voting. The MIT Press.

[16] Bandiera, O., Barankay, I., & Rasul, I. (2005). Social Preferences and the Response to Incentives: Evidence from Personnel Data. The Quarterly Journal of Economics, 120(3), 917–962.

[17] Baujard, Antoinette; Frédéric Gavrel; Herrade Igersheim; Jean-François Laslier and Isabelle Lebon. 2018. "How Voters Use Grade Scales in Evaluative Voting." European Journal of Political Economy, 55, 14-28.

[18] Baujard, Antoinette and Herrade Igersheim. 2011. "Reconstructing Preferences in a Voting Context: Some Clues from Experimental Data on Voting Rules," Routlege, 273-91.

[19] Baujard, Antoinette; Herrade Igersheim and Isabelle Lebon. 2020. "Some Regrettable Grading Scale Effects under Different Versions of Evaluative Voting." Social choice and welfare.

[20] Baujard, Antoinette; Herrade Igersheim; Isabelle Lebon; Frédéric Gavrel and Jean-François Laslier. 2014. "Who's Favored by Evaluative Voting? An Experiment Conducted During the 2012 French Presidential Election." Electoral studies, 34(C), 131-45.

[21] Becker, W., & Rosen, S. (1992). The learning effect of assessment and evaluation in high school. Economics of Education Review, 11(2), 107–118.

[22] Blais, André. 2000. To Vote or Not to Vote? : The Merits and Limits of Rational Choice Theory / André Blais. Pittsburgh, Pa: University of Pittsburgh Press.

[23] ____. 2006. "What Affects Voter Turnout?" Annual review of political science, 9(1), 111-25.

[24] Blais, André; Jean-François Laslier and Karine Van der Straeten. 2016. Voting Experiments. Cham: Springer International Publishing.

[25] Bolocofsky, D. N., & Mescher, S. (1984). Student characteristics: Using student characteristics to develop effective grading practices. The Directive Teacher 6, 11-23.

[26] Borgatti, S.P. 2003. The Key Player Problem. In: Breiger, R., Carley, K., Pattison, P. (Eds.). Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers. National Academy of Sciences Press, 241-252.

[27] Brams, S. J. and D. R. Herschbach. 2001. "The Science of Elections." Science (American Association for the Advancement of Science), 292(5521), 1449-49.

[28] Brams, Steven J. and Peter C. Fishburn. 2007. Approval Voting / Steven J. Brams, Peter C. Fishburn. New York: Springer.

[29] ____. 2005. "Going from Theory to Practice: The Mixed Success of Approval Voting." Social choice and welfare, 25(2/3), 457-74.

[30] Brownback, A. (2018). A classroom experiment on effort allocation under relative grading. Economics of Education Review, 62, 113–128.

[31] Bull, C., Schotter, A., & Weigelt, K. (1987). Tournaments and Piece Rates: An Experimental Study. Journal of Political Economy, 95(1), 1–33.

[32] Callaway, D. et al., 2000. Network Robustness and Fragility: Percolation on Random Graphs. Physical Review Letters 85, 5468-5471.

[33] Candelo, N., et al. 2014. Endogenous Formation of Dark Networks: Theory and Experiment. Manuscript.

[34] Cherry, T., & Ellis, L. (2005). Does Rank-Order Grading Improve Student Performance?: Evidence from a Classroom Experiment. International Review of Economics Education, 4(1), 9–19.

[35] Chowdhury, S. et al, 2013. An Experimental Investigation of Colonel Blotto Games. Economic Theory 52: 833-861.

[36] Cohen, R. et al., 2001. Breakdown of the Internet under Intentional Attack. Physical Review Letters 86, 3682-3685.

[37] Crooks, T. (1988). The Impact of Classroom Evaluation Practices on Students. Review of Educational Research, 58(4), 438–481.

[38] Daniel, L. Chen and Martin Schonger and Chris Wickens. 2016. "Otree—an Open-Source Platform for Laboratory, Online, and Field Experiments." Journal of Behavioral and Experimental Finance, 9, 88-97.

[39] Darmann, Andreas; Julia Grundner and Christian Klamler. 2017. "Election Outcomes under Different Ways to Announce Preferences: An Analysis of the 2015 Parliament Election in the Austrian Federal State of Styria." Public Choice, 173(1-2), 201-16.

[40] Davis, B. (2009). Tools for Teaching. Hoboken: John Wiley & Sons, Incorporated.

[41] Dechenaux, E., et al, forthcoming. A Survey of Experimental Research on Contests, All-Pay Auctions and Tournaments. Experimental Economics.

[42] Deutsch, M. (1979). Education and distributive justice: Some reflections on grading systems. American Psychologist, 34(5), 391–401.

[43] Dhillon, Amrita and Susana Peralta. 2002. "Economic Theories of Voter Turnout." The Economic Journal, 112(480), F332-F52.

[44] Roger L. Dominowski. (2014). Teaching Undergraduates (Vol. 75).

[45] Downs, Anthony. 1957. An Economic Theory of Democracy / Anthony Downs. New York: Harper.

[46] Duffy, John and Margit Tavits. “Beliefs and Voting Decisions: A Test of the Pivotal Voter Model.” American Journal of Political Science, 52(3), 603-618

[47] Enders, W., Jindapon, P. 2009. Network Externalities and the Structure of Terror Networks. Journal of Conflict Resolution 54, 262-280.

[48] Enders, W., Su, X., 2007. Rational Terrorists and Optimal Network Structure. Journal of Conflict Resolution 51, 33-57.

[49] Erickson, B., Peters, C., Strommer, D., & Erickson, B. (2006). Teaching first-year college students (Rev. and expanded ed.). San Francisco, CA: Jossey-Bass.

[50] Everton, Sean F. 2012. Disrupting Dark Networks. Cambridge: Cambridge University Press.

[51] Farley, J., 2010. Breaking Al Qaeda Cells: A Mathematical Analysis of Counterterrorism Operations (A Guide for Risk Assessment and Decision Making). Studies in Conflict and Terrorism 26, 399-411.

[52] Feddersen, Timothy J. 2004. "Rational Choice Theory and the Paradox of Not Voting." The Journal of Economic Perspectives, 18(1), 99-112.

[53] Feddersen, Timothy and Alvaro Sandroni. 2006. "A Theory of Participation in Elections." The American Economic Review, 96(4), 1271-82.

[54] Feinstein, J., Kaplan, E., 2010. Analysis of a Strategic Terror Organization. Journal of Conflict Resolution 54, 281-302.

[55] Felsenthal, Dan S. 1989. "On Combining Approval with Disapproval Voting." Behavioral Science, 34(1), 53-60.

[56] Forsythe, Robert; Thomas Rietz; Roger Myerson and Robert Weber. 1996. "An Experimental Study of Voting Rules and Polls in Three-Candidate Elections." International journal of game theory, 25(3), 355-83.

[57] François, Abel and Olivier Gergaud. 2019. "Is Civic Duty the Solution to the Paradox of Voting?" Public choice, 180(3-4), 257-83.

[58] Franke, J., Öztürk, T., 2015. Conflict Networks, Journal of Public Economics 126, 104-113.

[59] Freeman, L., 1979. Centrality in Social Networks. Social Networks 1, 215-239.

[60] Freeman, L.C., Roeder, D., Mulholland, R.R., 1980. Centrality in Social Networks: II. Experimental Results. Social Networks 2, 119-141.

[61] Gambil, G.C., 2002. Sponsoring Terrorism: Syria and Hamas. Middle East Intelligence Bulletin 4, 10.

[62] Gerdes, L., ed. 2015. Illuminating Dark Networks. Cambridge: Cambridge University Press.

[63] Geys, Benny. 2006. "Explaining Voter Turnout: A Review of Aggregate-Level Research." Electoral studies, 25(4), 637-63.

[64] Gibbons, R., & Murphy, K. (1990). Relative Performance Evaluations For Chief Executive Office. Industrial & Labor Relations Review, 43(3), 30.

[65] Goyal, S., Vigier, A., 2014. Attack, Defence, and Contagion, Review of Economic Studies 81, 1518-1542.

[66] Green, J., & Stokey, N. (1983). A Comparison of Tournaments and Contracts. Journal of Political Economy, 91.

[67] Güth, W., Levínský, R., Pull, K., & Weisel, O. (2016). Tournaments and piece rates revisited: a theoretical and experimental study of output-dependent prize tournaments. Review of Economic Design, 20(1), 69–88.

[68] Hall, J. (2006). Positive externalities and government involvement in education. The Journal of Private Enterprise, 21(2), 165–175.

[69] Harfst, Philipp; Damien Bol and Jean-François Laslier. 2021. "Designing Preference Voting." Electoral studies, 69, 102262.

[70] Higher education in the learning society. (1998, January 1). Great Britain National Committee of Inquiry into Higher Education, 1998. Retrieved from http://search.proquest.com/docview/59881345/

[71] Hong, S., 2011. Strategic Network Interdiction. Working Paper.

[72] Hoyer, B., De Jaegher, K., 2010. Strategic Network Disruption and Defense. Tjalling C. Koopmans Research Institute, Discussion Paper Series nr: 10-13.

[73] Kimenyi, M., Mwabu, G., & Manda, D. (2006). Human Capital Externalities and Private Returns to Education in Kenya. Eastern Economic Journal, 32(3), 493–513.

[74] Kosfeld, M., 2006. Economic Networks in the Laboratory: A Survey. Review of Network Economics 3, 20-42.

[75] Krebs, V., 2002. Mapping Networks of Terrorist Cells. Connections 24, 43-52.

[76] Krueger, A., & Lindahl, M. (2001). Education for Growth: Why and For Whom? Journal of Economic Literature, 39(4), 1101–1136.

[77] Lang, J. (2008). On course: A week-by-week guide to your first semester of college teaching. Cambridge, Mass: Harvard Univ. Press.

[78] Laslier, Jean-François. 1997. Tournament Solutions and Majority Voting / Jean-Francois Laslier. Berlin ;: Springer.

[79] Laslier, Jean-François and M. Remzi Sanver. 2010. Handbook on Approval Voting. Berlin, Heidelberg: Springer Berlin / Heidelberg.

[80] Laslier, Jean-François and Karine Van der Straeten. 2007. "A Live Experiment on Approval Voting." Experimental economics : a journal of the Economic Science Association, 11(1), 97-105.

[81] Lazear, E., & Rosen, S. (1981). Rank-order tournaments as optimum labor contracts. Journal of Political Economy, 89(5), 841–864.

[82] Ledyard, John. 1984. "The Pure Theory of Large Two-Candidate Elections." Public Choice, 44(1), 7-41.

[83] Ledyard, John O. 1981. "The Paradox of Voting and Candidate Competition: A General Equilibrium Analysis." 54-80.

[84] Levine, David K. and Thomas R. Palfrey. 2007. "The Paradox of Voter Participation? A Laboratory Study." The American Political Science Review, 101(1), 143-58.

[85] Marinakis, K (2018) “Absolute vs. Relative Incentives Under Peer Effects in Education” Working Paper

[86] McBride, M. Caldara, M. 2013. The Efficacy of Tables versus Graphs in Disrupting Dark Networks: An Experimental Study. Social Networks 35: 406-422.

[87] McBride, M. Hewitt, D. 2013. The Enemy You Can’t See: An Investigation of the Disruption of Dark Networks. Journal of Economic Behavior and Organization 93: 32-50.

[88] Medina Sierra, Luis Fernando. 2018. Beyond the Turnout Paradox : The Political Economy of Electoral Participation / Luis Fernando Medina Sierra. Cham, Switzerland: Springer.

[89] Michaels, J. (1977). Classroom Reward Structures and Academic Performance. Review of Educational Research, 47(1), 87–98.

[90] Miffen, T. et al., 2004. A Random Graph Model for Terrorist Transactions. IEEE Aerospace Conference Proceedings 5: 3258-3264.

[91] Milward, H.B. , Raab, J., 2006. Dark Networks as Organizational Problems: Elements of a Theory. International Public Management Journal, 9(3), 333-360.

[92] Moretti, E. (2004). Workers’ education, spillovers, and productivity: evidence from plant-level production functions. American Economic Review, 94(3), 656–690.

[93] Myerson, Roger B. and Robert J. Weber. 1993. "A Theory of Voting Equilibria." The American political science review, 87(1), 102-14.

[94] Nalebuff, B., & Stiglitz, J. (1983). Prizes and Incentives: Towards a General Theory of Compensation and Competition. Bell Journal of Economics, 14(1), 21–43.

[95] Natriello, G. (1987). The Impact of Evaluation Processes on Students. Educational Psychologist, 22(2), 155–175.

[96] Niedeble, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? (impact of workplace environment and gender on competition between employees). Quarterly Journal of Economics, 122(3), 1067–1101.

[97] Palfrey, Thomas R. and Howard Rosenthal. 1983. "A Strategic Calculus of Voting." Public choice, 41(1), 7-53.

[98] ____. 1985. "Voter Participation and Strategic Uncertainty." The American political science review, 79(1), 62-78.

[99] Raab, J., Milward., H.B., 2003. Dark Networks as Problems. Journal of Public Administration Research and Theory 13, 413-439.

[100] Riker, William H. and Peter C. Ordeshook. 1968. "A Theory of the Calculus of Voting." The American political science review, 62(1), 25-42.

[101] Roberson, B., 2006. The Colonel Blotto Game. Economic Theory 29, 1-24.

[102] Schotter, A., & Weigelt, K. (1992). Asymmetric Tournaments, Equal Opportunity Laws, and Affirmative Action: Some Experimental Results. The Quarterly Journal of Economics, 107(2), 511–539.

[103] Schram, Arthur and Joep Sonnemans. 1996. “Voter Turnout as a Participation Game: An Experimental Investigation.” International Journal of Game Theory 25 (September): 385-406.

[104] Shapiro, J., 2013. The Terrorist’s Dilemma. Princeton: Princeton University Press.

[105] Slavin, R. (1977). Classroom Reward Structure: An Analytical and Practical Review. Review of Educational Research, 47(4), 633–650.

[106] Simmel, 1906. “The Sociology of Secrecy and of Secret Societies. American Journal of Sociology. 11(4): 441-498.

[107] Spence, M. (1973). Job market signaling. Quarterly Journal of Economics, 87(3), 355–374.

[108] Stipek, D. (1981). Children's perceptions of their own and their classmates' ability. Journal of Educational Psychology, 73(3), 404–410.

[109] Tsoulouhas, T. (2017). Do tournaments solve the adverse selection problem? Journal of Economics & Management Strategy, 26(3), 675–690.

[110] Tullock, Gordon. 1967. Toward a Mathematics of Politics. Ann Arbor: University of Michigan Press.

[111] Van der Straeten, Karine; Jean-François Laslier; Nicolas Sauger and André Blais. 2010. "Strategic, Sincere, and Heuristic Voting under Four Election Rules: An Experimental Study." Social Choice and Welfare, 35(3), 435-72.

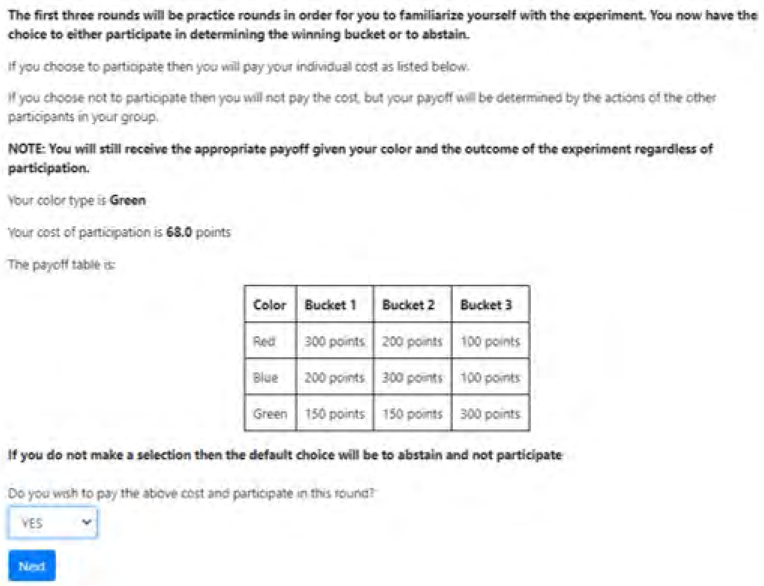
[112] Walvoord, B., & Anderson, V. (1998). Effective Grading: A Tool for Learning and Assessment (Vol. 28, p. 250). Jossey-Bass Publishers.

[113] Weiss, Andrew. 1995. "Human Capital vs. Signalling Explanations of Wages." Journal of Economic Perspectives, 9 (4): 133-154.

[114] Wu, S., Roe, B., & Sporleder, T. (2007). Mixed Tournaments, Common Shocks, and Disincentives: An Experimental Study. American Agricultural Economics Association>2007 Annual Meeting, July 29-August 1, 2007, Portland, Oregon. Retrieved from http://ageconsearch.umn.edu/record/9703/

A Voting Experiment Screenshots

This appendix contains screenshots from the voting experiment in Section 1. Figure 15 shows the participation decision that each subject faced at the beginning of each round.



The first three rounds will be practice rounds in order for you to familiarize yourself with the experiment. You now have the choice to either participate in determining the winning bucket or to abstain.

If you choose to participate then you will pay your individual cost as listed below.

If you choose not to participate then you will not pay the cost, but your payoff will be determined by the actions of the other participants in your group.

NOTE: You will still receive the appropriate payoff given your color and the outcome of the experiment regardless of participation.

Your color type is **Green**

Your cost of participation is **68.0** points

The payoff table is:

Color	Bucket 1	Bucket 2	Bucket 3
Red	300 points	200 points	100 points
Blue	200 points	300 points	100 points
Green	150 points	150 points	300 points

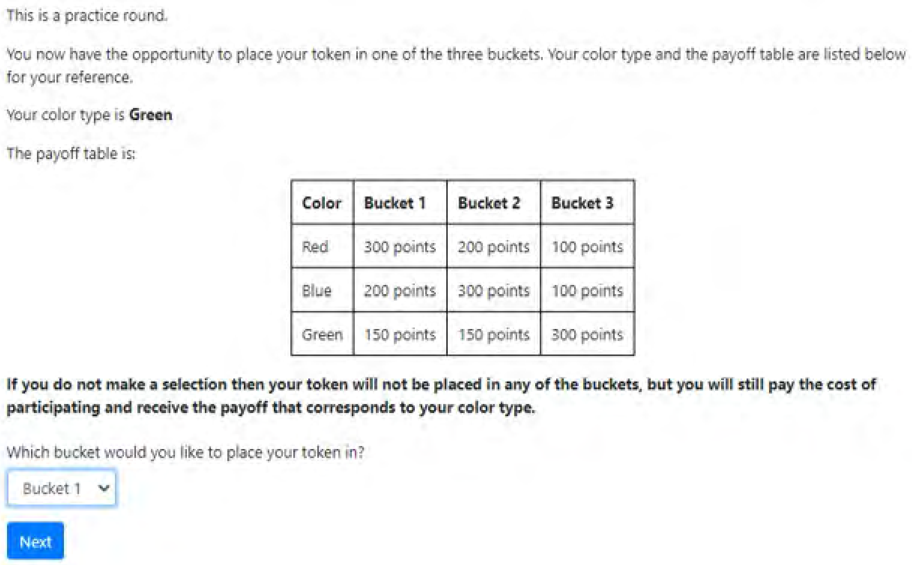
If you do not make a selection then the default choice will be to abstain and not participate

Do you wish to pay the above cost and participate in this round?

YES

Figure 15: Voting Participation Decision Screen

Figure 16 shows the player interface for the plurality phase while Figure 17 shows the interface for the evaluative phase.



This is a practice round.

You now have the opportunity to place your token in one of the three buckets. Your color type and the payoff table are listed below for your reference.

Your color type is **Green**

The payoff table is:

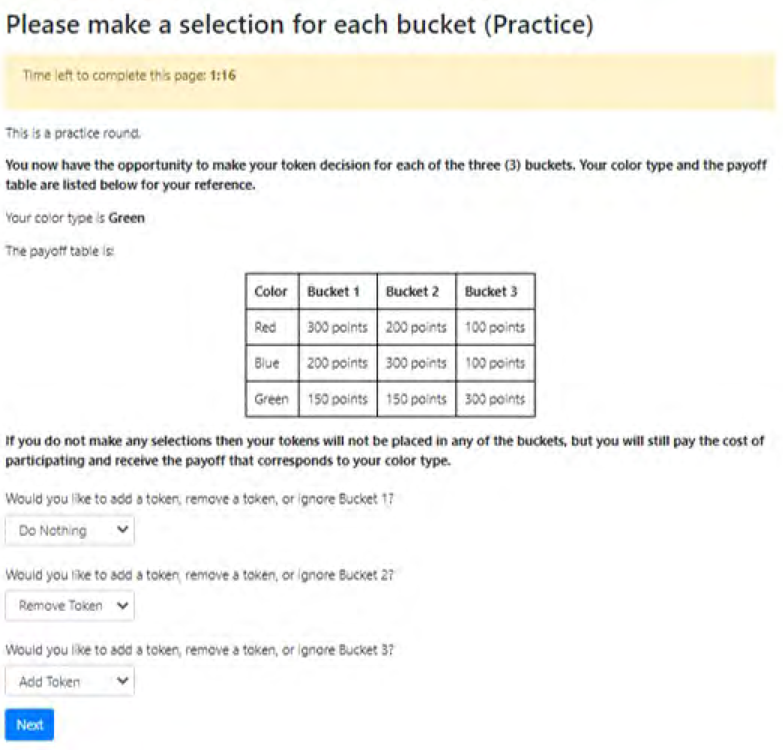
Color	Bucket 1	Bucket 2	Bucket 3
Red	300 points	200 points	100 points
Blue	200 points	300 points	100 points
Green	150 points	150 points	300 points

If you do not make a selection then your token will not be placed in any of the buckets, but you will still pay the cost of participating and receive the payoff that corresponds to your color type.

Which bucket would you like to place your token in?

Bucket 1

Figure 16: Primary Decision Screen (Plurality)


Please make a selection for each bucket (Practice)

Time left to complete this page: **1:16**

This is a practice round.

You now have the opportunity to make your token decision for each of the three (3) buckets. Your color type and the payoff table are listed below for your reference.

Your color type is **Green**

The payoff table is:

Color	Bucket 1	Bucket 2	Bucket 3
Red	300 points	200 points	100 points
Blue	200 points	300 points	100 points
Green	150 points	150 points	300 points

If you do not make any selections then your tokens will not be placed in any of the buckets, but you will still pay the cost of participating and receive the payoff that corresponds to your color type.

Would you like to add a token, remove a token, or ignore Bucket 1?

Do Nothing

Would you like to add a token, remove a token, or ignore Bucket 2?

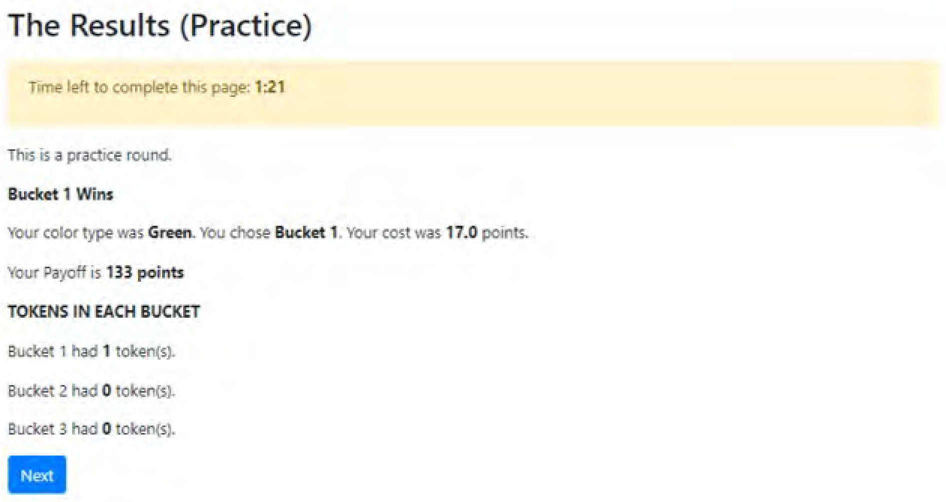
Remove Token

Would you like to add a token, remove a token, or ignore Bucket 3?

Add Token

Figure 17: Primary Decision Screen (Evaluative)

Figure 18 shows the results page that each subject observed at the end of each round and Figure 19 shows the payoff page that was displayed at the end of the experiment


The Results (Practice)

Time left to complete this page: **1:21**

This is a practice round.

Bucket 1 Wins

Your color type was **Green**. You chose **Bucket 1**. Your cost was **17.0** points.

Your Payoff is **133 points**

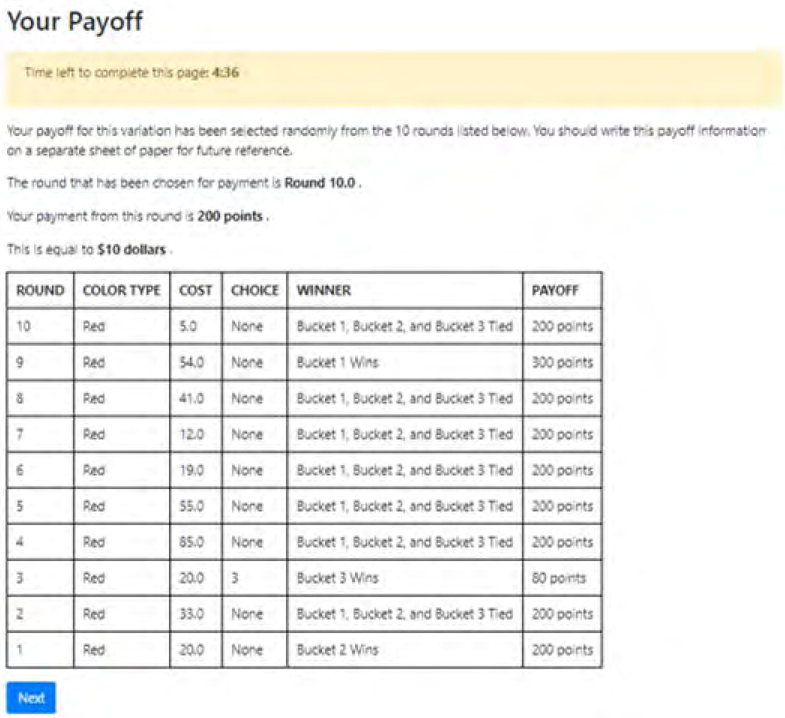
TOKENS IN EACH BUCKET

Bucket 1 had **1** token(s).

Bucket 2 had **0** token(s).

Bucket 3 had **0** token(s).

Figure 18: Results Screen



Your Payoff

Time left to complete this page: **4:36**

Your payoff for this variation has been selected randomly from the 10 rounds listed below. You should write this payoff information on a separate sheet of paper for future reference.

The round that has been chosen for payment is **Round 10.0** .

Your payment from this round is **200 points** .

This is equal to **$10 dollars** .

ROUND	COLOR TYPE	COST	CHOICE	WINNER	PAYOFF
10	Red	5.0	None	Bucket 1, Bucket 2, and Bucket 3 Tied	200 points
9	Red	54.0	None	Bucket 1 Wins	300 points
8	Red	41.0	None	Bucket 1, Bucket 2, and Bucket 3 Tied	200 points
7	Red	12.0	None	Bucket 1, Bucket 2, and Bucket 3 Tied	200 points
6	Red	19.0	None	Bucket 1, Bucket 2, and Bucket 3 Tied	200 points
5	Red	55.0	None	Bucket 1, Bucket 2, and Bucket 3 Tied	200 points
4	Red	85.0	None	Bucket 1, Bucket 2, and Bucket 3 Tied	200 points
3	Red	20.0	3	Bucket 3 Wins	80 points
2	Red	33.0	None	Bucket 1, Bucket 2, and Bucket 3 Tied	200 points
1	Red	20.0	None	Bucket 2 Wins	200 points

Figure 19: Payoff Screen