Undergraduate Research Journal

THE UNIVERSITY OF TEXAS AT AUSTIN™
Senate of College Councils

2016

Spring 2016 Volume 15 Number 1

The University of Texas at Austin Undergraduate Research Journal

Inquiries should be addressed to:

editorinchief@texasurj.com
The University of Texas at Austin
Senate of College Councils / URJ
2201 Speedway, Stop A6210
Austin, Texas 78712-1103
512-471-3166 Fax 512-471-3408

Volume 15, Number 1 Spring 2016

ISSN 1538-9421

Printed in the United States of America at
OneTouchPoint-Southwest Printing in Austin, Texas

THE UNIVERSITY
OF TEXAS AT AUSTIN
Senate of College Councils

Undergraduate Research Journal

Volume 15 Number 1 Spring 2016

The editors of the Journal wish to thank the Office of the Vice President for Research for its support. In particular, they thank Dr. Dan Jaffe, Vice President for Research, and congratulate him on an extraordinary first year in his new position.

They also thank the Division for Student Affairs for its continued belief in the mission of the UT Austin Undergraduate Research Journal. Most of all, they thank Dr. Gage Paine, who has recently left her role as Vice President for Student Affairs, for her many years of work to make UT Austin a safe and supportive environment for all students.

They also thank the University Co-op for its continued support, having been a friend of the UT Austin Undergraduate Research Journal since its inception.

They offer their appreciation as well to Ms. Veronica Cantu, Ms. Becky Carreon, Dr. Robert Reichle, the Office of Undergraduate Research, the Senate of College Councils, and the Office of the President for helping make The UT Austin Undergraduate Research Journal possible.

Finally, the editors thank the faculty reviewers and, most of all, the authors themselves, whose research continues to inspire all as they strive to change the world.

On behalf of the 2015-16 staff, I am delighted to present this year's edition of The University of Texas at Austin's *Undergraduate Research Journal*. Since 2001, the *Undergraduate Research Journal* has been committed to showcasing the best of the world-changing research being conducted by undergraduate students at UT Austin.

The 2016 issue brings together research from disciplines across the university. Several of our articles reveal truths about the human subject, whether through the lens of art or laboratory studies. Others present research into as diverse subjects as irrational powers of irrational numbers, wasp microbiomes, and patent policy. Still other articles model the obligations society has to the world at large, examining the effects of humanity on the environment. Each article offers a unique, important insight and has been thoroughly vetted by one or more leading faculty members to that effect.

Even as the *Journal* steadily continues its many years of service to the University and its community, the *Journal* itself has not remained static as an institution. Regular readers will note that many new roles have been introduced into the *Journal*'s staff, as the Executive Board seeks to utilize Longhorns' spectacular talents to the fullest extent possible. Events that have become mainstays of campus life (e.g., a business research panel, a reading-writing workshop, etc.) continued this past year, while the *Journal* also held its first humanities research panel and increased its support of the Office of Undergraduate Research's events during Research Week.

As our staff knows, this year has been a year for change. Therefore, I am immensely grateful for the patience, the hard work, and the independence demonstrated by our staff throughout this year. Working with the *Journal* has been an honor these past four years, and based on that experience I can say that the *Journal* is in its strongest position ever—much to the credit of the staff and probably in spite of myself. The *Journal* is a great institution because of its student authors, its faculty reviewers, and the everyday efforts of its staff members, whether an assistant editor that is new to the *Journal* or a seasoned veteran on the Executive Board.

I am excited to see where the *Journal* goes next, and I invite you to see how far we have come. To that end, I encourage you to browse our archive, as well as our regularly-updated research blogs, online at *texasurj.com*.

Sincerely,

Patrick Haley, Editor-in-Chief
Undergraduate Research Journal

Contents

UNDERGRADUATE RESEARCH JOURNAL | 2016

The Effects of Snorkel-Based Tourism on the Behavior of Reef Fishes

The University of Texas at Austin

Undergraduate Research Journal

Volume 15
Number 1
Spring 2016

Savannah Clapp, Lauren Rowsey, and Jordan Grant

Abstract

Coral reefs support a diverse array of marine life and have a specific set of conditions under which they can flourish. Factors such as the behavior of reef fish can impact a reef system's health. Tourism in Akumal Bay in Quintana Roo, Mexico, has experienced a rapid increase in the last five years. Tourists are perceived as predators, given their potential to alter the behavior of reef fish, such as movement, food intake, and reproduction. In this study, we investigated how snorkelers affect reef fish behavior as a measure of the impacts of increased tourism on reef health. We looked at four abundant Caribbean fish species in the

area: *Haemulon sciurus*, *Acanthurus coeruleus*, *Acanthurus bahianus*, and *Sparisoma viride*. We tested their behavior in response to tourism by measuring their flight initiation distance (FID), which is the distance between a tourist and a fish when the fish flees. To quantify snorkel-based tourism, we counted people in the water at set times of a day. We found that FID decreased as the number of tourists increased across all the species that were studied. Specifically, *S. viride* indicated the strongest correlation to tourism with an R^2 value of 0.1774 ($p = 0.0068$). In addition, we found a positive relationship between FID and structural proximity, or the distance to the nearest structure within which a fish could hide, and this relationship was highest in *S. viride* ($R^2 = 0.1904$; $p = 0.0049$). Our results suggest that tourism may be causing prey reef fish to become desensitized to consumptive predators, allowing the prey reef fish to invest more energy into reproduction and food intake. If prey reef fish have more energy for algae grazing, then reefs may be healthier due to less algae cover.

Introduction

Prey intimidation by a predator has as large an effect as direct consumption on prey demographics (Preisser et al. 2005). A prey's behavioral response to its predatory threats is an important factor that determines prey survival and fitness (Preisser et al. 2005). Such behavioral responses include defensive strategies that can result in different movement patterns, lowered food intake, and reduced reproduction (Madin et al. 2010). In particular, reduced energy intake through lowered feeding causes the prey to invest more energy into defensive maneuvers or strategies. This, in turn, reduces the success of mating and reproduction, thereby decreasing the population's size as a whole (Preisser et al. 2005).

Many natural predators are in decline in terrestrial and aquatic systems due to overharvesting (e.g., hunting and fishing; Estes et. al 2011). However, humans are increasing in many of these systems and could have important effects as predators. Bellwood et al. (2011) determined that the rise in human population and fishing pressure affect fish behavior in tropical coral reefs. Human visitors to coral reefs are capable of destroying coral habitat by trampling and anchoring in the area. These activities can alter shelter availability (rugosity of the reef) and increase fear responses in fish (Chabanet et al. 2005; Nunes et al. 2014). Tourists can also disrupt the environment through snorkeling or diving; although these activities may not directly destroy the environment, they increase fear responses in fish. The topic of human impacts on reef fish behavior is not widely studied and has been limited to studying the effects of fishermen, as opposed to the effects of non-predatory people, such as tourists, on behavior of prey fish. However, tourists have potential to alter prey fish behavior given their large size and disruptive behaviors such as spear-hunting. If prey fish are shown to change their behavior in response to tourists in this study, supposedly "harmless" human tourists could become a danger to the survival and reproductive success of coral reef prey fish.

Predatory effects on prey behavior have commonly been studied by measuring prey flight initiation distance (FID), which is the distance between a predator and a prey when

the prey flees. Studies done on FID in coral reefs have tested prey fishes' behavioral response to fishing pressure from fishers using spears or spear guns. There are several factors that influence a prey animal to flee or change behavior, including refuge availability, previous predation threat, and its size (Januchowski et al. 2011). Moreover, a predator's size can also affect a prey fishes behavior, because certain sized predators can be more intimidating to the prey (Madin et al 2010). Because they can potentially alter behavior in prey fish by influencing all of the mentioned factors, humans can be perceived as predators. One study done on humans as predators showed that as the level of perceived threat increased due to a rise in fishing pressure, the FID of prey fish increased (Januchowski et al. 2011). Another study concluded that behavioral responses to predation are scaled to the level of predation risk; the more severe the fishing pressure, the higher the FID. Furthermore, Januchowski et al. (2011) also discovered that fish in marine reserves showed higher FID when surrounding waters had increased fishing pressure. This indicates that fish in a protected zone, where predation risks are lower, can observe and adopt the behavior of fish from a nearby area that experienced predation from fishermen even though they may not have experienced this predation directly (e.g. fish within a marine reserve; Januchowski et al. 2014).

The effects that predators have on prey behavior can also be influenced by environment and habitat. Nunes et al. (2014) studied how habitat and structural complexity of a reef impacts anti-predator behaviors. They concluded that the more complex the habitat is, the less fear most prey fish showed, and that removing habitat increased fear and negatively impacted species survival and success (Nunes et al. 2014).

Akumal Bay, Mexico, is a site of major tourism for Mexico that has witnessed an increase in tourism from 78.1 million arrivals in 2013 to 81 million arrivals in 2014 (SECATUR n.d.). This high volume of tourism could have destructive effects on the reef systems in this area. The bay has experienced both a high spear-fishing pressure and an increase in snorkeling and marine recreational activities on the reefs that has led to disruptive practices such as grabbing sea turtles and standing on coral (Bellwood 2004, Garcia Salgado et al. 2008). Furthermore, Kramer (2002) showed that there had been a decrease in coral cover and an increase in macroalgal cover over the 10 years prior to 2002. Lacey (2013) also confirmed low coral cover in this area and dominant macroalgae in the bay as of 2012 (Harvell et al. 2007). Although many factors can influence macroalgae growth, its dominance can be an indicator that human "predators" are increasing FID in the area. As an increased FID is an indication that prey fish--mainly herbivores--are investing more time and energy into fleeing behaviors rather than feeding, this change could indicate that tourists are negatively influencing normal feeding patterns, reproduction, and movements by scaring the prey fish. These factors make Akumal Bay an important site to study the impact that snorkel-based tourism has on the FID of reef fishes.

This study focused on how tourists impact the response of three species of herbivorous reef fishes (*Sparisoma viride, Acanthurus coeruleus,* and *Acanthurus bahianus),* and one

non-herbivorous reef fish (*Haemulon sciurus*) to predatory threats. Little previous research has focused on this topic, which makes this paper valuable for better understanding the impact humans can have on the health of reef fish populations and benthic communities on coral reefs. The objective of this study was to determine 1) how the FIDs of these four species of reef fish are influenced by tourism, 2) how habitat affects FID, and 3) what other factors, such as group size and displayed behaviors, influence FID. We hypothesized that FID increases with increasing density of tourists in the bay and that more structural complexity in a habitat results in a lower FID.

Materials and Methods

Study site

This study was conducted in Akumal Bay in Quintana Roo, Mexico on the Yucatan peninsula from 21 May to 5 June of 2015, and it focused on the back reef area of the bay that spans both a high and a low tourist density area. Using Google Earth, Akumal Bay was divided into three horizontal rows of approximately 125 meter width that go from North to South. Using a Garmin GPS, reference points were created along each row for easy visualization during fieldwork. Each researcher estimated FID in each row in the morning and afternoon for six consecutive days. About 40 FID points were collected every day, resulting in a total of 214 data points over six consecutive days. Only four abundant species of reef fishes -- *Sparisoma viride, Acanthurus coeruleus, Acanthurus bahianus,* and *Haemulon sciurus* -- with body size within the range of 15 cm to 35 cm were considered during data collection.

Tourist Density

Tourist density was calculated in order to quantify the number of tourists in different regions of the Akumal Bay. The bay was divided into 36 total bins by first dividing the area of interest into 12 latitudinal bins of 50 meter width then dividing each latitudinal bin into three longitudinal bins. The first two longitudinal bins were 50 meters wide, whereas the last bin continued to the fore reef and was about 100 meters wide. To locate the bins easily while working on the shore, we navigated to a GPS point to establish bins on Google Earth then physically marked them using transect tape. We measured 25 meters on each side of each GPS point to create the latitudinal bin. We used a compass to establish a uniform heading by using a bearing from north of 127° and then used reference points to accurately visualize 50 m and 100 m distances offshore. We counted and recorded the number of tourists present in each cell every day using a mechanical counter at peak tourist hours of ca.1300, 1430, and 1600 h when the highest numbers of tourists are expected to be present in the water. Each FID point had a recorded GPS location, and it was assigned to an average tourist count of the tourist density bin that it was located in. Upon qualitative observation and later confirmation by quantifying tourist counts over the course of the

study, we determined the North side of Akumal Bay to be the highest tourist density area and the South side to be the lowest tourist density area.

Fish behavior and FID

In order to determine prey fish behavior in response to predatory threats, we measured prey fishes' FID in response to humans as a predatory threat. We recorded FID using methods developed from a previous study on prey fish responses to predatory threats from spear fishers (Januchowski et al. 2011). Because Januchowski's study focused specifically on spear-fishers as the human predator, protocols were modified in order to measure prey fish behavior in response to snorkeling tourists for the purpose of this study. First, a target fish of one of the four selected species was identified and approached from the surface. Every researcher mimicked tourist behavior by splashing their fins moderately and using their arms to swim forward at a uniform speed of 0.5 m/s. Because most tourist groups in Akumal are required to wear life jackets that prevent them from diving down, the observer remained at the surface until the fish fled. This behavior protocol was standardized across all observers who functioned as predators for the purpose of this study. FID was measured to the nearest centimeter using transect tape. All flight and non-flight behavior was recorded. Non-flight behavior was categorized as follows: 1) the target fish was still or possibly foraging at approach, and it remained still or kept foraging after approach (kept foraging), 2) the fish was swimming at approach, and there was no change in swimming speed or direction (no flight), and 3) the fish was either swimming or at a stop at approach and then swam towards the observer (kept watch). Flight behavior was categorized as follows: 1) the fish fled at a normal swimming speed in any direction (fled), 2) the fish fled after the observer was directly on top of it at the surface (fled after approach), 3) the fish fled into a coral structure (fled into coral), and 4) the fish was eating and then stopped to flee (stopped foraging). The observers categorized each fish to only one flight or non-flight behavior that identifies the prey fish behavior the best.

We recorded variables that address any relationship between habitat types and FID, such as target fish habitat, structural proximity, and percent cover of coral or seagrass. We measured target fish habitat by noting the location of the target fish in sand, seagrass, or coral habitat. Because there were very little seagrass samples, the seagrass habitat group was later merged into the sand habitat group during data processing. Structural proximity was measured as the distance between the target fish and a structure, which was defined as anything large enough that the fish could hide in or around of (e.g. coral or rocks). Percent cover of each habitat type was calculated using Google Earth and ImageJ by determining a 10 meter radius circle around the location of an FID point, calculating the area of coral or seagrass cover, and calculating the percent cover as a fraction of each habitat cover area over the total area of the 10 meter radius circle. Finally, a total of 214 FID data points were analyzed using statistical tools, such as linear regressions and Student's t-tests.

Results

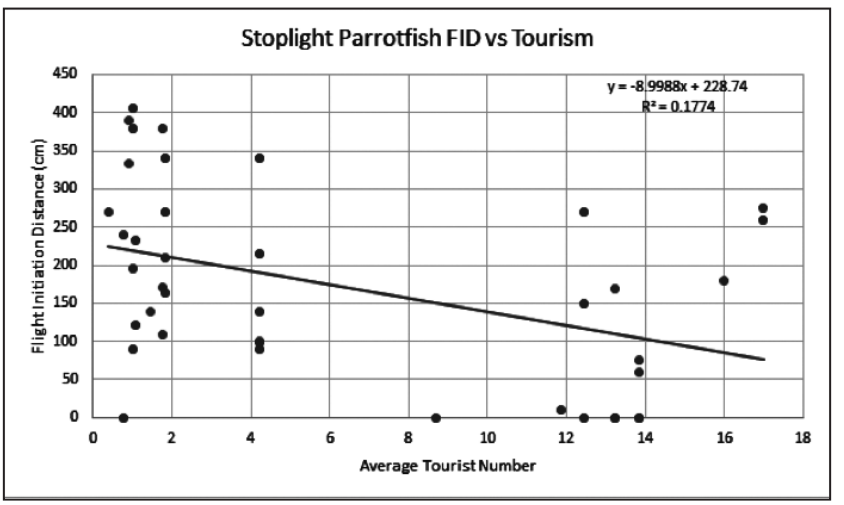
Tourist Density

Each observed species showed a negative relationship between FID and average tourist number, such that FID was higher in areas with low tourist counts and lower in areas with high tourist counts (Figure 1). We looked at effect sizes, which tell us the difference between the groups we were testing, for example FID in high tourist areas vs. low tourist areas. This is determined by the slope, the larger the slope the larger the effect size. We also looked at R^2 values, which look at how much variability is explained by our model, and how close our data fits to the regression line. The effect size and R^2 values were small across all species, which shows that our method to study fish behavior is variable. The highest to lowest R^2 values were *S. viride, A. bahianus*, *A. coeruleus*, and *H. sciurus*, respectively. The same pattern was observed for the effect size that was measured by the slope of the trendline. A significant relationship was acknowledged for data with p-value lower than alpha = 0.05.

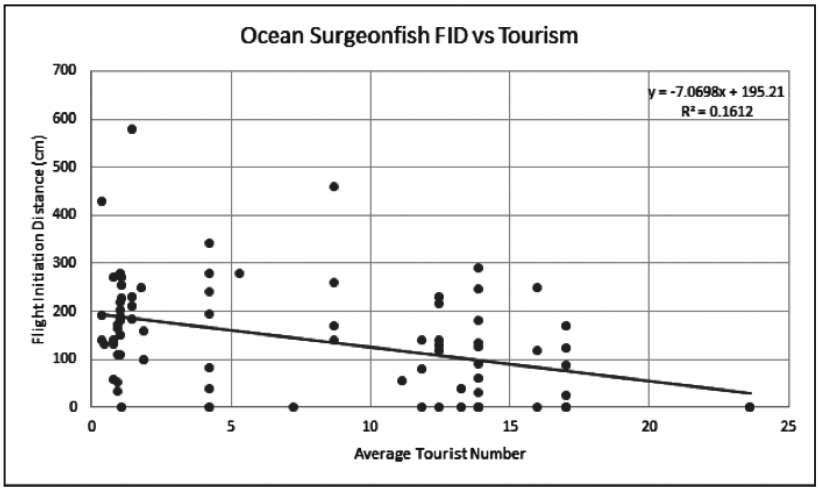
Linear regression analysis indicates that all species except *H. sciurus* showed significant results ($p < 0.05$) (Table 1). *S. viride* had the strongest relationship (r=0.1774, p=0.0068) followed in increasing order by *A. bahianus* (p=0.0309), *A. coeruleus* (p=0.0422), and *H. sciurus* (p=0.0619).

Figure 1: Relationship between FID and average tourist number. The analysis was separated by species in individual graphs: a) *S. viride*, b) *A. bahianus*, c) *A. coeruleus*, d) *H. sciurus*.

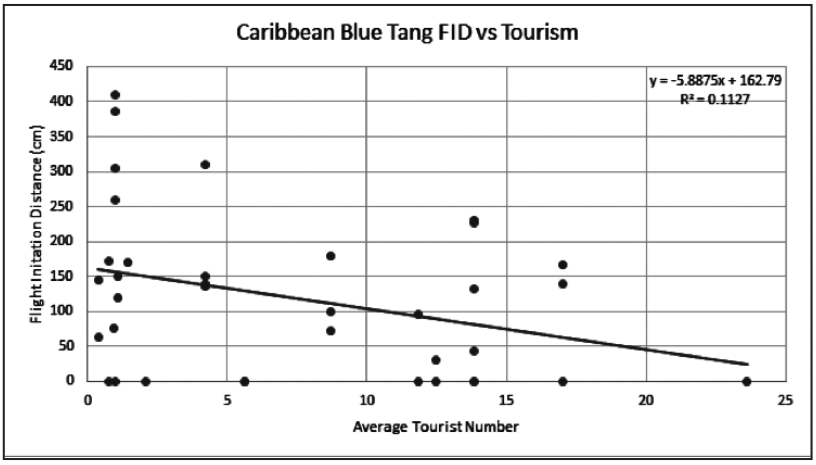
a.



b.


Ocean Surgeonfish FID vs Tourism
y = -7.0698x + 195.21
R² = 0.1612
Flight Initiation Distance (cm)
Average Tourist Number
0
100
200
300
400
500
600
700
0
5
10
15
20
25

c.


Caribbean Blue Tang FID vs Tourism
y = -5.8875x + 162.79
R² = 0.1127
Flight Initiation Distance (cm)
Average Tourist Number
0
50
100
150
200
250
300
350
400
450
0
5
10
15
20
25

d.

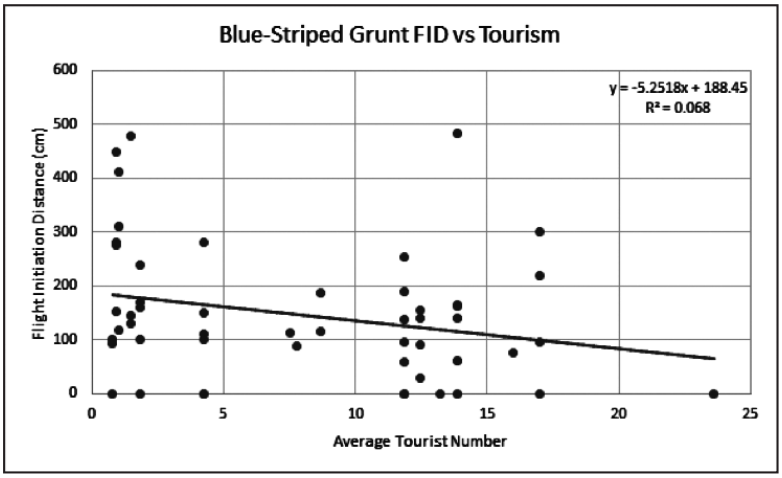

Blue-Striped Grunt FID vs Tourism
y = -5.2518x + 188.45
R² = 0.068
Flight Initiation Distance (cm)
Average Tourist Number
0
100
200
300
400
500
600
0
5
10
15
20
25

Table 1: R^2 and p-values of linear regression analysis of FID with average tourist number reported by species.

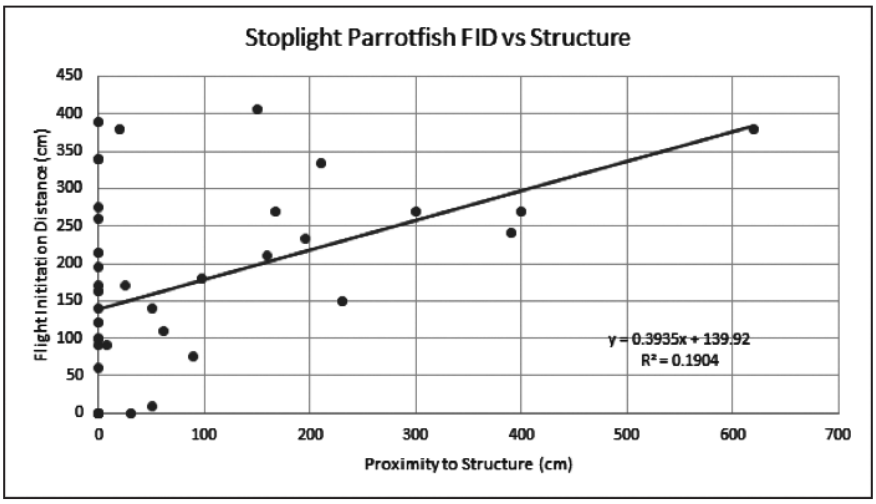
Species	R^2	*P*
S. viride	0.177	0.0068
A. bahianus	0.161	0.0309
A. coeruleus	0.113	0.0422
H. sciurus	0.068	0.0619

Structural Proximity

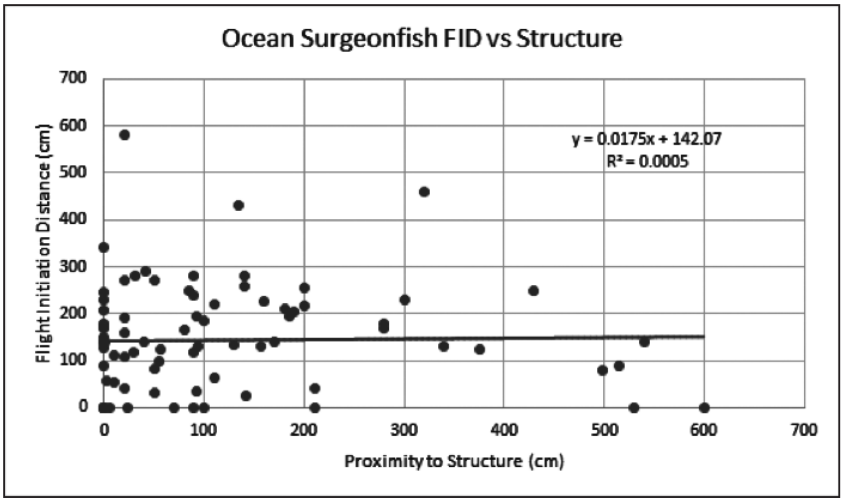
Overall, we observed a positive trend when comparing FID to structural proximity across species, such that a higher FID was associated with a larger distance from structure (Figure 2). *S. viride* had the largest effect size, or largest difference between the two groups measured (slope = 0.3935) and R^2 value at 0.1904, while *A. bahianus* had the lowest effect size (slope = 0.0175) and R^2 value at 0.0005 (Table 2). There also seems to be a difference in between herbivores and non-herbivores in the distance they traveled from structure. The herbivorous species of prey fish (*S. viride, A. bahianus*, and *Acanthurus coeruleus*) were observed to travel the farthest distance from structure at approximately 600 cm, whereas the non-herbivorous *H. sciurus* was observed stay within closer proximity to structure at around 450 cm.

Figure 2: Relationship between FID and proximity to structure, either coral, rocks, or man-made structure. The analysis was separated by species in individual graphs: a) *S. viride*, b) *A. bahianus*, c) *A. coeruleus*, and d) *H. sciurus*.

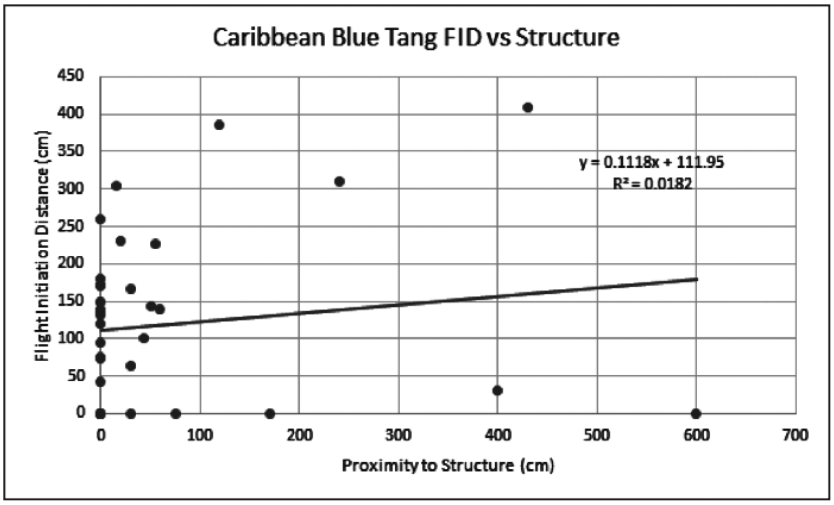
a.



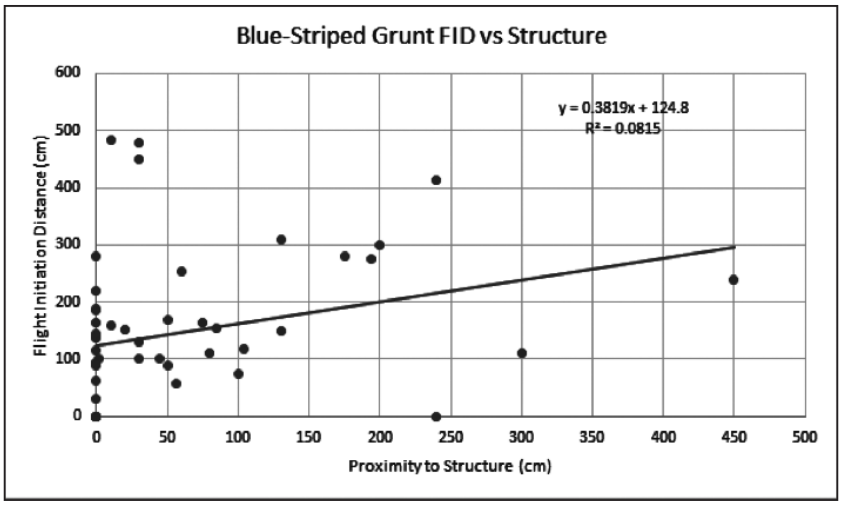
b.


Ocean Surgeonfish FID vs Structure
y = 0.0175x + 142.07
R² = 0.0005
Flight Initiation Distance (cm)
Proximity to Structure (cm)
0
100
200
300
400
500
600
700

c.


Caribbean Blue Tang FID vs Structure
y = 0.1118x + 111.95
R² = 0.0182
Flight Initiation Distance (cm)
Proximity to Structure (cm)
0
50
100
150
200
250
300
350
400
450
500
600
700

d.


Blue-Striped Grunt FID vs Structure
y = 0.3819x + 124.8
R² = 0.0815
Flight Initiation Distance (cm)
Proximity to Structure (cm)
0
50
100
150
200
250
300
350
400
450
500
600

A linear regression testing the relationship between FID and proximity to structure revealed that only *S. viride* and *H. sciurus* showed significant results (Table 2). *S. viride* showed the strongest relationship between FID and proximity to structure ($R^2 = 0.0049$) followed in increasing order by *H. sciurus* ($R^2 = 0.0403$), *A. coeruleus* ($R^2 = 0.4263$), and *A. bahianus* ($R^2 = 0.7268$).

Table 2: R^2 and p-values of linear regression analysis of FID with proximity to structure (in cm) reported by species.

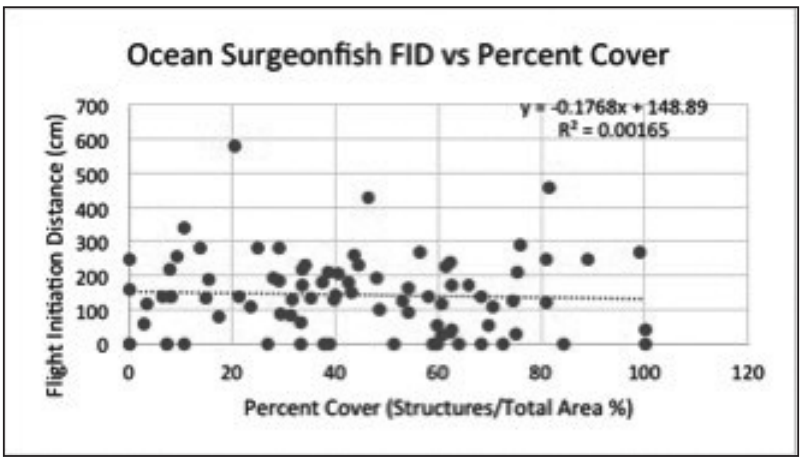
Species	R^2	*P*
S. viride	0.190	0.0049
A. bahianus	0.001	0.7268
A. coeruleus	0.018	0.4263
H. sciurus	0.081	0.0403

Percent Cover

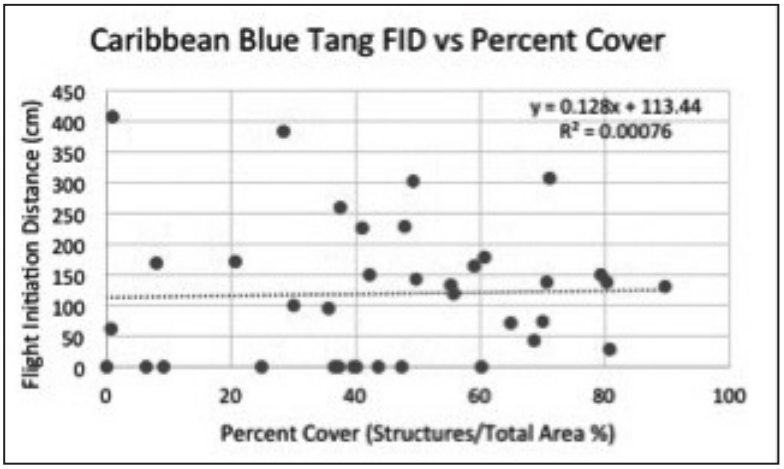
No significant relationship between percent cover of surrounding coral or seagrass habitat and FID was observed. Generally, this relationship appeared to be noisy. *S. veride* had the highest R^2 value at 0.0292, which is still very low. In all species except *A. coeruleus*, FID slightly decreased with an increase in percent cover of surrounding structural habitat within 10 meters, but there was no to little overall relationship (Figure 3). A linear regression analysis for each species for FID against percent cover found no significant relationship (Table 3).

Figure 3: Relationship between percent coverage of 10 meter radius of surrounding area and FID. The analysis was separated by species in individual graphs: a) *S. viride*, b) *A. bahianus*, c) *A. coeruleus*, and d) *H. sciurus*.

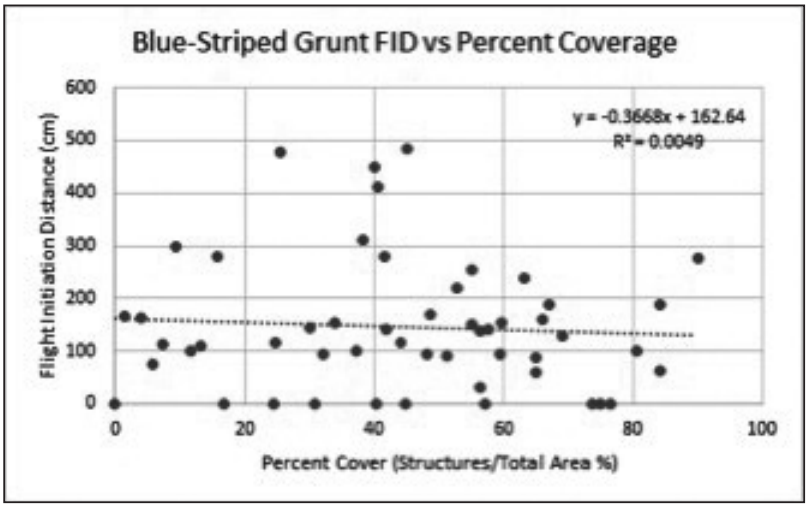
a.



b.


Caribbean Blue Tang FID vs Percent Cover
y = 0.128x + 113.44
R² = 0.00076
Flight Initiation Distance (cm)
Percent Cover (Structures/Total Area %)
0
50
100
150
200
250
300
350
400
450
0
20
40
60
80
100

c.


Blue-Striped Grunt FID vs Percent Coverage
y = -0.3668x + 162.64
R² = 0.0049
Flight Initiation Distance (cm)
Percent Cover (Structures/Total Area %)
0
100
200
300
400
500
600
0
20
40
60
80
100

d.

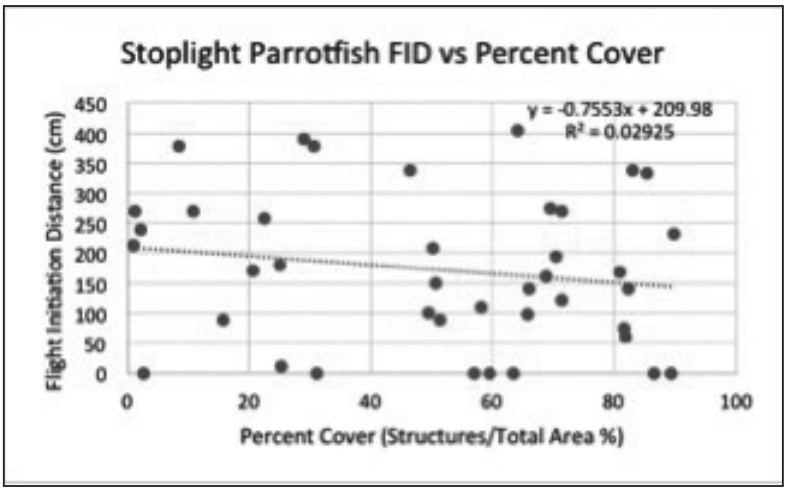

Stoplight Parrotfish FID vs Percent Cover
y = -0.7553x + 209.98
R² = 0.02925
Flight Initiation Distance (cm)
Percent Cover (Structures/Total Area %)
0
50
100
150
200
250
300
350
400
450
0
20
40
60
80
100

Table 3: R^2 and p-values of linear regression analysis of FID with percent coverage reported by species.

Species	R^2	*P*
S. viride	0.02924	0.62303
A. bahianus	0.00165	0.71198
A. coeruleus	0.00075	0.87148
H. sciurus	0.00486	0.62303

Habitat Type

Habitat type was recorded for every FID point and determined to be coral, seagrass, or sand. For statistical purposes, sand and seagrass habitats were grouped together due to the low instances of finding a target fish in a seagrass habitat. FID was lower in coral habitats than both sand and seagrass habitats combined (Figure 4). The presence of coral reduced FID the most for *H. sciurus*, followed by *S. viride* and *A. coeruleus*, and reduced the FID the least for *A. bahianus*. Student's t-test was used to determine if the FID data for coral/sand habitats were significantly different across species. Only *H. sciurus* showed a significant relationship between habitat type and FID (Student's T-test; p= 0.006412).

Figure 4: The number shown is the average FID for each species of fish in its respective habitat type: coral or sand/seagrass. Average FID of coral habitat population is shown in blue and that of sand/seagrass habitat population is shown in orange. Standard deviation error bars are shown with the average FID for each habitat below. The x-axis is species type, while the y-axis is the average FID in centimeters.

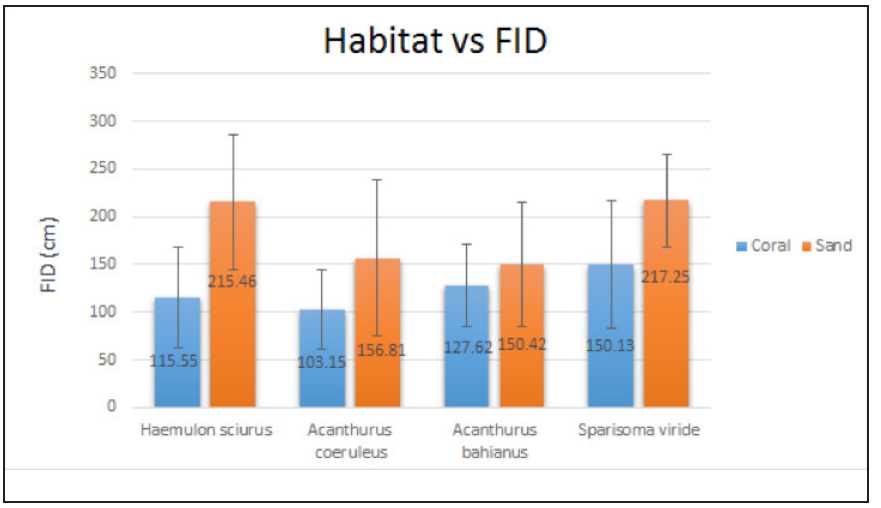


Table 4: Two-tailed p-values of coral habitat against sand habitat from a Student's t-test for each species.

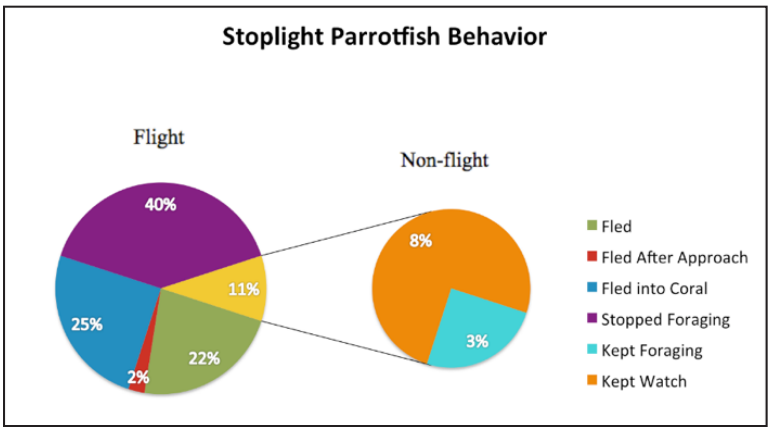
Species	*P (two-tailed)*
H. sciurus	0.006412
A. bahianus	0.373887
A. coeruleus	0.192659
S. viride	0.115689

Non-Flight and Flight Behaviors

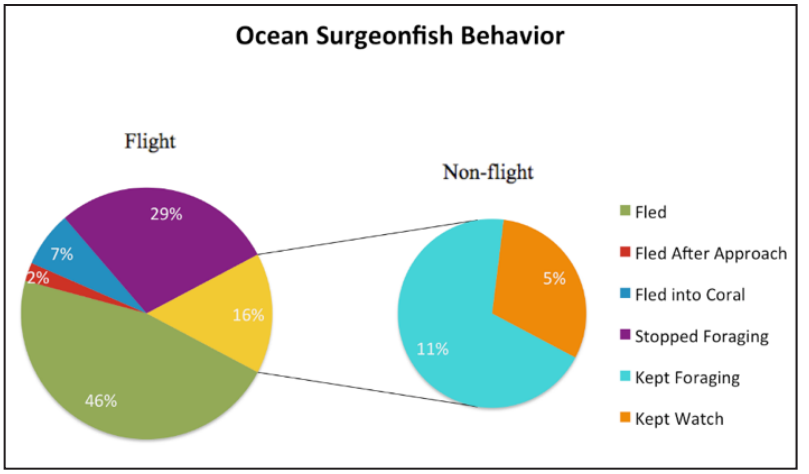
Flight behaviors were divided into the categories listed in the methods. While the fish could exhibit more than one behavior, only the most distinctly observed behavior was recorded. When the specific flight or non-flight behavior of the prey fish was categorized, *S. viride* and *A. coeruleus* were observed to have the largest percentage of flight behavior of fleeing into coral, where they were typically found foraging (Figure 5). *A. bahianus* and *H. sciurus* had very different flight behaviors and were more frequently found swimming away from reefs in sandy habitats (Figure 5). All species but the *S. viride* exhibited fleeing as the most prominent behavior (Figure 5). Non-flight behaviors were near the same percentage for every species. Fleeing into coral was most prominent with *S. viride* and *A. coeruleus*. Species that remained near coral (*S. viride* and *A. coeruleus*) were more likely to flee into the coral or stop foraging. *A. coeruleus* displayed the most frequent non-flight behavior (Figure 5).

Figure 5: Prey fish behavior divided into flight and non-flight behaviors. Flight behaviors are divided into: fled, fled after approach, fled into coral, and stopped foraging. Non-flight behaviors are divided into kept foraging and kept watch. Each species was separated by graph: a) *S. viride*, b) *A. bahianus*, c) *A. coeruleus*, and d) *H. sciurus*.

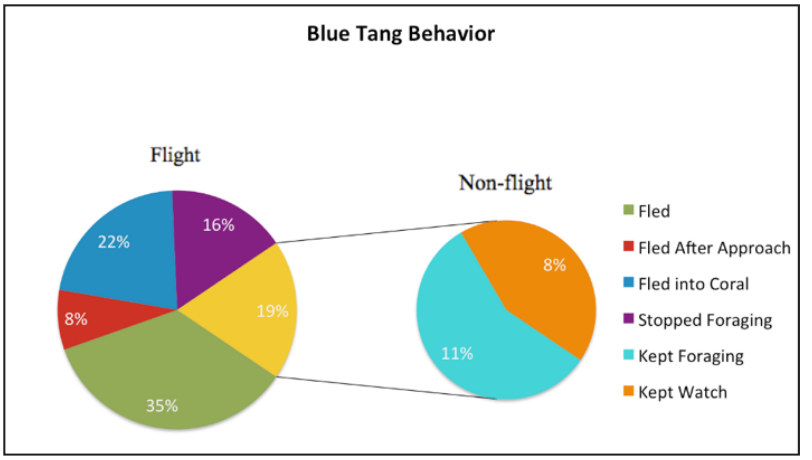
a.



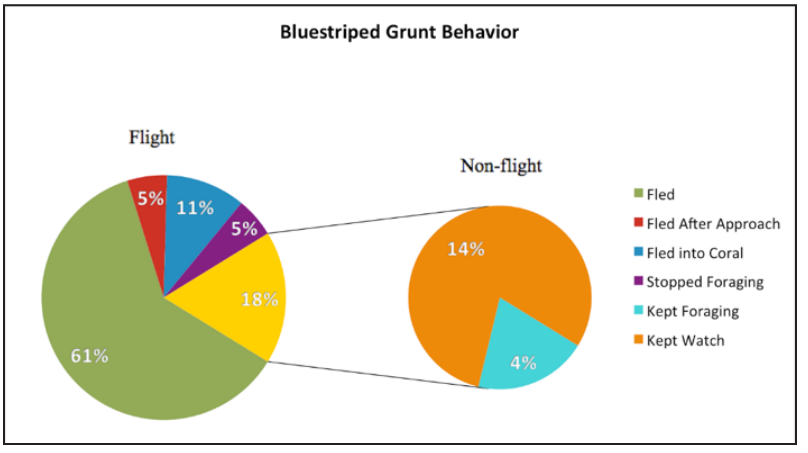
b.


Ocean Surgeonfish Behavior
Flight
Non-flight
29%
7%
2%
16%
46%
5%
11%
Fled
Fled After Approach
Fled into Coral
Stopped Foraging
Kept Foraging
Kept Watch

c.


Blue Tang Behavior
Flight
Non-flight
16%
22%
8%
19%
35%
8%
11%
Fled
Fled After Approach
Fled into Coral
Stopped Foraging
Kept Foraging
Kept Watch

d.


Bluestriped Grunt Behavior
Flight
Non-flight
5%
11%
5%
18%
61%
14%
4%
Fled
Fled After Approach
Fled into Coral
Stopped Foraging
Kept Foraging
Kept Watch

Discussion

Contrary to results seen in previous studies that focused on human fishers as predators, this study found a higher FID in areas with fewer human tourists (Januchowski et al. 2011). This rejects our initial hypothesis that higher FID would be found in areas with higher tourist counts. There are many potential reasons for this unexpected result. First, the fish could have acclimated to the tourists' presence. Over the last 10 years, Akumal Bay has grown in the number of visiting tourists, and fish populations may have already adapted to the change since the tourists are not actually threatening them in the same manner as fishers. A second explanation is that human presence scares off larger predatory species such as barracuda, resulting in a safer environment for prey fish. Although not experimentally tested, a lower density of larger-sized predator fish was observed in the high tourism area compared to the low tourism area. Prey fish would feel safer in the high tourism area, an area that predators avoid, and would perhaps even venture further out to feed. This could alter the availability of prey for the predators in the low tourism area. A follow-up study that addresses this new hypothesis will confirm whether tourists have influences on the normal feeding practices of both prey and predatory fishes.

Many of our FID results exhibited small effect sizes and R^2 values, but they are reasonable as animal behavior in nature is often highly variable. We suspect that tourist counts in Figure 1 may have been less accurate due to a tropical storm on the Pacific side of Mexico, which caused large storms in Akumal over many days of the study. We chose not to omit the data collected during these bad weather days due to limitation on time to collect more data. However, this decision may have led to an unusually low tourist numbers in the high tourism areas. Low tourism areas also showed a decrease in numbers due to bad weather, so the general assumption of high and low tourist areas used in our study was not affected.

Lower FID was found to be associated with lower structural proximity (Figure 2). Only *S. viride* and *H. sciurus* had significant values for the structural proximity experiment, but the effect was in the same direction for all of the species. This supports our hypothesis that FID would decrease as the fish moves further away from the safety of a structure, suggesting a lower fear response towards the perceived predator when the target fish was nearer to structure. Structures give prey fish an area to flee to or hide in, offering protection from larger predators. This lowers the risk of foraging for the prey and would make the fish feel safer. Prey species are more likely to forage longer near a structure even if predators are present nearby. Although a range in distances from the coral was observed, there were some trends in some species. *S. viride* were typically nearer to the coral and therefore more bold due to their proximity. *A. bahianus* exhibits a unique behavior by not following the trend of a lower FID with a closer proximity to a structure, indicating a lack of fear towards tourists. This is also depicted in Figure 4, where *A. bahianus* exhibited the smallest difference in average FID between sand and coral habitats.

Percent cover of coral or seagrass within 10 meters of each FID site turned out to be an ineffective way of relating coral or seagrass coverage to FID. A larger data set could have demonstrated significant values because FID did decrease slightly with increasing percent cover, which supports our hypothesis that there would be a lower FID with higher structural complexity (Table 3 and Figure 3). One possible problem was that Google Earth photographs of Akumal Bay were from 2010 and were outdated by 5 years. Even though the coral structure has probably stayed relatively stable, the seagrass beds may have changed drastically. Another problem was light reflection in some areas of the bay that obscured the shading of the darker coral areas on Google Earth photographs. Furthermore, the Garmin GPS used was a slightly older model and was not always precise to the nearest meter. Overall, structure proximity was a better variable for measuring structural complexity because it was directly observed. If there is truly no significant trend between percent cover and FID, that could suggest that prey reef fish rely more on a close proximity to structure than to a complex structural habitat such as a seagrass bed or reef.

Group size and habitat type had very little to no correlation to FID. When considering habitat type, *S. veride* and *H. sciurus* were primarily seen on or around coral habitats, indicating a closer relationship between these two species and coral. Both group size and habitat could have been limited methodologically due to inconsistencies between observers. It was difficult to count fish within a 2 meter distance that was not measured at the time of taking FID. As the observer, we could have unknowingly selected target fish that were not present in groups, which would greatly affect the results for that variable. In relation to habitat type, there could have been discrepancies for what was considered sand and coral. Areas with some coral cover in the middle of a large sandy area could have been labeled differently by each observer. In future work, a more uniform and less arbitrary method for recording these variables would be implemented.

Structural proximity and habitat type within the coral reefs showed a relationship to the behavior displayed at flight or non-flight. Species that were found more often in the sand over coral demonstrated a lower percentage overall that did not flee and a much lower percentage that fled into the holes, except for *S. viride* that remained near the coral and fled into it more often (Figure 5). *A. coeruleus* had lower FID closer to the coral, and they displayed the highest non-flight behavior, so, perhaps due to their proximity to coral, they were less wary (Figure 2). When the different species' flight and non-flight behavior were examined in terms of types of fish, we found that non-herbivorous *H. sciurus* exhibited the behavior "stop foraging" only 5% of the time, while the herbivorous fishes exhibited that behavior 16 to 40% of the time (Figure 5).

Overall, the herbivorous fish species seemed to show similar behavior to each other compared to the non-herbivorous fish. The three herbivorous fishes all showed a significant relationship between tourist count and FID, while *H. sciurus* did not (Table 1). However, *H. sciurus* was the only species that showed a significant difference in the average FID between

sand and coral habitats ($p = 0.0064$) (Table 4). The behavioral differences seen between the herbivorous target species and the *H. sciurus* could potentially be due to their feeding patterns. The herbivorous fishes we observed feed mainly on algae, while *H. sciurus* feeds on small crustaceans and bivalves and therefore would naturally spend less time foraging and more time looking for prey. Different lifestyles would result in different behaviors, possibly resulting in the differences in the results for this species.

Conclusion

The health of an ecosystem is measured by species diversity, primary productivity, and other physical, chemical, and biological factors. Healthy ecosystems are incredibly important in terms of various economies that they support, whether they are fisheries or ecotourism destinations. Behavioral responses can have an impact on the dynamics and success of the coral reef. Tourists seem to be increasing FID, contrary to what was initially hypothesized, which indicates that perhaps prey fish feel safer in high tourism areas. As the wariness of prey fish decreases, they venture out further from their shelters in order to feed and graze on the algae. Increase in algae coverage due to reduction in algal grazing can result in further loss of corals as algae outcompete corals (Madin et al. 2010; Januchowski et al. 2014). Therefore, decreased wariness of prey fish could possibly mean better foraging and a healthier coral reef system.

On the other hand, tourists could be altering the trophic hierarchy of the reef by decreasing the amount of predator density and creating a downward shift. This combined with other factors, such as the influx of nutrients into the bay due to increased development around Akumal Bay, could shift the ecosystem of the bay to something completely different than what is was before (Estes et. al 2011, Lacey et. al 2013). While this seems as if it would be a negative shift, whether or not the effect is negative or positive for the reef is inconclusive and require future follow-up research. This research could be continued in many directions: predatory fish behavior in response to tourism, algae growth due to possible lowered herbivorous fish density, and even specific behavioral trends of individual species that showed unique patterns (*H. sciurus* and *A. bahianus*). Further studies and data analyses will provide further unique insights into the relationship between tourism and reef fish behavior, of both prey and predatory species.

Literature Cited

Bellwood DR, Hughes TP, Folke C, Nystrom M (2004) Confronting the coral reef crisis. Nature 429:827-833

Bellwood, D. R., Hoey, A. S., Hughes, T. P. (2011). Human activity selectively impacts the ecosystem roles of parrotfishes on coral reefs. Proc. R. Soc. B. 282, 1-9.

Chabanet, P., Adjeroud, M., Andrefouet, S., Bozec, Y.M., Ferraris, J., Garcia-Charton, J.A., Estes, J.A., Terborgh, J, Brashares, J.S., Power, M.E., Berger, J, Bond, W.J., Carpenter, S.R., Essington, T.E., Holt, R.D., Jackson, J.B.C., Marquis, R.J., Oksanen, L, Oksanen, T, Paine, R.T., Pikitch, E.K., Ripple, W.J., Sandin, S.A., Scheffer, M, Schoener, T.W., Shurin, J.B., Sinclair, A.R.E, Soule, M.E., Virtanen, R, and Wardle, D.A. (2011). Trophic Downgrading of Planet Earth. Science 333 (6040), 301-306.

Schrimm (2005). Human-induced physical disturbances and their indicators on coral reef habitats: a multi scale approach. Aquat. Living Resour. 18, 215-230.

Garcia-Salgado M, Nava-Martinez G, Bood N, McField Mand others (2008) Status of coral reefs in the Mesoamerican region. Status of Coral Reefs of the World: 2008:253- 264

Harvell D, Jordan-Dahlgren E, Merkel S, Rosenberg Eand others (2007) Coral diseases, environment drivers, and the balance between coral and microbial associates. Oceanography 20:172-195

Januchowski-Hartley, F.A., Graham, N.A.J., Feary, D.A., Morove, T., Cinner, J.E. (2011). Fear of Fishers: Human predation explains behavioral changes in coral reef fishes. PLoS ONE. 6, 1-9.

Januchowski-Hartley, F.A., Graham, N.A.J., Cinner, J.E., Russ, G.R. (2014). Local fishing influences coral reef fish behavior inside protected areas of the Indo-Pacific. Bio. Cons. 182, 8-12.

Kramer, P. R. and Kramer, P.A. (2002). Ecoregional Conservation Planning for the Mesoamerican Caribbean Reef (MACR). World Wildlife Fund. 1-147.

Kulbicki, M. (1998). How the acquired behaviour of commercial reef fishes may influence the results obtained from visual censuses. Journal of Experimental Marine Biology and Ecology. 222, 11-30.

Lacey E.A, Fourqurean J.W., Collado-Vides L. 2013. Increased algal dominance despite presence of *diadema antillarum* populations on a caribbean coral reef. Bulletin of Marine Science. 89, 603-620.

Madin, E. P., Gaines, S. D., Warner, R. R. (2010). Field evidence for pervasive indirect effects of fishing on prey foraging behavior. Ecology. 91, 3563-3571.

Nunes, J.D.A.C.C., Sampaio, C.L.S. & Barros, F., 2015. The influence of structural complexity and reef habitat types on flight initiation distance and escape behaviors in labrid fishes. *Marine Biology*, 162, pp.493–499.

Preisser, E.L., Bolnick, D.I., Benard, M.F. (2005). Scared to Death? The Effects of Intimidation and Consumption in Predator-Prey Interactions. Ecology. 86, 501-509.

Risk, M. J. (1972). Fish diversity on a coral reef in the Virgin Islands. Atoll Res. Bull.153, 1–7.

SECTUR (2014). Compendio Estadístico del Turismo en México 2014. Datatur. http://www.datatur.sectur.gob.mx/SitePages/CompendioEstadistico.aspx

People Broken Into Pieces Trying to Join: Byzantine Erotica and the Provocative Paradox

The University of Texas at Austin

Undergraduate Research Journal

Volume 15
Number 1
Spring 2016

Kendall DeBoer

In the first thirty-six pages of *Infinite Jest*, David Foster Wallace refers to "Byzantine erotica" twice: Hal's professional conversationalist (his father) asks him if he would like to discuss Byzantine erotica, and there is "a triptych of high-quality Byzantine erotica" in the medical attaché's apartment.[1] Additionally, Hal has "an old print of a detail from the minor and soft-core Alexandrian mosaic *Consummation of the Levirates*" in his room at Enfield Tennis Academy.[2] These conceptually-absurd-if-not-altogether-impossible "racy mosaics" persist throughout the novel.[3] Like many of Wallace's unexpected details, Byzantine

erotica has a purpose beyond absurdity. Byzantine erotica reflects the constant thematic paradoxes throughout *Infinite Jest*, and thereby provides a frame of reference to which one can compare and thereby better understand the formal and thematic qualities of the novel. Byzantine art is that which relates to Constantine I's "New Rome" in the East, founded in 324 and conquered by Ottoman Turks in 1453.[4] Byzantine art's rich mosaic ornamentation is one of its defining characteristics (see fig. 1). These mosaics usually depict static scenes. Figures maintain calm, regular postures and inhabit their spaces in a "uniformed arrangement" to bolster "compositional equilibrium" and symmetry.[5] Strict regulations demand a "hierarchy of precedence" and an obedience to the laws of decoration, as Byzantine art's purpose is to embody the "triumphant Church's spiritual and didactic demands" through visuals.[6] Orthodox Christian doctrine and Byzantium's imperial politics become manifest in these formulaic works with "rigid formality."[7] Balance, order, logic, power, discipline, spirituality—these concepts define Byzantine art.

Erotic art's definition is less specific in that it has no association with a particular time period or region. Whether a piece is ancient, contemporary, eastern, or western is irrelevant. Beyond its defiance of temporal, spatial, or cultural associations, erotic art also refuses to privilege any specific form. The genre does not privilege paintings, sculptures, woodblocks, collages, etc. Erotic art has no steadfast regulations. The only requirement (if this strict word can apply to such a liberal genre) of erotic art is that an erotic work must involve love, sexuality, or desire (see fig. 2). Accordingly, erotic art is often dynamic, passionate, and imbalanced as it tries to emulate sexual experiences. Sexual themes may provide "social and therapeutic functions" or provide an outlet for expression.[8] Thus, free from constraints, erotic art is often expressionist.

When Wallace references his fabricated Byzantine erotica in *Infinite Jest*, he references a concept that is inherently self-contradictory. In fact, if one were to look through works from every movement and genre of art history, Byzantine art would be one of the only movements devoid of direct erotica. In most surveys of erotic art, scholars acknowledge that before Victorian period, European erotic art was very rare. Scholars typically omit Byzantine art altogether in the context of erotica. The Kronhausens' survey typifies this tendency:

> The West has no similar tradition [like that of Eastern erotic art], at least not on such a large scale. One might say that at certain periods, such as the Renaissance in Italy or the eighteenth century in France, erotic art was appreciated by a portion of the upper class as part of a gracious style of living. But this appreciation by a small group of wealthy and well-educated art lovers in Europe can not really be equated with the widespread acceptance of erotica in the East.[9]

Though Byzantine art is the art of Eastern Rome, the Kronhausens use the terms 'West' and 'East' to refer to Europe and Asia, so this passage indicates the dearth of European erotic art. The Kronhausens argue that the suppression of erotic art rose with the "morbid antisexualism of the Apostle Paul" and other early Church leaders "who never tired of

Figure 1: *Justinian, Bishop Maximianus, and attendants*. Mosaic from the north wall of the apse in San Vitale, Ravenna, Italy, ca. 547.
Notice the rigid frontality of the figures, the balance and symmetry of the composition, the symbols of religious-political authority, and the intricate details of the mosaic.

Figure 2: "Primitive, almost animalistic couple" by Karel Appel, ca. 1960.
Scan from Kronhausen's *Erotic Art: A Survey of Erotic Fact and Fancy in the Fine Arts* (57).
Notice the motion of the figures, the asymmetry of the composition, the simplicity of content, and the expressive carnal/material elements of the canvas.

exhorting their followers against the temptations of the flesh."[10] This antisexual tradition developed alongside the rise of a specific type of Christianity upheld by the state. Thus, Orthodox Christianity ensured that the Middle Ages "produced almost no erotic art," except for the indirect eroticism of Madonnas and female saints in religious ecstasy, and even those depictions were rare.[11] The Fathers of the Church denounced the "orgiastic eroticism" of Roman cults.[12] The religion and culture of Byzantium wholly refute the idea of erotic art.

Further, the two artistic genres oppose one another in every way in terms of content and form. As established, Byzantine art upholds a specific political doctrine of authoritarianism; erotic art flouts any sort of political associations and often ascribes to an anarchistic free-love ideology. Where Byzantine art is rigid, symmetrical, and frontal, erotic art is expressionistic, asymmetrical, and lacking a clear perspective. Byzantine art has delineated rules and functions while erotic art exists to defy those ideas.

But this opposition is meaningful, and feels appropriate in the context of the other oppositions in the novel. Wallace employs undecidable paradoxes throughout *Infinite Jest*, such as in his example of a quiz from Enfield Tennis Academy's Ms. Thode's course: "you are pathologically driven to go out and steal, steal, steal," but "you are pathologically driven to not ever leave home [and] you live alone [...] Meaning, you must go out, into the marketplace to satisfy your overwhelming compulsion to steal, steal, steal."[13] Ms. Thode refers to this conundrum as a Double-Bind. The coexistence of these two qualities seems impossible, but Wallace places them together. A broader paradox Wallace returns to through the novel is the coalescence of good and bad, pleasure and pain. As drug addicts recover, "the way it gets better and you get better is through pain."[14] Wallace conflates and combines temporal perceptions too. During Gately and Lenz's fight, "It's always that everything always speeds up and slows down both."[15] And the major problem with psychic depression, Wallace posits through the thoughts of Kate Gompert, is that it is "a sort of double bind in which any/all of the alternatives we associate with human agency--sitting or standing, doing or resting, speaking or keeping silent, living or dying--are not just unpleasant but literally horrible."[16] An erotic Byzantine mosaic does not seem so different from an agoraphobic kleptomaniac, a betterment through hurt, a speeding up and slowing down, a distaste for both poles of a binary, in that these examples all present paradoxical and inconclusive cases. The paradoxes of Byzantine erotica highlight the major themes of coinciding, conflicting dualities that shape *Infinite Jest*.

Consider bodies. In Byzantine mosaics, "the organic body has dematerialized," rendering the featured humans "solemn spirits" who do not have "solid bodies."[17] So, Byzantine mosaics downplay and even negate the body in favor of abstract qualities such as spirituality and obedience. In erotica, the body's importance is clear. Fleshly, sexual acts define the genre. The body feels pleasure, which takes precedence over other ideals. Byzantine erotica, then, negates and celebrates the body, offering a fluid, multivalent mode that defies strict binary categorization.

The tennis philosophies of Head Coach and Athletic Director of Enfield Tennis Academy, Schtitt, and Hal's father, J.O.I. Sr., parallel the body paradox introduced by Byzantine erotica. Schtitt believes tennis is about "learning to sacrifice the hot narrow imperatives of the Self—the needs, the desires, the fears, the multiform cravings of the individual appetitive will—to the larger imperatives of the team (OK, the State) and a set of delimiting rules (OK, the Law)."[18] This quotation makes the sacrifice of the Self sound like the sacrifice of one's physical (bodily) wants for the sake of some abstract thing greater than the body. Playing tennis well requires discipline and control of the body to achieve something larger, outside of itself. The team and the rules (or the State and the Law) take precedence over the physicality of the body. Schtitt's beliefs regarding negation of the physical self for the sake of a more important concept continues throughout *Infinite Jest* in varying forms, as drug addicts deny their bodies' wishes for a moral or spiritual aim of sobriety, or as Gately denies pain-killers in the hospital so he can abide. Byzantine mosaics similarly favor regulations and adherence to Orthodox Christianity over depictions of the body.

Conversely, J.O.I. Sr. praises bodies and physicality, advising his son to "touch things with consideration," reminding his son he's "a machine a body an object."[19] He idealizes a tennis ball as "the ultimate body" with its "total physicality," "no revving head," and "complete presence."[20] Though by the end of this section he suffers an emotional collapse, J.O.I. Sr. first asserts the superiority of the body over the mental, abstract, spiritual, and the like, just as erotic art puts physicality before all else. The primacy of the body recurs in *Infinite Jest*; The Entertainment is powerful because of its absolute control and mesmerization of the body. The first sentence of the novel is "I am seated in an office, surrounded by heads and bodies," which emphasizes the synecdotal physical parts that comprise humans rather than the humans' interiority.[21] Byzantine erotica's simultaneous encapsulation of body-sacrifice and body-celebration finds its parallel in these philosophies of Schitt and J.O.I. Sr., and these philosophies are indicative of motifs that continue throughout the work.

Though they present themselves as contradictory, there is a constant threat that these ideologies regarding the body will flow into one another. Schitt may uphold tennis as a means for transcending the limitations of the body, but ironically, this transcendence is only available through ritualistic practice of a physical sport.

> You compete with your own limits to transcend the self in imagination and execution. Disappear inside the game: break through limits: transcend: improve: win. Which is why tennis is an essentially tragic enterprise, to improve and grow as a serious junior, with ambitions. You seek to vanquish and transcend the limited self whose limits make the game possible in the first place.[22]

Though transcendence is the aim of the sport, the limited body is necessary to enact the process to achieve this transcendence. And J.O.I. Sr. admits to the failings of the body. He says, "the body betrays you and down you go."[23] Then, he begins berating Hal because he

claims, "We're just bodies and shoulders and scarred knees and big bellies and empty wallets and flasks to you."[24] If bodies are as amazing as J.O.I. Sr. claims earlier in this episode, then it would not be negative that Hal perceives him as a body. Neither a pro-body or an anti-body argument can exist independently. These two theoretically opposing worldviews--the body-as-the-means and the body-as-the-end--interact with one another; they blend. Byzantine erotica illustrates well this intricate and complicated mixture.

Byzantine erotica also anticipates the tensions in *Infinite Jest* between ideologies of authority and freedom. Because Byzantine art is authoritative, it reinforces "dogmatic concepts by means of symbolic representations" through its rigorous adherence to religious-political programs.[25] The Orthodox Christian doctrine asserted that the Byzantine emperor had divine right and could exercise "the ultimate spiritual [and] temporal authority," and the art reflects this doctrine.[26] Conversely, erotic art expresses "the demand for sexual freedom" which is crucial for individual happiness, and which requires a high degree of political and economic freedom.[27] Erotic art is free from authority. Byzantine erotica espouses neither authority nor freedom while also espousing both authority and freedom. This fabricated art form enacts the interplay and collapse of seemingly oppositional ideals, and provokes the reader to consider the complexity and nuance of these slippery ideologies.

Another undecidable paradox arises when Marathe, member of the Wheelchair Assassins and quadruple agent against the AFR, and Steeply, worker for the Office of Unspecified Services, have a conversation in which the two advocate, respectively, authority and freedom. Marathe, criticizing the United States for its citizens, who would choose death by pleasure, states that "Someone or some people among your own history sometime killed your U.S.A. nation already, Hugh. Someone who had authority, or should have had authority and did not exercise authority."[28] Marathe believes guiding authority is important for a nation's well-being. He goes on to point out that sometimes people "forget how to choose," so they need leaders who will use their authority to indicate the proper "temples." Here, because "temples" signifies places of worship, Marathe draws a direct link between authority and spirituality. Alcoholics Anonymous doctrine echoes Marathe's beliefs, because one must surrender to the authority of the program and choose to remain sober:

> ... and then Gately seems to find out AA turns out to be the very loyal friend he thought he'd had and then lost, when you Came In. And so you Hang In and stay sober and straight, and out of sheer hand-burned-on-hot-stove terror you heed the improbable-sounding warnings not to stop pounding out the nightly meetings even after the Substance-cravings have left and you feel like you've got a grip on the thing at last and can now go it alone, you still don't try to go it alone, you heed the improbable warnings because by now you have no faith in your own sense of what's really improbably and what isn't [...] and when people with AA time strongly advise you to keep coming you nod robotically and keep coming [...] and you keep getting ritually down on your big knees every morning and night asking for help from a sky that still seems a burnished shield against all who would ask aid of it--how can you pray to a 'God' you believe only morons believe in, still?--but the old guys say it doesn't yet matter what you believe or don't believe, Just

> Do It they say, and like a shock-trained organism without any kind of human will you do exactly like you're told, you keep coming and coming, nightly...[29]

In the AA program, the doctrine dictates participants' choices. These choices have a certain religiosity about them, which Gately describes as "Blind Faith."[30] Again, Wallace couples authoritative doctrine with spirituality. Likewise, Byzantine art enforces the authority of the emperor who will lead his citizens to Christian salvation. Marathe's support for a guiding figure (i.e. the State) to lead citizens in choosing their temple and AA's authoritative abstinence-and-spirituality mirror Byzantine art's representation of emperors who rule all aspects of religious-political life.

Against Marathe, Steeply argues for personal freedom, arguing, "one cannot be human without freedom."[31] He contends that neglecting one's own wishes and desires for the State and following "instructions on choosing" are fascist, shudder-inducing ideas. Steeply elaborates later, stating, "We want choice. A sense of efficaciousness and choice. To be loved by someone. To freely love who you happen to love," which Marathe summarizes as "utilitarienne. Maximize pleasure, minimize displeasure: result: what is good. This is the U.S.A. of you."[32] Steeply's argument for freedom extends into other more hedonistic themes in *Infinite Jest*. The motif of entertainment throughout plays with the idea of freedom to indulge. Hal, contemplating television, notes that "viewers had been conditioned [...] to associate the Freedom to Choose and the Right to Be Entertained with all that was the U.S. and true."[33] Wallace iterates again and again that a distinctly American view upholds the liberty to partake in pleasure. In supporting freedom, Steeply and others exemplify the ideals of erotic art. Steeply and erotic art declare that one should have the right as an individual to choose to pursue happiness and pleasure. Therefore, Byzantine erotica's paradoxical promotion of both obedience to authority and freedom to pursue one's pleasure foreshadows the continuous argument between Marathe and Steeply. Just as Byzantine erotica mixes the authority of Byzantium with the freedom of sexual expression without preferring either ideology, Marathe and Steeply never arrive at a decisive conclusion.

Again, these presented dichotomies are blurry, not categorical. The authority promoted by Marathe ultimately folds when Steeply provokes him to admit "Canadians also, we would choose dying for this, the total pleasure of a passive goat," demonstrating that Canadians also do not choose the right temples.[34] The point of the doctrinal AA program is that in spite of all the rules and order, by the end "you're free"; so sacrificing personal choices ultimately leads to freedom.[35] Conversely, indulging in personal choices and freedom can also lead to enslavement or loss of agency, as illustrated by the myth of the woman covered in long blonde hair who is "too exotic and intriguing or seductive to resist," but "the pleasure's too intense," leading to death: "even the ones who know the pleasure of it will kill them, they go ahead anyway."[36] In indulging and watching the Entertainment, one must view it "again and again" forever; people continue to watch even at the cost of "one digit from the

Subject's extremities" per reviewing or even when they are aware of its fatal consequences.[37] Authority and freedom are not separate and distinct in *Infinite Jest*; they play off one another and refuse to settle into static ideas. Byzantine erotica provides a beautiful template for this notion in its simultaneous embodiment of authoritarianism and libertarianism, spirituality and hedonism.

Reading *Infinite Jest* can be frustrating because Wallace's prose and ideas feel sometimes disparate. English-major pedagogy calls for finding an overarching theme, or a novel's thesis—and the faster one can identify these key concepts, the better the reading experience should be, theoretically. In *Infinite Jest*, Wallace leads the reader towards one thesis-like conclusion, but then negates that conclusion in a vignette or a scene pages later. Wallace presents opposing themes throughout, such as transcendental Schtitt v. material J.O.I. Sr. and authoritarian Marathe v. libertarian Steeply, and refuses to present a clear, consistent ideology. Every scene, character, moment has the complexity of paradoxical Byzantine erotica. But despite the fragmentation of hypothetical Byzantine erotica in form and content, all the shards of colorful glass and all the conflicting themes provide a cohesive, engaging, and innovative work of art. Byzantine erotica is provocative because it allows perceived polar binaries to coexist in one space without favoring a singular interpretation or mode. *Infinite Jest* as whole is the same way; all the fragmentary bits of narrative and all the paradoxical thesis-moments come together, resulting in an artistic experience that challenges traditional practices of deciding and settling on a novel's central point. *Infinite Jest* invites inquiry and thought about immense and urgent questions regarding the human condition, but does not provide a definitive solution or answer. Because human experience is nuanced, *Infinite Jest* is nuanced. Wallace's invention and repeated invocation of Byzantine erotica reminds the reader of the complexity central to the novel.

In this sense, *Infinite Jest* proffers multivalence as a mode of thought well-equipped for grappling with abstract philosophies and ideologies. Wallace rejects the clean precision of dualities by illustrating their coexistence. Wallace's incorporation of Byzantine erotica and other undecidable oppositions without any distinct denouement emerging asserts the importance of complexity. Byzantine erotica acts a microcosm of a bigger post-modern ideology advanced throughout *Infinite Jest*: simplistic, black-and-white binaries do not account for instances in which opposing poles collapse into one another. The binaries of spirit v. body and authority v. liberty do not account for Byzantine erotica, because it inhabits both positions even as it negates both positions. Through this lens, *Infinite Jest* is an artistic deconstruction of hierarchical notions of binary divisions. Instead, Wallace presents us with provocative, complex, and indecipherable Byzantine erotica. *Infinite Jest* pushes its readers to continually reform, revise, and extend interpretations of knowledge and meaning, and how those interpretations are categorized.

Focusing solely on Byzantine erotica, one can apprehend aspects on the seemingly self-contradictory nature of *Infinite Jest*. But one can also apprehend that self-contradiction

does not negate the work of art. Wallace's work suggests that art does not need a clear, singular thesis. Provocative paradox can be the point. In fact, the contradictory nature of Byzantine erotica and *Infinite Jest* can motivate more nuanced considerations of traditional dualistic worldviews. Wallace's insertion of Byzantine erotica in *Infinite Jest* anticipates opposing ideologies throughout the novel, and demonstrates that opposing ideologies combine to create a new, innovative artistic experience. Like the sharp fragments of a Byzantine mosaic, Wallace's oscillating and self-contradicting themes join together, resulting in a moving and provocative work of art.

Bibliography

Bayet, Charles. *Byzantine Art*. Translated by Anne Haugen and Jessica Wagner. New York: Parkstone, 2009.

Brendel, Otto J. "The Scope and Temperament of Erotic Art in the Greco-Roman World." In *Studies in Erotic Art*, edited by Theodore Bowie and Cornelia V. Christenson, 3-108. New York: Basic, 1970.

Chatzidakis, Manolis. *Studies in Byzantine Art and Archaeology*. London: Variorum Reprints, 1972.

Kleiner, Fred S., and Christin J. Mamiya. "Rome in the East: The Art of Byzantium." In *Gardner's Art through the Ages*, 325-56. 12th ed. Belmont: Wadsworth, 2005.

Kronhausen, Phyllis, and Eberhard Kronhausen, comps. E*rotic Art: A Survey of Erotic Fact and Fancy in the Fine Arts*. New York: Bell, 1968.

———, comps. *Erotic Art 2*. New York: Grove P, 1968.

Lucie-Smith, Edward. *Eroticism in Western Art*. London: Thames and Hudson, 1972.

Wallace, David Foster. *Infinite Jest*. New York: Back Bay, 2006.

Endnotes

1 David Foster Wallace, *Infinite Jest* (New York: Back Bay, 2006), 29, 36.
2 Ibid. 171.
3 Ibid. 29.
4 Fred S. Kleiner and Christin J. Mamiya, "Rome in the East: The Art of Byzantium," in *Gardner's Art through the Ages*, 12th ed. (Belmont: Wadsworth, 2005), 325.
5 Charles Bayet, *Byzantine Art*, trans. Anne Haugen and Jessica Wagner (New York: Parkstone, 2009), 38.
6 Manolis Chatzidakis, *Studies in Byzantine Art and Archaeology* (London: Variorum Reprints, 1972), 2, 4.
7 Kleiner and Mamiya, "Rome in the East," in *Gardner's Art through the Ages*, 326, 333, 336.
8 Phyllis Kronhausen and Eberhard Kronhausen, comps., *Erotic Art: A Survey of Erotic Fact and Fancy in the Fine Arts* (New York: Bell, 1968), 8.
9 Phyllis Kronhausen and Eberhard Kronhausen, comps., *Erotic Art 2* (New York: Grove P, 1968), VIII.
10 Ibid. XI.

11 Kronhausen and Kronhausen, *Erotic Art: A Survey*, 90.
12 Edward Lucie-Smith, *Eroticism in Western Art* (London: Thames and Hudson, 1972), 32.
13 Wallace, *Infinite Jest*, 307-308.
14 Ibid. 446.
15 Ibid. 612.
16 Ibid. 696.
17 Kleiner and Mamiya, "Rome in the East," in *Gardner's Art through the Ages*, 336.
18 Wallace, *Infinite Jest*, 82-83.
19 Ibid. 158, 159.
20 Ibid. 160.
21 Ibid. 3.
22 Ibid. 84.
23 Ibid. 167.
24 Ibid. 168.
25 Chatzidakis, *Studies in Byzantine Art and Archaeology*, 5.
26 Kleiner and Mamiya, "Rome in the East," in *Gardner's Art through the Ages*, 326.
27 Kronhausen and Kronhausen, *Erotic Art: A Survey*, 8.
28 Wallace, *Infinite Jest*, 319.
29 Ibid. 350.
30 Ibid. 351.
31 Ibid. 320.
32 Ibid. 423.
33 Ibid. 412.
34 Ibid. 474.
35 Ibid. 351.
36 Ibid. 528.
37 Ibid. 725, 727.

The Legacy of 1830 Land Cover on Present-Day Massachusetts Forest Composition

The University of Texas at Austin

Undergraduate Research Journal

Volume 15
Number 1
Spring 2016

Sofie McComb

Abstract

European settlers in the eastern United States cleared three-quarters of New England's forests for agriculture and intensively logged the rest. Peak agricultural land cover in Massachusetts occurred in approximately 1830, after which farmland abandonment led to a protracted natural reforestation. Concerns regarding deforestation prompted an 1830 statute that required towns in Massachusetts to survey land-cover and create maps depicting cultural, woodland, and agriculture features. Using these maps in conjunction with modern forest inventory data, the potential for a long-term impact of land-use (i.e. forest clearing)

was analyzed. Specifically, sites that were forested in 1830 were compared to sites that were cleared in 1830 to determine if they had a different contemporary species composition. Additionally, the 1830 land-use map was compared to historical agricultural censuses as well as modern land cover maps to attempt to account for and describe uncertainties in the accuracy of the map, events occurring after 1830, and land cover transitions over time. Averaging across the full state scale, there were few differences in tree species composition between sites that were cleared versus forested in 1830 ($A < 0.01$ based on Multi-response Permutation Procedure (MRPP)). Mid-successional eastern white pine (*Pinus strobus*) and red maple (*Acer rubrum)* accounted for half of the basal area in both historically forested and cleared areas. Stratifying and focusing on data-rich parts of the state revealed suggestive compositional and functional trait differences, including a lack of seed persistence, more species with resprout ability, and more shade tolerant eastern hemlock in the historically forested areas as compared to the historically cleared areas. However, the MRPP and Monte Carlo randomization tests revealed few significant patterns. Historical land-use change and disturbance, and the subsequent forest recovery, have led to a homogenized landscape dominated by rapidly growing, seed abundant, climatically tolerant, and generalist tree species. Overall, this analysis suggests that the natural and anthropogenic processes that have been impacting the landscape since 1830 have largely erased the colonial land-use legacy on modern forest composition.

Introduction

European colonization of Massachusetts throughout the 17th and 18th centuries brought extensive deforestation, logging, and burning to New England forests (Foster 1998). The switch to individual ownership of private land as well as a cultural evolution led to market-oriented intensive agriculture and extensive clearing of the land for farms and pastures (Foster et al. 2008). Even areas not cleared for agriculture were harvested intensively, predominately for fuel, but also for tannins, lumber, fences, and charcoal (Foster and O'Keefe 1998). The peak of agricultural clearing in Massachusetts occurred around 1830, when the percent of forested cover was reduced to ~30% (Foster and O'Keefe 1998). After 1830, a shift in economy and culture led to the abandonment of many farmlands and, in turn, to natural reforestation and succession of forests. The modern landscape of Massachusetts is approximately 63% forested (Thompson et al. 2013). However, historic human land-use disturbance left a lasting mark on the New England forests, altering the forest structure, species composition, and landscape dynamics (Forest et al. 2008). The extensive land use and land cover history of the New England region, particularly Massachusetts, allows for an evaluation of the forest composition and dynamics response to an intensive and broad anthropogenic disturbance regime (Foster et al. 1998).

There is a long precedence of using historic land cover maps and inventory data in ecological research. Global historic cropland inventory data sets were used alongside a land

cover change model to determine the change of global land cover to agricultural lands from 1700 to 1992 (Ramankutty and Foley 1999). A variety of historical data sources, including land cover maps, deeds, and sale records, were used to determine the change in land cover and vegetation patterns on a Central Massachusetts land tract from 1730 to 1990 (Foster et al. 1992). Non-geometric cadastral maps from the 17th and 18th centuries alongside 20th century aerial photographs were used to analyze the land cover change and ecosystem transformations in southeast Sweden over a 300-year time scale (Cousins 2001). Historic maps are necessary for studies of an ecosystem's history and are indispensable in systematic studies of vegetation, as they easily portray at a glance what pages of descriptions can fail to explain (Clements 1905). The maps outline the archaeological geography of a region, and allow for a wider and time-encapsulating vision of a landscape (Beresford 1971). A series of maps need to be studied in order to determine the cultural context of the maps and to distinguish any superfluities or flaws in the maps (Egan and Howell 2001). Furthermore, historical maps are potentially prone to human error and must be used in conjunction with other historic sources to assess accuracy of the map such as: census data, diary entries, newspaper articles, and aerial photography. Agricultural censuses serve as estimates of historic town land use and land cover data, which are useful for understanding how anthropogenic disturbances shape the landscape over time (Hall et al. 2002).

The present Massachusetts forest can be divided into two categories: secondary forest based on land that was historically cleared for agriculture and then recovered to forest, and primary forest that was never fully cleared, but may have been subject to logging and other land use since 1830 (Foster and O'Keefe 1998). Forest compositional change at a regional scale within Massachusetts is most strongly associated with the historical extent of agricultural clearing (Motzkin et al. 1996, Thompson et al. 2013). At the site scale, vegetation is most impacted by local land-use patterns, while at the landscape scale, the vegetation is associated with a patchy combination of a variety of landowner and land-use patterns (Motzkin et al. 1996, Foster et al. 2008). Maps made in 1830, however, do not indicate what has happened to the land afterwards. Although the year 1830 is considered the peak of land clearing in Massachusetts, any individual parcel of land that was forested in 1830 could have been cleared afterwards.

While many of the historic tree species have persisted to modern times, the relative abundance and distribution of tree species has been heavily altered due to agriculture (Thompson et al. 2013, Hall et al. 2002). The modern forest has a more homogenous composition and is summarized by early and mid-successional species such as eastern white pine and red maple, as opposed to the forest of the past where late successional and shade tolerant species such as eastern hemlock and American beech prevailed (Foster 1992, 1998). Furthermore, agricultural clearing left a legacy on species functional traits, as areas that were historically plowed were found to have significantly altered species composition than areas that were unplowed, due to diverse species historically out-competing each other in

the highly dissimilar conditions (Motzkin et al. 1996). While historical land use can have a significant legacy on modern composition, forest composition of both historical and modern forests is very strongly related to environmental conditions. Furthermore, environmental conditions played an important role in historic land-use selection. Forests were generally left on more elevated slopes, more acidic or sandy soils, and in climates generally unsuited for agriculture, which would cause us to see environmental differences today between historically forested and cleared areas (Foster et al. 2008). The relative importance of historical versus environmental factors in controlling forest composition patterns is strongly scale dependent.

This study is built upon a large body of research conducted at the Harvard Forest regarding colonial deforestation and subsequent reforestation, and how these changes in land use impacted long-term species composition and structure. Specifically, this research continues the study of the Harvard Forest 1830 map of forest cover throughout Massachusetts, which depicts the land cover and cultural features found during the year 1830 (Foster and Motzkin 2009). The detail of this map indicates individual pastures and woodlots, the scale at which land-use practices were performed (Hall et al. 2002). Furthermore, the map provides a spatially explicit snapshot of land use at a time of transition from an agriculture-dominated economy to one of agricultural abandonment. This study focused on the present day difference in forest composition between areas that were forested in 1830 and areas that were cleared, as defined by the 1830 map.

The objective of the study was to describe the history of land cover and human land-use change throughout Massachusetts after colonial settlement by testing if colonial land clearing and the history of land use had a lasting legacy on Massachusetts forest composition. I analyzed compositional differences between historically forested and cleared areas to understand how the magnitude and pattern of compositional change varies across the region and reflects the land-use history. Specifically, I assessed the present day difference in forest composition, in terms of tree species abundance and species functional traits, between areas that were forested in 1830, and areas that were cleared in 1830. I also analyzed the 1830 map independently as well as alongside other historical and modern data sources in order to summarize historic land cover patterns and their transition to modern land-use configurations.

Specific questions explored were:

1. What were the modern differences in species composition between areas that were cleared in 1830 and areas that were forested? Were the modern differences in species composition aggregated by specific species functional traits?
2. Could the inherent uncertainty in the accuracy of the 1830 land cover map be ameliorated by restricting the analysis to the highest quality land cover patches and by comparing the map to other historical datasets to account for discrepancies?

3. Could the ambiguity in land use after 1830 be taken into account, especially the clearing of forest cover after 1830, through the analysis of other historical datasets?
4. How did modern land cover distribution compare to historic land cover?

Methods

Study Area

The state of Massachusetts is 21,000 km² in area in the Northeastern United States (69.9-73.5 E, 41.3-42.9 N). The state is largely forested and contains five distinct vegetation zones: spruce-fir-northern hardwoods, northern hardwoods-hemlock-white pine, transition hardwoods, central hardwoods, and pitch pine-oak (Foster et al. 2008). Only seven percent of modern Massachusetts land cover is farmland, which includes pasture, hay, and tilled land (Hall et al. 2002). The state contains 351 municipalities within 15 EPA Level IV ecoregions (Griffith et al. 1994). Massachusetts captures much of the physical, biological, and cultural variation of the New England region, and is therefore a useful landscape for studying regional processes and disturbances (Foster et al. 2008). Furthermore, Massachusetts has unparalleled historical spatial data due to the rich history of forest ecology and land-use change research.

Environmental Data

In order to analyze the role of environmental factors on modern forest vegetation, I assembled several statewide map layers representing environmental variables. Comparing areas that were forested in 1830 and areas that were cleared in 1830, I assessed elevation, temperature and precipitation, and the surficial geology, in terms of percent sand. I used a digital elevation model (DEM) created by joining existing Geographic Information System (GIS) data layer grids (USGS 1993) in ArcMap GIS (ESRI 1996) and then projecting to the Massachusetts State Plane North American Datum 1983 coordinate system. Climatic gradients (but not absolute climatic conditions) across Massachusetts are assumed to have remained relatively constant since 1830 and all analyses were conducted using modern climate data, which was supplied by the modeled Growing Degree Day (GDD) and Precipitation Maps in the Harvard Forest Data Archive (Hall 2003). MassGIS Surficial Geology data layers were used to model geologic features and variation across Massachusetts (MassGIS 1999). Environmental Protection Agency (EPA) Level IV Ecoregion shapefiles were utilized for ecoregion-based landscape analyses (EPA 2014). Monte Carlo random distribution tests were conducted to analyze significant differences between historically forested and cleared areas for each environmental variable. I performed Monte Carlo tests by calculating the difference between group means while scrambling the group membership randomly into two groups of the same size for as many replications as specified, and determining the p-value by the proportion of events that a difference like the one measured could be obtained by chance with the exact distribution of the data or with no assumption of normality (Mikhailov 2001).

Historical Land Use and Land Cover Data

In 1830, a Massachusetts statute mandated that each town produce a town plan depicting its ecological and cultural features such as the roads, buildings, meadows, and surrounding woodlands (Resolves 1831). These original land cover maps are housed at the Massachusetts state archives. The original 1830 town maps were created by different surveyors, leading to varying levels of accuracy within each map. Furthermore, not all towns made land cover maps, and therefore some towns are missing from the aggregated 1830 map. The 1830 forest cover map that I used was created by Harvard Forest researchers, and was produced by digitizing and aggregating the original town land cover maps together. These maps were georeferenced to US Geologic Survey topographic quadrants using a zoom-transfer scope and then digitized in ArcView GIS based on a matchup of road locations, which have changed little since 1830 (Hall 2002). The digital copy of the map can be found in the Harvard Forest Data Archive as the 1830 Map of Land Cover and Cultural Features in Massachusetts (Foster and Motzkin 2009). The map is in Massachusetts Mainland State Plane Projection with NAD 83 in meters (Figure 1). The 1830 map is the foundation of this paper and was used to define historic land cover for the comparison to modern species distribution. I also analyzed the 1830 map to determine the percent of open and cleared land in each town to further understand the land-use patterns in specific areas of the state, and to better compare the 1830 map to other datasets.

In addition to the 1830 map, Harvard Forest has compiled the Massachusetts agricultural censuses and forest clearing records for various years between 1801-1920, including the year 1830. The files are stored in the Harvard Forest Archives (Foster et al. 2003). The agricultural census data originates from Tax Valuations (1801-1860), state agricultural census records (1865-1905), and state forester reports (1920); the major categories of the census differentiate between the census years: 1801, 1830, 1860, 1885, 1905, and 1920 (Foster et al. 2003). Each agricultural census was aggregated by town, unlike as in the 1830 map, and the area of each land cover classification is given for each town. I reclassified the census categories in order to have consistent groupings for comparison to other sources; the major categories were open land, wetland, woodland, missing, and transitional. I compared the agricultural records to the 1830 map in order to determine differences in land cover types and to assess the accuracy of the 1830 map. The agriculture censuses can also be used to help understand the data gap left by towns that were missing 1830 town maps. The 1830 agricultural census data were compared with the 1830 map to determine the relationship between the two historical datasets and the discrepancies between their measurements of historic land use. I compared the agricultural census for 1830 to the 1830 map by regressing the area of land classified as forested in the census against the area of land classified as forested in the map. The same comparison was made for cleared land in the agricultural census versus the 1830 map. Any towns that had missing data, in either the agricultural census or map, were excluded from the analysis in order to prevent skewing of the relationship and to prevent any zero area results

from being shown in the comparison. I also used the agricultural censuses to analyze which towns had continuous forest expansion and which did not from 1830 to 1920, as clearing events after 1830 distort the analysis of modern species distribution based on past land use. I studied forest expansion by comparing the amount of forested area per town (kilometers squared) between the years 1830, 1860, 1885, 1905, and 1920. I defined continuous forest expansion as those towns that had increased forest area for each year-to-year comparison.

Modern Forest Composition and Land Cover Data

I used the USDA Forest Service Forest Inventory and Analysis (FIA) Massachusetts dataset for the present-day inventory and tree plot data (USDA 2014). Tree species composition, among many other forest attributes, is included in the FIA data in extensively arranged plots throughout Massachusetts (and the rest of the United States). FIA data are available on the USDA FIA website (USDA 2014). Within each FIA plot, the diameter at breast height (DBH) was measured for each tree > 25.4 cm DBH. For each plot, I calculated the species basal area, which was used as a measure of species abundance within each field plot for the top 20 species of the region. Geographic latitudinal and longitudinal coordinates for the inventory plots were obtained from the US Forest Service pursuant to a Memorandum of Understanding between the US Forest Service and Harvard University.

I used land cover data from MassGIS to define modern (2005) land cover of the entire state (MassGIS 2009). I reclassified the specific MassGIS land cover categories into broader categories. The MassGIS data layer with the broad land cover categories was compared to historic land use, as defined by forested and cleared areas on the 1830 Map, in order to determine what modern land uses the past land cover had transitioned into (Figure 3). I compared the difference in current land cover in an attempt to understand how persistent and resilient land covers were between historically cleared and forested lands.

Analysis of Modern Forest Composition based on Historic Land Use

I compared modern forest composition to the 1830 map by extracting 1830 land cover values from the 1830 map using the FIA plots unaltered geographic coordinates. The 1830 land cover values allowed the modern species composition data to be subset into two land-use legacy categories (cleared and forested) and compared for present day differences in forest composition. The analysis of the forest composition and structure was performed using R statistical language and ESRI ArcGIS products. I displayed species basal area on Cleveland dot plots and performed multivariate statistical techniques, including Bray-Curtis dissimilarity indexes and Multiple-Response Permutation Procedures (MRPP), on the forest composition data. Bray-Curtis dissimilarity is a statistic used to quantify the dissimilarity between two community compositions (Faith et al. 1987). The values range from 0 to 1, with 1 being completely dissimilar and 0 being identical. MRPP analysis is a nonparametric statistical analysis that tests the hypothesis of no difference between communities and

determines whether the communities are significantly different in terms of their species composition, based on α=0.01 (Mielke and Berry 2001). The MRPP chance-corrected within-group agreement statistic (A) describes the observations of within-group homogeneity of the communities, compared to expectations based on chance (McCune et al. 2002). The values range from 0 to 1, with A=1 meaning that the communities are identical; values are commonly less than 0.1 for community ecology studies and for this study, A=0.05 for biological significance of community difference (McCune et al. 2002).

In an attempt to ameliorate the inherent uncertainty of the 1830 land cover map accuracy, I buffered the field plots to select plots that were greater than 200 meters from each 1830 land cover polygon edge. I used the buffers as an objective criterion for excluding plots near an edge of a land cover class, where map uncertainty is assumed to be the highest. Multiple buffer distances were explored to gain an understanding of how changing the distance from edge impacts the forest composition differences between historically forested and cleared areas. I chose 200 meters as the buffer range to balance increasing certainty and losing field plots, and completed the same analysis on the buffered and subset FIA plots as on the non-buffered FIA plots. I also explored species compositional differences within EPA level IV ecoregions (Figure 2), in order to account for spatial, environmental, and data variability across the state. I performed the same analysis on the ecoregion FIA plots as on the statewide FIA plots. I selected data rich ecoregions for further analysis based on multiple factors: the amount of complete data present in the ecoregion in 1830, the present amount of modern forest cover in the ecoregion based on the MassGIS layer, and the number of FIA plots within the ecoregion. Furthermore, I selected ecoregions with a lack of significant environmental differences between historically cleared and forested areas, in order to reduce the possibility that differences observed were due to differences in the environment, as opposed to differences in historical land use. I also used the towns with continuous forest expansion, as defined by the agricultural census analyses, to subset the FIA plots and run the same analysis. The subset allowed an analysis of field plots in areas where there was less likely to be major clearing events after 1830.

I assembled and defined species functional traits from a combination of the USDA Plant Characteristics and Silvics manuals (USDA and NRCS 2014, Burns et al. 1990). The functional traits examined included shade tolerance, growth rate, resprout ability, fire tolerance, and seed abundance and persistence. Shade tolerance is a plant's ability to tolerate low light levels and was ranked in five qualitative categories from very intolerant to very tolerant. Growth rate is the vertical rate of growth after successful establishment and was ranked in three categories from slow to rapid. Resprout ability is vegetative recovery, or a tree's ability to survive and resprout, and was ranked in two categories, either having the ability or not. Fire tolerance is a species' ability to reestablish and regrow after a fire disturbance and was ranked in four categories from none to high. Seed abundance is the amount of seeds the plant produces and was ranked in three categories from low to high. Seed persistence is if the fruit or seed is described as being persistent or has potential for dormancy, and was ranked in two categories, either

being persistent or not. I performed Monte Carlo randomization tests to compare differences in species basal area between historically forested and cleared areas for each category of each species functional trait, with α=0.05. I included a table defining the hypothesized differences between historically cleared and forested areas for each species functional trait, along with a caption explaining the rationale behind the hypotheses (Table 1).

Results

Modern Species Composition between Historically Forested and Cleared Areas

Aggregating all available data within the state of Massachusetts, I found minimal significant differences in composition between sites that were cleared versus forested in 1830 (Figure 4). Within Massachusetts, four species generally dominated the landscape in both cleared and forested areas: eastern white pine (*Pinus strobus*), red maple (*Acer rubrum*), Northern red oak (*Quercus rubra*), and eastern hemlock (*Tsuga canadensis*). I found 9 out of the top 20 species had different basal area compositions between historically forested and cleared areas, although the differences were small. The more prevalent species in historically forested areas were Northern red oak, Scarlet oak (*Quercus coccinea*), American Beech (*Fagus grandifolia)*, Pitch Pine (*Pinus rigida)*, and White oak (*Quercus alba)*; in historically cleared areas, eastern White Pine, Black Cherry (*Prunus serotina)*, White Ash (*Fraxinus americana)*, and Paper Birch (*Betula papyrifera)* were more prevalent. The mean species basal area for historically cleared areas was 1.19 m^2 per hectare, while it was 1.27 m^2 per hectare for historically forested areas. The range of species basal area for 1830 cleared areas was 0.10-6.01 m^2 per hectare, while for 1830 forested areas it was 0.02-5.32 m^2 per hectare. The Bray-Curtis dissimilarity value revealed that the two communities were slightly different (d=0.15). The MRPP analysis revealed that there was a statistically significant difference between communities, although the agreement statistic signified a very small effect size (p=0.001, A=0.0026). There was a lack of significant difference for environmental variables at the statewide level between forested and cleared areas. The Monte Carlo random distribution test provided p-values above the 95% confidence level for elevation, base soil, GDD, and precipitation (p=0.55, p=0.22, p=0.05, p=0.25).

After aggregating species basal area by functional traits, I found a small number of significant differences between 1830 forest and open land for the statewide combined data. However, the results from the Monte Carlo random distribution test indicated minimal significant functional trait patterns. The most noticeable patterns were that the most dominant species in 1830 forested areas had resprout ability, and the most prevalent species in 1830 cleared areas had relatively high growth rates and seed abundance. Looking at all species, there was far more basal area with intermediate shade tolerance than the other tolerance categories. There were significantly more slow-growing trees in the 1830 forested areas than cleared areas. However, on average, there was two times the amount of rapidly growing trees than the slow and moderate-growing trees. Furthermore, there was significantly more species

basal area with the ability to resprout in the 1830 forested areas than the cleared areas and the landscape as a whole contained more species with the ability to resprout. There was significantly more low fire tolerance species basal area in the 1830 forested areas compared to the cleared areas and there was more species basal area in the low fire tolerance category than the other categories. Most species had medium and high seed abundance. Additionally, there were significantly more species basal area without seed persistence in historically forested areas than cleared areas and more species basal area lacking seed persistence overall.

The ecoregion scale showed scattered and mixed patterns, which seem to be based more on other factors, such as ecoregion differences, species distribution, and the dominance of particular species, than on land-use legacy. Furthermore, the clearest patterns were between the western and eastern sections of the state. The western Berkshire ecoregions, which are known to have more old growth and less forest disturbance, had overall more tolerant species, slower growing trees, more species with resprout ability, low to medium fire tolerance, high seed abundance, and low seed persistence. The eastern section tended to show opposite or less distinctive patterns, as it is more developed than the western section both in the past for agriculture and the present for urbanization.

Western New England Marble Valleys and the Gulf of Maine Coastal Lowland ecoregions were the most data-rich ecoregions and the environment variables were reasonably well distributed between the two historical land cover classes (Figure 5). The Worchester/Monadnock Plateau and Lower Worcester Plateau-Eastern Connecticut Upland ecoregions were the next best ecoregions for analysis as they met all credentials except for environmental variables (Figure 5).

Focusing on one of the two most complete ecoregions, the Western New England Marble Valleys had less than 20% missing data on the 1830 map, a high modern forest cover, a fair amount of dispersed FIA plots, and no significant environmental variables. The species basal area graph showed much more eastern hemlock and American beech in the forested plots, with much more white ash and eastern white pine in the cleared areas (Figure 5). The Bray-Curtis dissimilarity value of 0.65 was high and showed compositional difference between the areas, especially being driven by the difference in eastern hemlock and eastern white pine. MRPP gave a significant p-value and an agreement statistic that implied community difference ($p=0.001$, $A=0.067$). The shade tolerance graph showed much more intermediate to very tolerant species within forested areas, while cleared areas had higher amounts of intolerant to very intolerant species, which is what was expected with a land-use legacy driving a difference between forested and cleared areas. Furthermore, there were significantly more high tolerance species in the 1830 forested areas than in the 1830 cleared areas. The growth rate graph displayed more rapidly growing trees than slow growth trees, although there were significantly more slow growth species in the 1830 forested areas than the cleared areas. There was no significant difference between historically forested and cleared areas for resprout ability although there were more species with the

ability to resprout overall, especially in forested areas. High fire tolerance was significantly more in historically forested areas than cleared; however, the most dominant trait across the landscape was low fire tolerance in both land-use areas. There was no significant difference between the two land-use categories for seed abundance, although medium and high seed abundance species were predominant. The seed persistence graph showed there were no significant differences between historically forested and cleared areas for seed persistence although there were more species lacking persistence overall.

Both the Worchester/Monadnock Plateau and Lower Worcester Plateau-Eastern Connecticut Upland ecoregions had significant results for elevation, temperature, and precipitation, with only soil not differing between historically forested and cleared areas. Species basal area results were almost identical to those seen in the Western New England Marble Valleys and the Gulf of Maine Coastal Lowland ecoregions, and species functional trait results reflected the change of the most abundant species.

Minimizing Uncertainty of 1830 Map Accuracy

While some field plots were securely located within their land-use polygons, others were directly on the edge of polygons. The distance to edge subset revealed greater compositional differences, with more separated values for some species and a higher Bray-Curtis dissimilarity value of 0.19 (Figure 6). Similar to the un-buffered analysis, performing a MRPP test revealed that while there was a statistically significant difference between communities, the agreement statistic signified a very small effect size (p=0.001, A=0.0066). Due to lack of difference, I did not complete species functional trait and ecoregion analyses for the buffered plots.

The comparison of the 1830 agricultural census data and the 1830 map for open land revealed a moderate correlation (R^2 value=0.58) (Figure 7). The 1830 map had higher area measurements, as the 1830 map values were approximately double the agricultural census values. The comparison for woodland revealed a low correlation (R^2 value= 0.25) (Figure 7). The 1830 map again had much higher measurements than the agricultural census, as the 1830 map area values were approximately three times the agricultural census area values. The correlation line was somewhat skewed by a few very high map measurements that made most of the points clump near the zero mark, although there were no zero area measurements. Overall, the 1830 map showed much higher area measurements than the census as a fairly consistent trend for all towns.

Accounting for Land-Use Changes after 1830

The highest percentage of cleared land by town was in Central Massachusetts in the Worcester/Monadnock Plateau Ecoregion. However, the western town of Ashfield had the highest percent cleared at 95.5% cleared. The highest percentage of forested land by town was in the far west and southeast areas of the state in the Berkshire and Cape Cod ecoregions.

Plymouth, in the Cape Cod ecoregion, was the town with the highest percentage of forested area at 77.6% forested. The Green Mountains and Berkshire Highlands ecoregion had the most towns with missing data, as approximately 75% of the towns within the ecoregion were missing data (Figure 2). The total open area in the 1830 map was 11,798 km^2 and the total forested area in the 1830 map was 4,778 km^2.

The forest expansion map showed the towns with forest expansion between subsequent censuses from 1830 to 1920 (Figure 8). Furthermore, the map showed the towns that lost forest between two of the censuses, and towns that were missing data either for the map or the agricultural censuses (Figure 8). While towns with continuous forest expansion were spread across the state, a majority were located in the western portion of the state, especially the southern Berkshire area. There were more than 10 towns with continuous forest expansion that were covered by the missing 1830 data. Furthermore, the Cape Cod region was covered with towns that lost forest between 1830-1920, which was the location for a prevalent amount of the forest on the 1830 map (Figure 1).

For the towns that had continuous forest expansion from 1830 to 1920 (Figure 8), I found some compositional differences. Specifically, there were changes in the relationships of species basal area between historically forested and cleared areas, and a major reduction of certain species such as pitch pine. However, there were only approximately 1/5 the number of plots as the original analysis and thus larger standard errors. The Bray-Curtis dissimilarity analysis on these plots indicated very little change in compositional difference (d=0.16), and I did not perform a MRPP, ecoregion, nor species functional trait analysis due to low sample size.

Fate of 1830 Land Cover

The comparison of the 1830 land cover map to the 2005 MassGIS land cover map revealed the land-use transitions between the two time periods. The graph of 1830 open area to modern land use showed that approximately 53% of the 1830 open area became modern forest, while 10% remained open land; 27% became developed land, 8% became Wetland and 2% became Water (Figure 9). The graph of 1830 woodland to modern land use showed that approximately 67% of the 1830 forest stayed forest in modern times, while 4% switched to open land; 18% became developed, 9% became Wetland, and 2% became Water (Figure 9).

Discussion

Modern Species Composition between Historically Forested and Cleared Areas

The lack of significant differences in environmental variables at the statewide level implied that environmental differences were likely not the cause for any difference seen between historically forested and cleared areas. Differences observed were more likely due to historical land-use legacy. The most important patterns I discovered at the statewide scale were centered on differences in species functional traits. Mid-successional species such

as the shade intermediate eastern white pine and shade tolerant red maple accounted for half of the basal area in both historically cleared and forested areas due to their ability to grow in wide range of climatic conditions. The landscape was dominated by rapidly growing trees, which include the prolific red maple and eastern white pine. The predominance of these species and the homogenization of the modern landscape tended to overwhelm any of the landscape legacies that might be observed (Foster et al. 2008). Furthermore, the landscape contained more species with the ability to resprout, and, in historically forested areas, vegetative recovery appeared to be a more prevalent and important ability. The ability to resprout might have developed more prominently in forested areas, as there was more likely to be a base for the tree to regenerate from and extensive disturbance from logging would promote the relative ability of a species to resprout (Ickes et al. 2003). Areas cleared for agriculture, on the other hand, would clear away any species with the ability. The predominance of low fire tolerance trees in the area can be related to the lack of a natural fire regime in Massachusetts. As fire tolerance development is unnecessary for most species, there was a natural selective bias toward less fire tolerant species, such as red maple, that put energy into other factors such as rapid growth or seed abundance (Foster et al. 2008). Seed abundance was prolific in both historically cleared and forested areas, and the high reproductive capability of the few species that dominate the landscape outweighed any potential differences in seed abundance. Seed persistence and dormancy appeared to be less imperative in historically forested areas than cleared areas. This difference might be due to the protection and rapid sprouting of seeds on and within the forest floor, while cleared areas were open and less persistent seeds could not survive as well as they could within the forested areas. Furthermore, there was an overall predominance of species that lack persistent seeds, which could be explained by an inverse relationship of seed abundance and persistence. The abundance of seeds, especially in already highly forested areas, made the need for seed persistence lower and a waste of valuable resources (Venable and Brown 1988). The accumulation of species with persistent seeds in seed banks was favored in highly cleared areas. Where there is little natural habitat, the dispersal rate is low, and there is a high likelihood of reproductive failure (Venable and Lawlor 1980).

Overall, there were not many significant differences at the statewide level, of which there are several potential explanations. It is possible that data averaging across the state skewed recognizable differences given the large variability in forests across the state. Furthermore, inaccurate and variable historic surveying methods could lead to a lack of difference between forested and cleared areas due to imprecise and often erroneous delineation of which areas were forested in 1830 and which were not, even after performing a buffer distance analysis to attempt to account for inaccuracies in historic surveying (Hall et al. 2002). Another possible factor is the amount of missing data from the original 1830 map, which leaves much of the western portion of the state out of the analysis. Ultimately, 180 years of natural and anthropogenic processes may have erased much of the land-use legacy.

At the ecoregion scale, the results tended to be mixed and scattered, not revealing distinctive patterns other than those previously mentioned. Furthermore, the significance of environmental variables for many of the ecoregions implied that environmental differences might be the cause for the differences seen between forested and cleared areas, and not historic land use. Therefore, attributing differences in species compositional traits between historically forested and cleared areas to historic land-use legacy is controversial. I performed an ecoregion analysis in an attempt to overcome: data averaging across the state, the influence of missing data overriding differences, and the impact of the east-west environmental gradient. However, there were too few plots in each ecoregion to make valid assumptions. The larger ecoregions had enough FIA plots for analysis, but some of the smaller ecoregions had less than 50 plots, with one ecoregion only containing 14 plots for both forested and cleared areas together. The lack of plots, as well as the continual problem of surveyor accuracy and landscape changes occurring after 1830, made it difficult to distinguish any landscape legacies.

More than 180 years of natural and anthropogenic processes have largely erased colonial land-use legacy differences between historically forested and cleared areas on modern forest composition. However, the lack of current significant differences between 1830 cleared and forested areas points to the resilience of the Massachusetts landscape and the ability of species to adapt and recover in new conditions. The resilience of the Massachusetts forests as a whole is evident, as many of the historic trees species have persisted to modern times (Thompson et al. 2013). The agricultural legacy on the land most prominently affected the relative abundance and distribution of tree species (Hall et al. 2002). Furthermore, as the landscape is currently fairly homogenized and dominated by species generalists, it would be difficult to see patterns based on historic landscape legacies even if they did still exist.

Minimizing Uncertainty of 1830 Map Accuracy

For statewide analysis I selected the 200 meter buffer distance because it was a good balance between increasing certainty and losing field plots; however, the analysis reduced the total FIA plots by 50% and there were few profound differences. Nonetheless, the distance to edge analysis helped to further illustrate the variant differences in accuracy of past surveyors, and how that might have impacted the accuracy of the map. For future work, to improve the accuracy and usability of the map, the exactitude and precision of the surveyor for every town should be qualified or quantified and each town should be given a rank based off of this surveyor accuracy.

In the comparison of the 1830 map to the 1830 agricultural census, there was a large difference in area measurements for woodland comparisons. This discrepancy was most likely due to the difference in how the two datasets measured and defined forest area. The 1830 map documented forest land, while the agricultural census had a woodland category as well as unimproved and unimprovable categories. The unimproved and unimprovable categories

were said to be lands that were cleared for agriculture but were transitioning to forested land. These categories were in a separate category from woodland, but some of the area might belong in the woodland category and might account for the discrepancy between the two datasets (Hall et al. 2002). A more thorough analysis of the agricultural census datasets, and their variable and inconsistent land-use categorizations over the years, is necessary in order to perform a useful and meaningful comparison of the census datasets with the 1830 map. This analysis should be explored in future research, as the agricultural censuses have the potential to reveal some of the inaccuracies of the map and may also provide data in areas previously lacking measurements, which was the reason I first examined them.

Accounting for Land-Use Changes after 1830

Examining the 1830 land cover map by town revealed more distinct trends of cleared and forested lands in Massachusetts. This visualization of historic land-use trends was useful when analyzing the 1830 map against both modern and historic data. The 1830 percent forested and cleared maps detailed the major trend that a higher percentage of cleared areas were located in the central and eastern portions of Massachusetts, while a higher percentage of forested areas were located in the western and southeastern portions of the state. Regardless, high percentages of cleared areas were across the entire state, predominating the overall low forest percentages, even in many of the areas where forest levels were higher.

As most areas were allowed to recover forest by the 1920s, the study of continuous forest expansion from 1830 to 1920 revealed potential areas where old growth would most likely be found. One of the most interesting patterns I uncovered was the loss of forest in the Cape Cod region between the time period, as the Cape region had a high proportion of the forest area on the 1830 map. The change in Cape forests occurred shortly afterwards, as approximately half of the forest was cleared and plowed in the mid-19th century for agriculture, and much of the remaining forest extensively logged (Eberhardt et al. 2003). After agricultural abandonment in the area, pitch pine expanded, predominately on former plowed sites as the species thrives in disturbed and open areas (Motzkin et al. 1996). The present predominance of the species in the region illustrates the extent and intensity of the clearing that occurred after 1830. The southwest Berkshire region appears to be the best area to analyze in terms of continuous forest expansion, as it had a substantial amount of the 1830 forest as well as had a number of the continuous forest towns, and therefore is expected to give a better representation of any landscape legacy patterns, if they do exist. The focus on the Western New England Marble Valleys in the ecoregion analysis depicts this, as the ecoregion was firmly embedded in the Berkshire area. However, the ecoregion had very few FIA plots, and a modern study of just the southern part of the ecoregion using a higher sampling size, still in comparison to historic land uses as detailed by the 1830 map, could potentially reveal a landscape legacy from agricultural clearing.

The use of towns with continuous forest expansion to subset FIA plots was rather inconclusive and did not unveil any patterns other than those already seen. Other than the species basal area analysis, I did not explore further assessment due to low sample size of plots and the species basal area output was too similar to previous results. However, the overall analysis of towns with continuous forest expansion highlights future potential areas of study and could be augmented by analyzing when each town stopped losing forest in comparison to gaining forest, and how that affects the understanding of the staggered impact of historic landscape change over time.

Fate of 1830 Land Cover

The comparison of the current land cover map provided by MassGIS to the 1830 Map revealed that the majority of the land cleared in 1830 became forest and that most of the historically forested areas stayed forested. The MassGIS land cover map also revealed that both historically forested and cleared land became developed land as the second highest land use after forest, the predominant usage being for low and medium density residential areas. Additionally, much of the eastern half of the state became residential, while the western half of the state largely remained forested. This pattern can be traced back to the historic patterns observed, as a predominance of the historically forested area was in the west or southeast, while the east and central areas of Massachusetts were heavily cleared, especially the Boston area. Lastly, the modern analyses show the drastic changes solely between two snapshots, 1830 and 2005, and do not show the variable changes between the years. While the agricultural censuses were a good foundation for doing so, more land-use maps between the 1920s and modern times need to be made available digitally and incorporated into the analyses.

Conclusion

There is a lack of overall significant differences between forested and cleared areas in terms of land-use legacy. However, land use greatly restructured the entire landscape as a whole, moving to a more homogeneous forest as well as a large amount of mid-successional species. The high levels of disturbance, including Dutch elm disease and the chestnut blight, opened up the landscape for species such as red maple, which respond rapidly to disturbance regimes and can out-compete more slowly growing species in disturbed areas (Foster et al. 2008). The most prominent species appear to be generalists that grow and establish rapidly as well as those that have wide tolerance and range to climatic conditions. Massachusetts has been overall highly resilient to the disturbance regimes, although there has been a shift to more rapid growing, mid-successional, and shade-intermediate species such as red maple and eastern white pine, and a decline in longer lived and slower maturing species such as eastern hemlock and American beech (Foster et al. 2008). Fire suppression in the landscape could also play a role in the prominence of these species, as the lack of a fire regime favors

species such as white pine and red maple over trees like pitch pine and white oak (Foster et al. 2008). The east-west environmental gradient also predominately influences the modern distribution of species composition, and the capacity for environmental factors and ecological processes to overpower certain land-use impacts, especially over a long period of time, cannot be ignored from analysis (Hall et al. 2002).

The differences seen between statewide and ecoregion analyses illustrate that ecological questions are scale dependent, and need to be examined on the proper scale. There are different patterns found at site, landscape, and regional levels, due to diverse factors being prevalent at each scale. Certain scales tend to display a homogenization of ecological characteristics due to either uniform practices or an averaging of a variety of practices, while other scales illustrate heterogeneous patterns due to patchy and variable land-use patterns (Foster et al. 2008). Focusing on another scale is often necessary to analyze the appropriate factors and find the patterns being searched for. Focusing on another scale, perhaps at the town level or a conglomeration of ecoregions, might have revealed greater legacy differences.

The 1830 map has certain limitations, which need to be understood when performing an analysis based on it. Different town maps were created by different surveyors, who had different levels of accuracy, as shown by the distance to edge analyses. There are more remaining 1830 woodlands in the western half of the state, especially in the Berkshires, which happen to have the most missing data on the 1830 map. The analysis and map do not take into account the difference between pastured lands, hayfields, and tilled lands in term of cleared areas, which greatly impacts which species can grow and thrive in the area, proven by the long-term species composition legacies between plowed and unplowed areas (Motzkin et al. 1996). Furthermore, more than 180 years have passed since 1830. Agricultural clearing did not end until around 1885, and events such as the intensive 1885 logging and the 1938 hurricane are not displayed in the map (Hall et al. 2002). Performing the analysis with an equivalent 1885 land-use map, if made available, would be more useful, as 1830 was the beginning of the peak of agricultural clearing, while 1885 was the end, and would account for more of the major land cover changes during the shift towards agricultural abandonment.

This study utilized the 1830 map in combination with other historical sources as well with modern land-use datasets. It reveals many of the potential problems with using the map, and attempts to outline a methodology of ways to correct and account for some of the issues that come along with using historic data sets. Furthermore, the utilization of the map and analyzing the problems that come along with it has highlighted the paramount possible uses for the map. The 1830 map should be used as a tool to select useful regions and areas for analysis, as it gives a glimpse into past land use and can be used side by side with other sources to pick areas of analysis based on the prerequisites of the study. The identification of areas with primary and old growth forest would be greatly supported by the use of the map to select potential locations. Furthermore, the map should also be part of site overviews and used in conjunction with field evidence in any study of Massachusetts forest composition

in order to provide the essential background and historic land-use information.

The study of historic land-use legacy on modern species composition is increasingly important. Urban sprawl, which has been developing in Massachusetts since 1950, has led to land cover patterns similar to historic agricultural clearing, and has the potential to lead to a new surge of forest fragmentation and decline (Foster et al. 2008). The 1830 map and other historical data sets are useful tools in dealing with these new conservation challenges. We can learn from the legacies left by past land use in order to predict and prepare for future changes in the landscape.

Author Note

All tables and figures referenced in the paper can be found in the Appendix section. All data tables, reclassification tables, and additional graphs and figures can be found in a supplementary document, which can be provided by the author upon request.

Acknowledgements

This research was supported in part by the National Science Foundation Harvard Forest Long Term Ecological Research Program (Grant No. NSF-DEB 12-37491). This research was also made possible by MOU (14-MU-11242305-025) between Harvard University and USFS Northern Research Station to use the unfuzzed FIA plot locations.

References

Beresford, M. 1971. History on the ground: Six studies in maps and landscapes. London: Methuen.

Burns RM, and H Barbara. 1990. Silvics of North America: 1. Conifers; 2. Hardwoods. Agriculture Handbook 654. vol.2, p. 877. U.S. Department of Agriculture, Forest Service, Washington, DC.

Clements, FE. 1905. Research Methods in ecology. Lincoln, NE: University Publishing.

Cousins, SAO. 2001. Analysis of land-cover transitions based on 17th and 18th century cadastral maps and aerial photographs. Landscape Ecology, 16: 41-54.

Eberhardt R, Foster DR, Motzkin G, and B Hall. 2003. Conservation of Changing Landscapes: Vegetation, land-use history and fire of Cape Cod National Seashore. Ecological Application, 13: 68-84.

Egan, D and EA Howell (ed.). 2001. The Historical Ecological Handbook: A Restorationist's Guide to Reference Ecosystems. Washington D.C.: Island Press.

EPA. 2014. Level III and IV Ecoregions of the Continental United States (http://www.epa.gov/wed/pages/ecoregions/level_iii_iv.htm#Level%20IV, 24 September 2014). US Environmental Protection Agency, Corvallis, OR, USA.

ESRI. 1996. ArcView GIS 3.2. Environmental Systems Research Institute, Inc., Redlands, CA, USA.

Faith DP, Minchin P, and L Belbin. 1987. Compositional dissimilarity as a robust measure of ecological distance. Vegetation, 69: 57-68.

Foster DR. 1992. Land-Use History (1730-1990) and Vegetation Dynamics in Central New England, USA. Journal of Ecology, 80(4):753-771.

Foster DR. 1998. Changes in the countryside of Thoreau: insights into the ecology and conservation of the New England landscape. Cambridge (MA): Harvard University.

Foster DR, Motzkin G, and B Slater. 1998. Land-use history as long-term broadscale disturbance: regional forest dynamics in central New England. Ecosystems, 1:96-119.

Foster DR and J O'Keefe. 1998. An Ecological History of Massachusetts Forests. Arnoldia, 58(2):3-31.

Foster DR, Hall B, and J Burk. 2003. Massachusetts Historical Landcover and Census Data 1640-1999. (http://harvardforest.fas.harvard.edu:8080/exist/xquery/data.xq?id=hf014, 24 September 2014). Harvard Forest Data Archive: HF014.

Foster DR, Donahue B, Kittredge D, Motzkin G, Hall B, Turner B, and E Chilton. 2008. New England's Forest Landscape: Ecological Legacies and Conservation Patterns Shaped by Agrarian History. In: Redman CL, Foster DR, Agrarian Landscapes in Transition: Oxford University Press, Inc., New York, NY.

Foster DR and G Motzkin. 2009. 1830 Map of Land Cover and Cultural Features in Massachusetts. (http://harvardforest.fas.harvard.edu:8080/exist/xquery/data.xq?id=hf122, 24 September 2014). Harvard Forest Data Archive: HF122.

Griffith GE, Omernik JM, Pierson SM, and CW Kiilsgaard. 1994. The Massachusetts Ecological Regions Project. US Environmental Protection Agency Environmental Research Laboratory, Corvallis, Oregon, USA.

Hall B, Motzkin G, Foster DR, Syfert M, and J Burk. 2002. Three hundred year of forest and land-use change in Massachusetts, USA. Journal of Biogeography, 29:1319-1335

Hall B. 2003. Massachusetts Growing Degree Day and Precipitation Maps. Harvard Forest Data Archive: HF088.

Ickes K, Dewalt SJ, and SC Thomas. 2003. Resprouting of woody saplings following stem snap by wild pigs in a Malaysian rain forest. Journal of Ecology, 91: 222-233.

MassGIS. 1999. Surficial geology datalayer description. Mass-GIS, Commonwealth of Massachusetts. Executive Office of Environmental Affairs, Boston, MA, USA.

MassGIS. 2009. Land Use (2005) datalayer. Mass-GIS, Commonwealth of Massachusetts. (http://www.mass.gov/anf/research-and-tech/it-serv-and-support/application-serv/office-of-geographic-information-massgis/datalayers/lus2005.html, 26 November 2014). Executive Office of Environmental Affairs, Boston, MA, USA.

McCune B, Grace JB, and DL Urban. 2002. Analysis of Ecological Communities. MjM Software Design.

Mielke PW and KJ Berry. 2001. Permutation Methods: A Distance Function Approach. Springer Series in Statistics. Springer.

Mikhailov GA. 2001. Monte-Carlo method. Encyclopedia of Mathematics. (http://www.encyclopediaofmath.org/index.php?title=Monte-Carlo_method&oldid=15336, 26 November 2014).

Motzkin G, Foster DR, Allen A, Harrod J, and R Boone. 1996. Controlling Site to Evaluate History: Vegetation Patterns of a New England Sand Plain. Ecological Monographs, 66(3):345-365.

Ramankutty N and JA Foley. 1999. Estimating historical changes in global land cover: Croplands from 1700 to 1992. Global Biogeochemical Cycles, 13(4): 997-1027.

Resolves of the General Court of the Commonwealth of Massachusetts... (https://archive.org/details/actsresolvespass182831mass, 24 September 2014). Boston: Dutton and Wentworth, 1831, pp.270-272.

Thompson JR, Carpenter DN, Cogbill CV, and DR Foster. 2013. Four Centuries of Change in Northeastern United States Forests. PLoS ONE 8(9): e72540.doi:10.1371/journal.pone.0072540

USDA. 2014. FIA DataMart. FIADB version 5.1 (http://apps.fs.fed.us/fiadbdownloads/datamart.html, 24 September 2014). US Department of Agriculture, Forest Inventory and Analysis, Arlington, VA, USA.

USDA, NRCS. 2014. The PLANTS Database (http://plants.usda.gov, 24 September 2014). National Plant Data Team, Greensboro, NC 27401-4901 USA.

USGS. 1993. Digital elevation models data users guide 5. US Department of the Interior, US Geologic Survey, Reston, VA, USA.

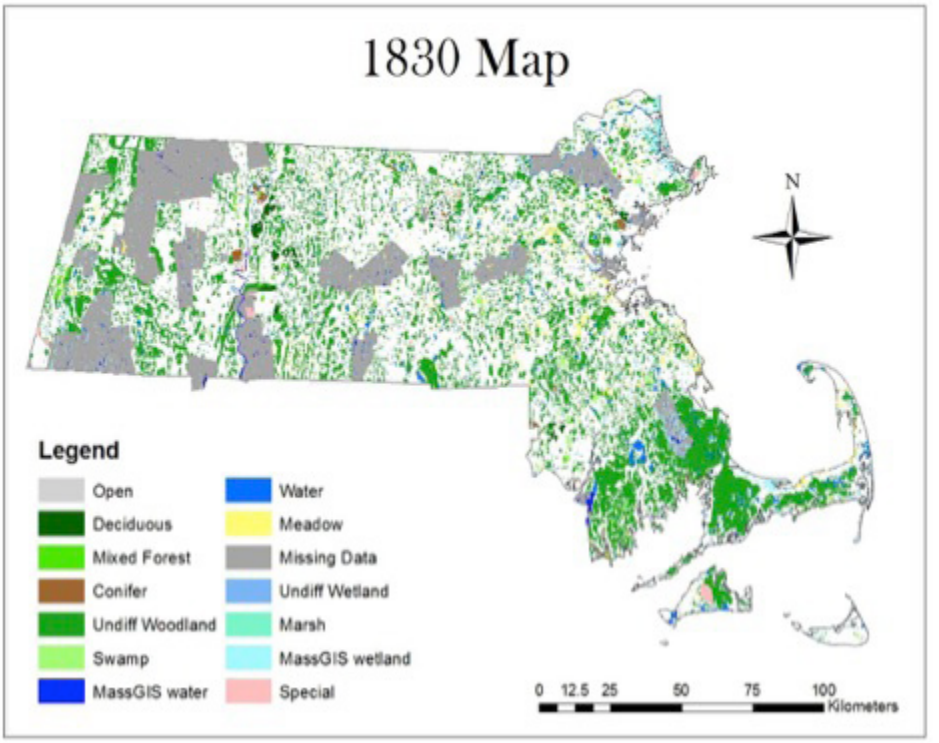
Venable DL and JS Brown. 1988. The selective interactions of dispersal, dormancy, and seed size as adaptations for reducing risk in variable environments. American Naturalist, 131: 360-384.

Venable DL and L Lawlor. 1980. Delayed germination and dispersal in desert annuals: escape in space and time. Oecologia, 46: 272-282.

Appendix: Tables and Figures

The map depicts the land covers found in 1830. The map was digitized and aggregated by Brian Hall at the Harvard Forest and can be found in the Harvard Forest Archive (Foster and Motzkin 2009).

Figure 1: Harvard Forest 1830 Map.



The ecoregion reference map depicts the 15 EPA Level IV ecoregions in Massachusetts. The number on each ecoregion is its ecoregion identification number as detailed by the EPA and the legend has the descriptive name and identification number for each ecoregion (Griffith et al. 1994).

The table lists the general hypotheses for the differences in species functional traits between historically forested and cleared areas. I based the hypotheses on past land-use studies and the successional re-establishment of species.

Shade intolerant and rapidly growing species should dominate historically cleared areas, as there theoretically should be more early successional species; late successional species theoretically prevail in historically forested areas, meaning more shade tolerant and slow growing trees (Foster 1992). Historically forested areas should have more species with resprout ability as it would give species an established advantage in a forest and would

accumulate over time; historically cleared areas would have cleared all trees completely, and establishment ability would be more important than resprout ability (Ickes et al 2003). There should be lower fire tolerance in historically cleared-areas as investment would be in establishment and not fire tolerance, as the area lacks a steady fire regime; within historically forested areas, however, species would be able to grow thicker and more fire tolerant bark due to more resources available and a greater opportunity cost (Foster et al. 2008). The inverse nature of seed abundance and persistence means historically forested areas should have higher seed abundance, due to prevalence of nearby species and need to out-compete by quick establishment, while historically cleared areas should have lower seed abundance due to the distance from forested areas and would need more persistent seeds that could remain in the soil until the opportunity to sprout arose (Venable and Lawlor 1980).

Figure 2: Ecoregion Reference Map.

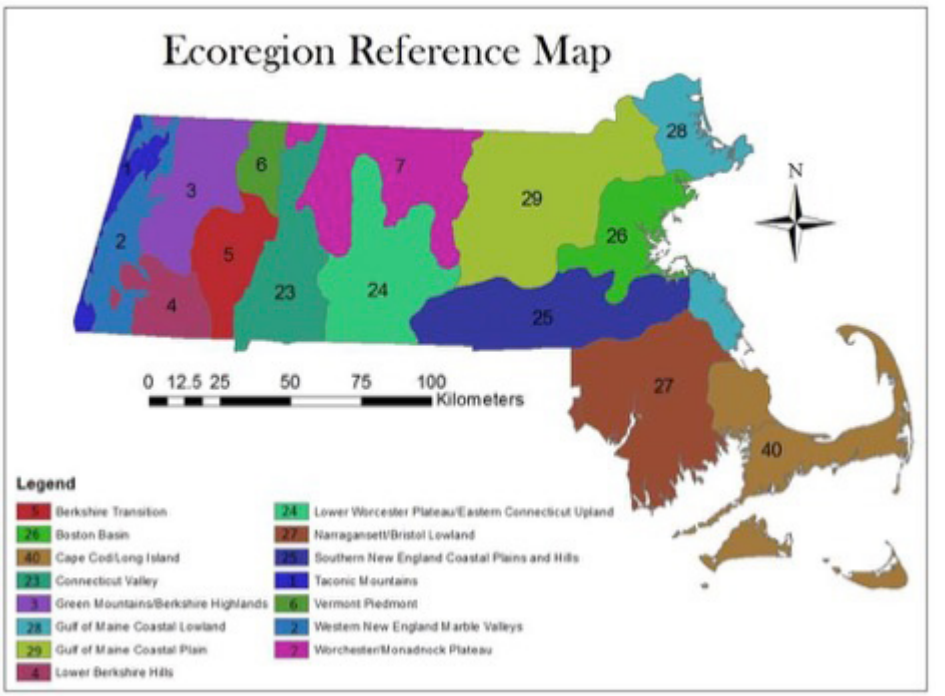
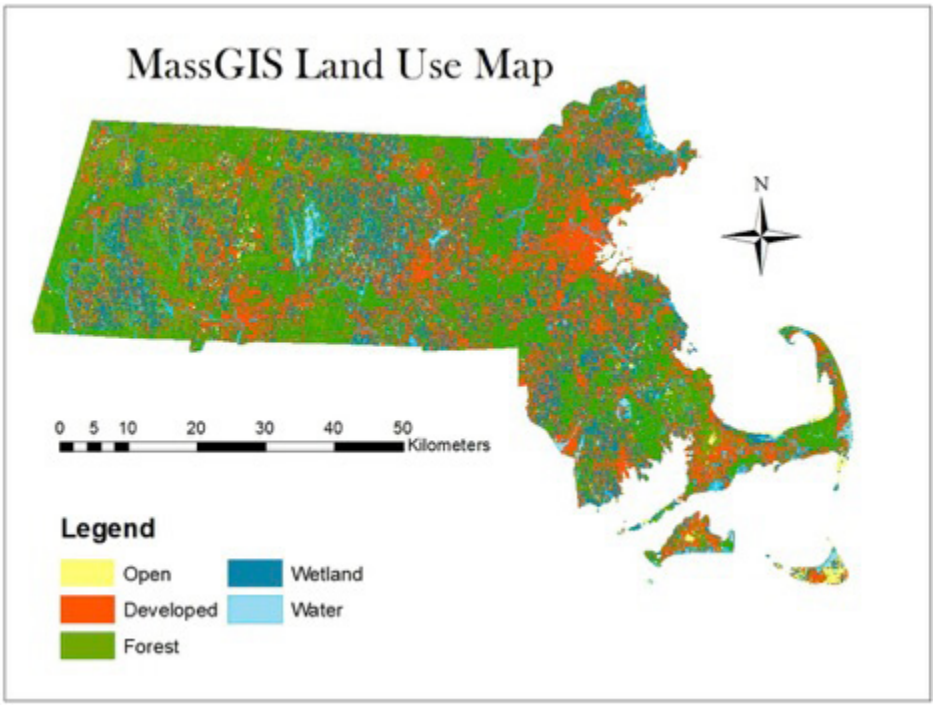


Table 1: Species Functional Trait Hypotheses.

Shade Tolerance	Growth Rate	Resprout				
Ability	Fire Tolerance	Seed Abundance	Seed Persistence			
Cleared	Intolerant	Rapid	No	Low	Low	Yes
Forested	Tolerant	Slow	Yes	High	High	No

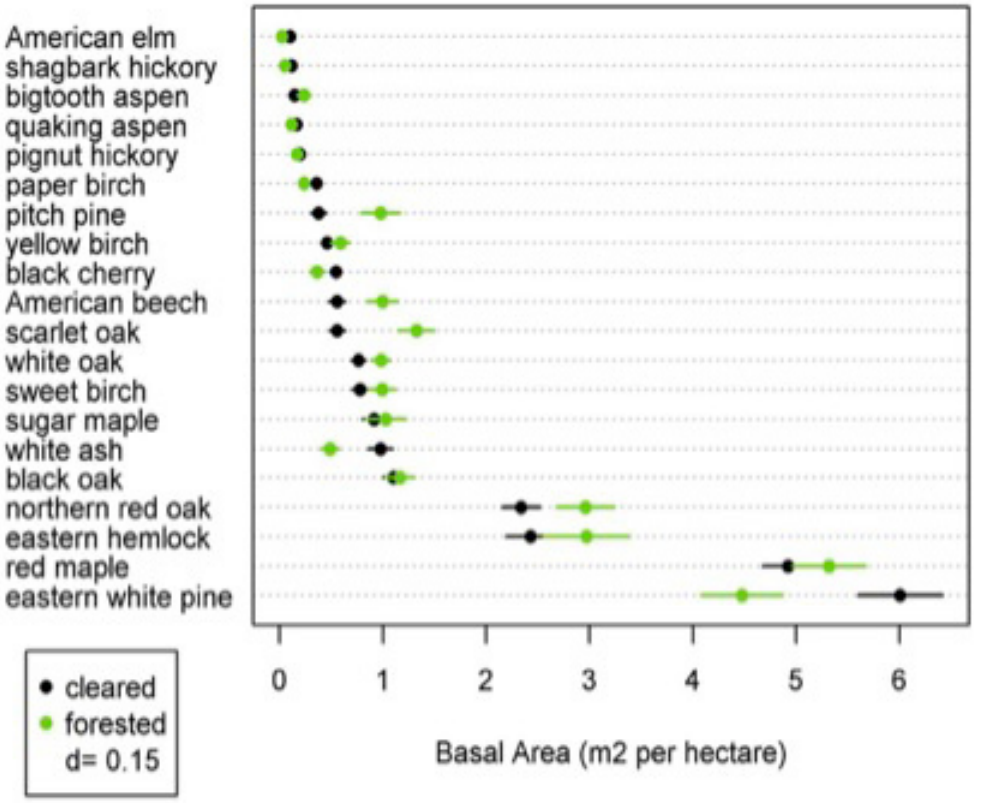
The map depicts the land covers found in 2005, re-categorized into broader categories then those classified by MassGIS (MassGIS 2009). The reclassified categories were used for the MassGIS map and for analyzing the transformation of the 1830 map to the modern landscape. The open category was based on agriculture or naturally open land use, while the developed category was based on industrial and commercial land use.

Figure 3: MassGIS Land Use Layer Map.



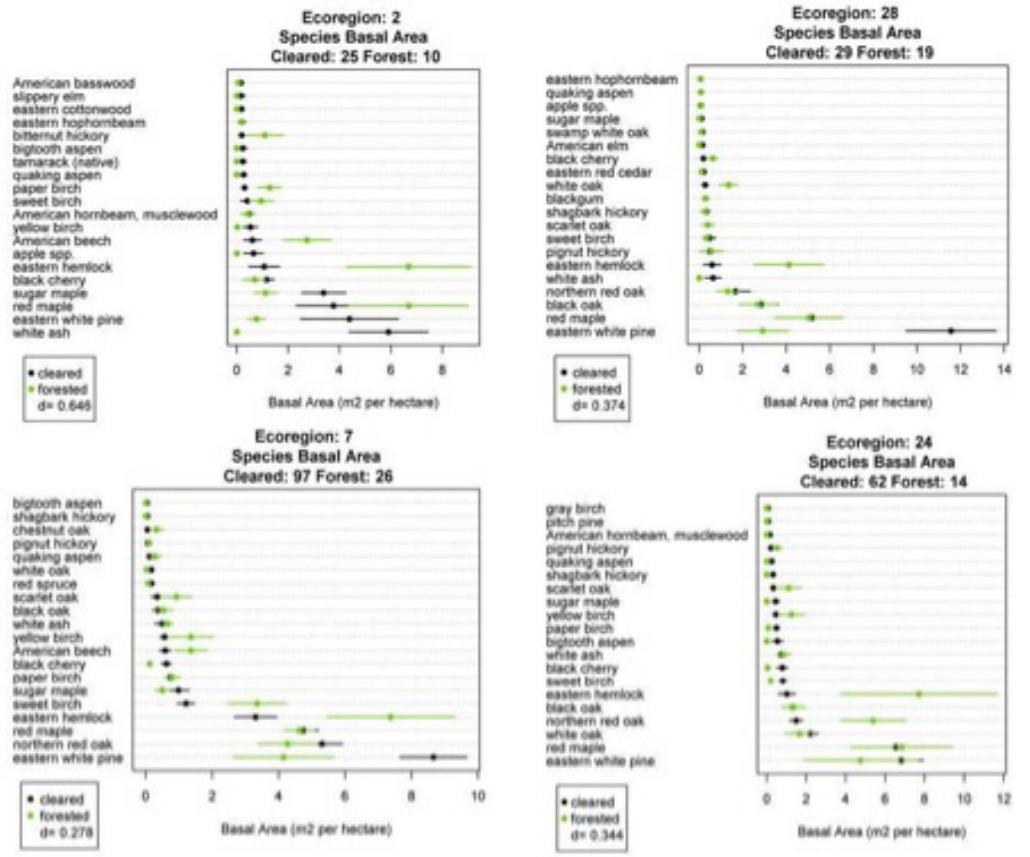
The Cleveland dot plot displays the basal area in meters squared per hectare for the predominant tree species across Massachusetts. The green dots represent modern forest species composition on sites that were forested in 1830, while the black dots represent the modern forest composition on sites that were cleared in 1830. There are 601 FIA plots in historically cleared areas, and 302 FIA plots in historically forested areas. The d value is the Bray-Curtis dissimilarity value, which expresses the dissimilarity between the forested and cleared community compositions.

Figure 4: Massachusetts Modern and Historic Species Basal Area Differences.



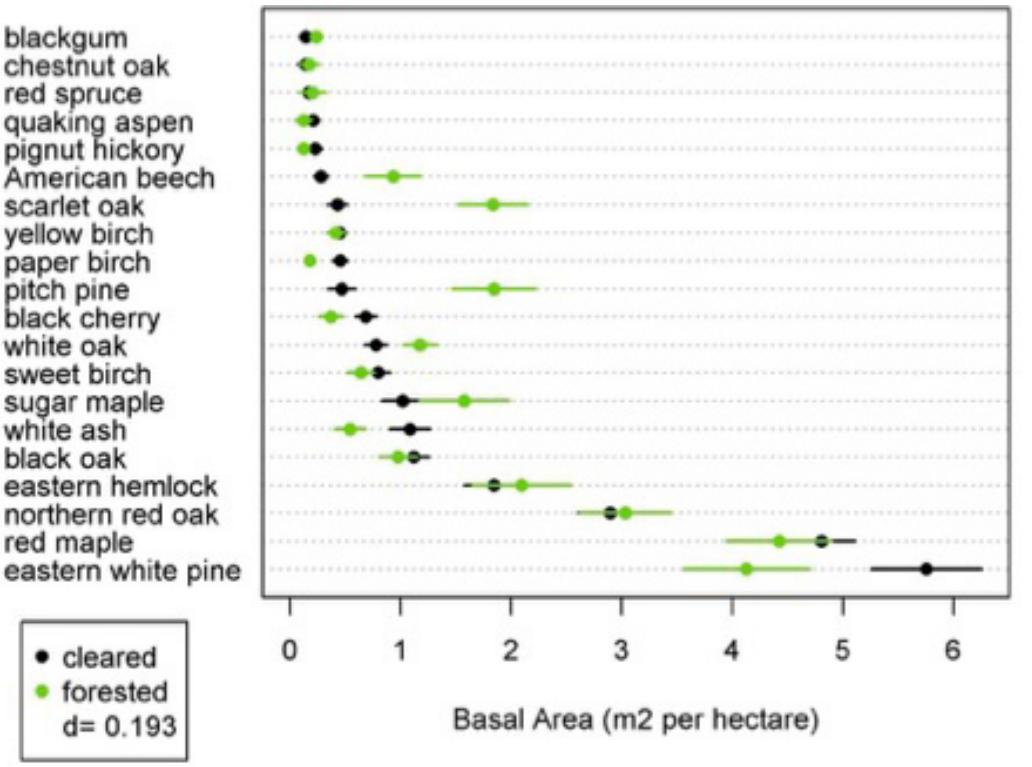
The Cleveland dot plots display the basal area for the predominant tree species for the most representative ecoregions. The green dots represent areas that were forested in 1830, while the black dots represent areas that were cleared in 1830. The number after Cleared is the number of FIA plots in historically cleared areas, while the number after Forested is FIA plots in historically forested areas. The d value is the Bray-Curtis dissimilarity value.

Figure 5: Ecoregion Species Basal Area Graphs.



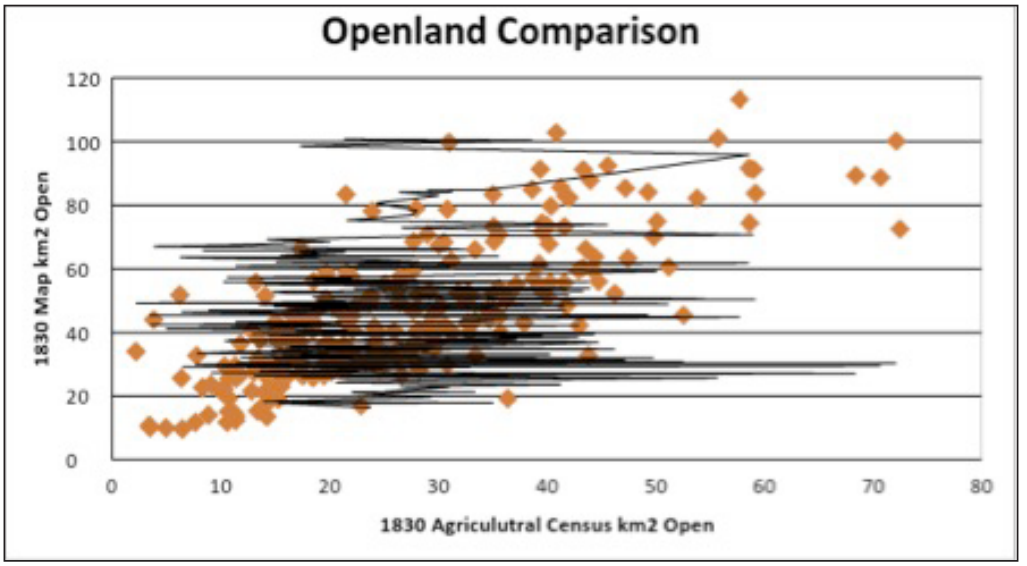
The Cleveland dot plot displays the basal area for the predominant tree species at the statewide level with a 200 meter buffer. The green dots represent areas that were forested in 1830, while the black dots represent areas that were cleared in 1830. There are 301 FIA plots in historically cleared areas, and 137 FIA plots in historically forested areas. The d value is the Bray-Curtis dissimilarity value.

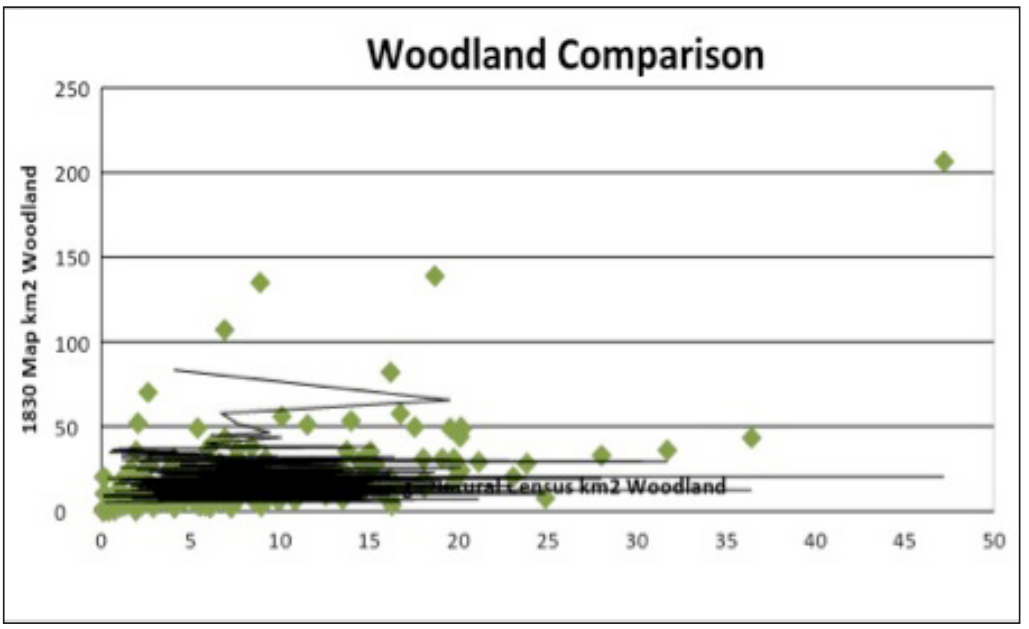
Figure 6: Statewide 200 meter Buffer Species Basal Area.



The graphs compare the kilometers squared of open or forested land found in each town in the 1830 Map versus the kilometers squared of open or forested land found in each town in the 1830 Agricultural Census dataset. The regression line and R^2 value are depicted. Towns with more than 50% missing data from either dataset are not shown.

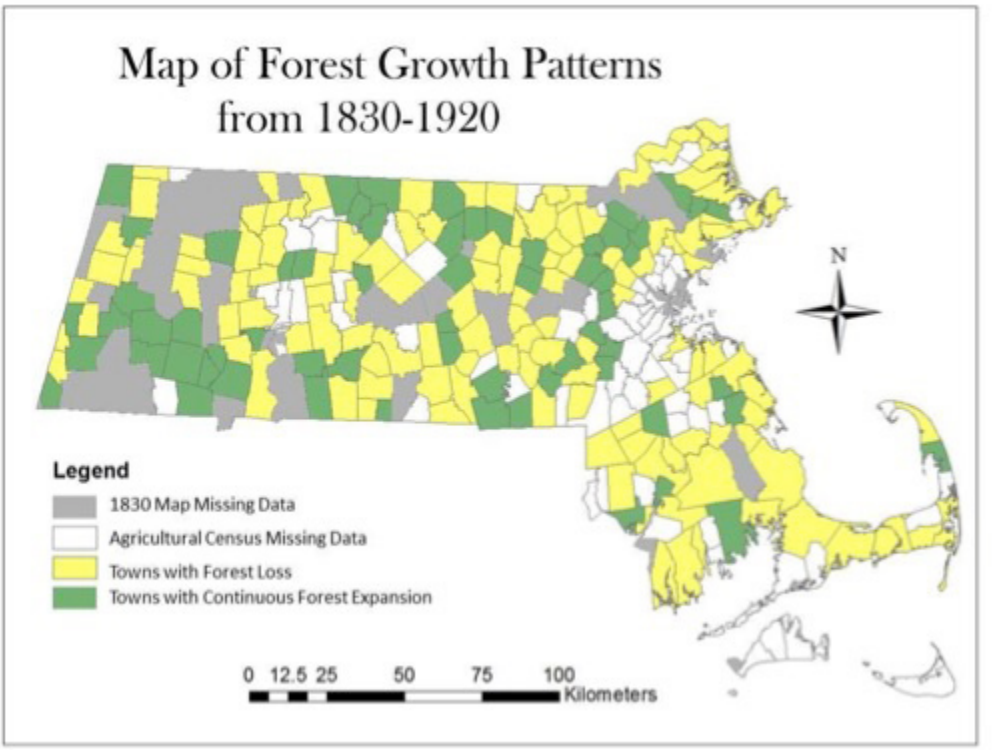
Figure 7: Comparison between 1830 Map and 1830 Agricultural Census by Town.





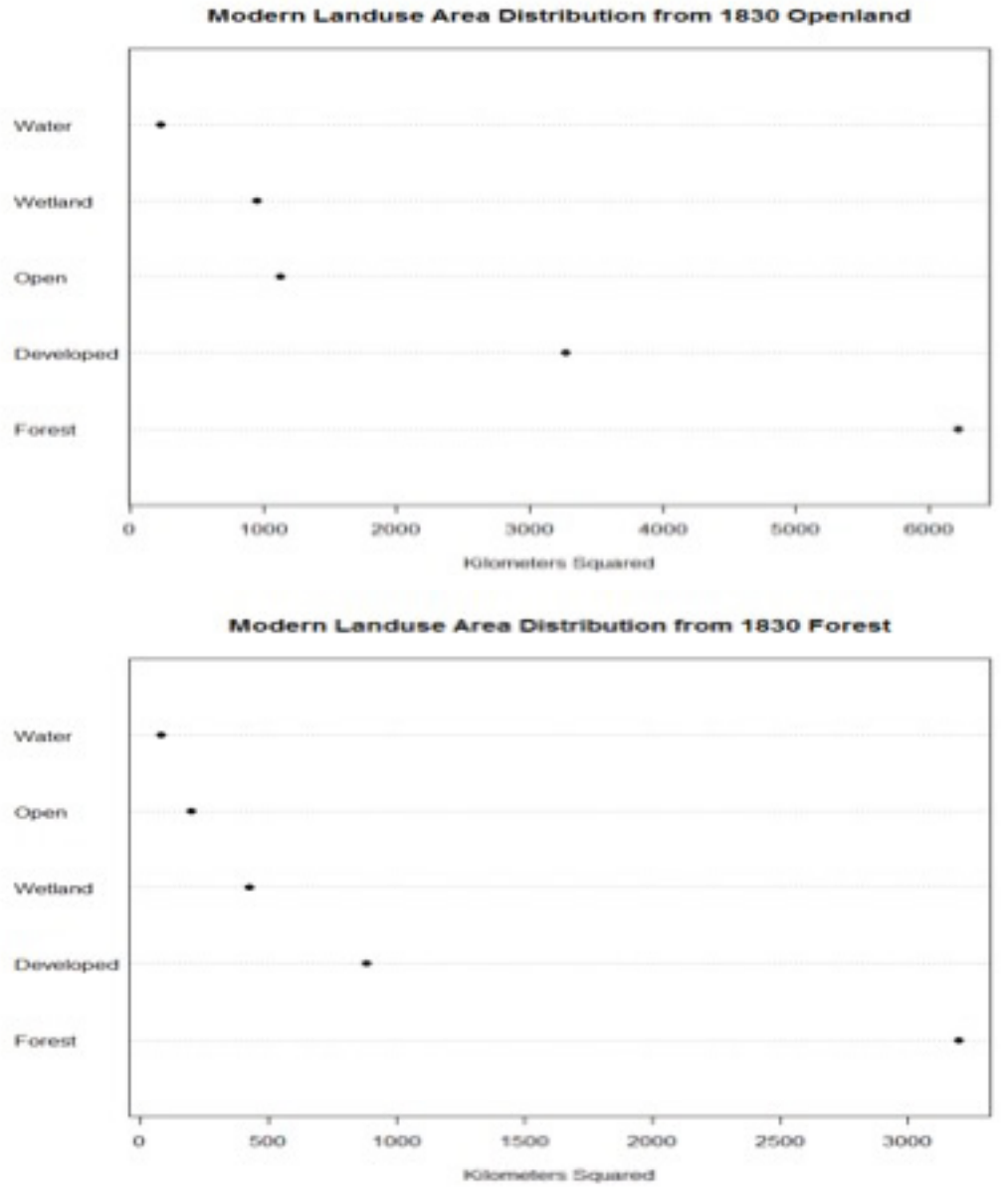
The map shows which towns had continuous forest expansion from 1830 to 1920 (in green) as determined by the agricultural censuses, and which towns had forest loss between the years 1830 and 1920 (in yellow) as determined by the agricultural censuses (Foster et al. 2003). Towns with missing 1830 Map data are in gray and the towns missing from the agricultural censuses are in white (Foster and Motzkin 2009).

Figure 8: Map of Forest Expansion Patterns from 1830-1920.



The graphs show the modern land uses that the land defined as open or forested by the 1830 land cover map transformed into. It depicts how many kilometers squared of 1830 open or forested land transformed into broad modern land use categories, recategorized from the more specific MassGIS categories.

Figure 9: Land Use Change from the 1830 Land Cover Map to the Modern MassGIS Land Cover Map for Historically Open and Forested Land.



The Role of Labels in Making Inferences about Group Categorization

The University of Texas at Austin

Undergraduate Research Journal

Volume 15
Number 1
Spring 2016

Subhashini Madhavan

Abstract

The purpose of this study was to investigate the effect of group labeling on children's willingness to learn new information from similar versus dissimilar informants. Cues to similarity were manipulated by assigning children aged 5 to 7 years old different labels to measure their preference of learning from two puppet informants, one of whom shared the same label and one that had a dissimilar category label. Results indicated that children used social category labels to attend to membership and guide inferences about names for novel

objects, largely choosing to endorse labels provided by the puppet informant that shared the same category label over the puppet informant that had a dissimilar category label.

The Role of Labels in Making Inferences about Group Categorization

The creation of social categories is an aspect of human nature that allows us to make sense of the world. These categories are arbitrary cultural constructions that are not grounded on objective facts (Diesendruck & Deblinger-Tangi, 2013). Individuals hold expectations of categories and are biased to believe that members of categories conform to certain shared behaviors (Shutts, Roben & Spelke, 2013). Moreover, the expectation that behaviors are shared within a category allows us to predict, perceive and explain other's actions (Powell & Spelke, 2013; Shutts, Roben, & Spelke, 2013). In this sense, constructing social categories provides us with one of the most prominent tools of social cognition, the ability to make inductive inferences. An inductive inference is the ability to extend our knowledge from what we know to making judgments about novel instances (Waxman, 2010). The judgments that individuals make may be due in part to the labels assigned to a category. People tend to expect that members of a category that share the same label share the category's underlying behaviors as well (Graham, 2008)

Children are known to form basic categories of gender, age and race at a young age and are able to use a category members underlying similarity, features and behaviors to make inferences (Master, Markman, & Dweck, 2012; Birnbaum, Deeb, Segall, & Ben-Eliyahu, 2010). However, when children are presented with a novel category that does not provide indicators about perceptual similarity or shared behaviors of members, children use the category label, in the absence of other information to make inductive inferences.

In selecting between discordant information provided by two informants children prefer to accept information from informants whose have been accurate in past statements or whose statements align with their own previous knowledge base (Corriveau & Harris, 2009; Reyes-Jaquez & Echols, 2013). This is especially true when children are presented new information (Chen & Waxman, 2013; Dunham, Stepanova, Dotsch, & Todorov, 2014). When it comes to learning new information, young children rely more on social cues, such as physical and psychological similarities, to make judgments (Reyes-Jaquez & Echols, 2013). Labels are one potential cue to similarity for category membership, and very little research has been done to understand if children will prefer to learn, in the absence of any other information, from a source that shares a similar category label.

The following literature review presents current findings relating to the formation of categories, language structure and labeling. First, I will discuss the importance of social categorization and how children form social categories. Then, I will discuss the initial set of social categories that develop during childhood. Next, I will explain the negative consequences of social categorization. Following this, I will examine the importance of labeling in relation to making inferences about novel categories. In addition, I will discuss the

importance of language and its relation to social categorization and labeling. Finally, I will examine the role that labeling plays in generating preferences and learning new information.

The Importance of Categorization

An important part of social cognition is the general tendency for people to segregate members of the society into specific groups based on overarching traits such as gender, race, and culture. Separating individuals into categories is predicated upon the notion that certain individuals will appear similar, conform to in-group behaviors or share certain traits that warrant grouping into an assigned category (Waxman & Chen, 2013; Gelman, 2004). This propensity for segregating members into separate groups is known as social categorization (Gelman & Meyer, 2011). Categorization serves a society in terms of organizing information in coherent and meaningful ways, by allowing members of society to make inductive inferences about a category's members (Allport, 1954). Categories can be used to make inductive inferences when little or nothing is known about a certain individual, such that the category in which an individual is placed allows us to form judgments about that individual (Waxman, 2010). For example, if someone is placed into the socially-constructed, gender-based category of being a female, individuals tend to form inductive inferences like: she will like pink things, will have a soft spoken voice, will be family oriented and enjoy delicate things.

People believe that, like animal categories, human categories have underlying unique properties attributed to them that gives them their identity (Gelman, 2003). This belief is called *psychological essentialism*. Psychological essentialism is the belief that social categories are innate and causal, membership within a category is absolute and permanent, and categorization of individuals allows us to make inductive inferences (Gelman, 2004; Birnbaum et al., 2010). Children's ability to recognize social categories develops very early on. Beginning in infancy, children show the capacity to categorize individuals on the basis of factors like gender, race, and age.

Children's ability to attend to such information at an early age allows us to speculate on the dimensions that children rely on when evaluating other people (Shutts, Roben, & Spelke, 2013). To date, in studies examining the formation of social group categories in children, few topics have been as extensively studied as race and gender based categorization. Preference of individuals of the same race develops as early as 3 months old. Children attend to information provided by same race individuals more often, as a result of greater familiarity and exposure to members of their race (Kelly et al., 2005).

Construction of Racial and Gender Based Categories

Research in natural and laboratory based studies have indicated that even toddlers selectively favor informants that are similar to them or are within a similar category. The first categories that children are able to comprehend are racial and gender based categories

(Dunham et al., 2014). At a young age, children indicate a preference for individuals that have the same race, gender, age, and accent (Martin & Fabes, 2001; Bigler, Jones & Lobliner, 1997; Kinzler, Shutts, & DeJesus, 2009; Waxman, 2010). Shutts, Banaji & Spelke (2010) sought to understand to what extent children used race and gender categories to determine preferences for novel objects. They presented 3-year-olds with objects or activities that were endorsed by unfamiliar people that differed in race (black or white), gender (male or female) and age (child or adult). Children were then asked to select an object or activity to play with that was endorsed by members of a different race, gender or age. The primary findings were that 3-year-old children showed robust preferences for novel objects that were endorsed by their own gender regardless of racial category. Children had a relatively weak preference for objects endorsed by their own race.

As young children often lack exposure to races other than their own, the first consideration that children take into account is the category of gender. The initial basis of social categorization is centered on attending to gender differences, even in the absence of salient visual cues or outright category labels. As gender is the social category that develops first, once children are informed of the differences in gender they base their decisions on social categories they already know or social categories they are made aware of. As children get older, the properties attributed to a certain category can also be used to make inferences about new members (Yamauchi & Markman, 2000).

Making Inferences about Racial Categories

When an individual or entity belongs to a category, people tend to derive judgments about members based primarily on the category to which they belong. In making inferences and conclusions about a person, we emphasize their social group membership (Yamauchi & Markman, 2000; Gelman & Markman, 1986). In an experimental setting, Waxman (2010) showed that children were more willing to extend a novel property presented to them (e.g., a card showing a black woman), to members within the same race group (e.g., associating the card with members of the African-American race) rather than across race groups (not associating the card with members of the Caucasian race). This suggests that category membership plays a vital role in how individuals, specifically children, learn to make conclusions about different people. Although categorization has been shown to have innumerable benefits, it can also result in negative consequences, including loss and distortion of information (Master, Markman, & Dweck, 2012)

Negative Consequences of Categorization

Placing members of society in different categories can result in negative consequences, like stereotyping and prejudice (Rothbart & Taylor, 1992). When an individual is assigned to a social category, they have a general proclivity to prefer members of their own in-group compared to members of the out-group (Powell & Spelke, 2013).

To understand the role of stereotyping as a consequence of categorization, Bigler & Liben (2007) manipulated group categorization to discern if it affects the formation of social stereotypes. These researchers postulated a theoretical model, known as developmental intergroup theory (DIT) that explained how stereotyping may emerge in early childhood. According to the DIT, manipulation of group categories can be based on perceptually salient attributes (e.g., gender, age, and race) or psychologically salient attributes (e.g. color, texture, shape, reading ability). Bigler & Liben (2007) manipulated psychologically salient attributes by assigning 6 to 11years old children to novel groups that wore different colored t-shirts. After three weeks, experimenters assessed in-group bias amongst these children. Results indicated that children showed bias towards members of their own group by preferring to engage in daily activities with those of their own category as opposed to out-group members. In-group contact between members of a category increases familiarity, interaction and a basis for preference. As such, merely forming categories can sometimes trigger the construction of stereotypes (Over & Carpenter, 2012; Powell & Spelke, 2013). When children form inferences about members of a category, whether positive or negative, their inferences can be based on various properties, one of which is the noun label given to a category. Explicit labeling of a category makes the category, and members associated with that category, more salient. As children use labeling to make inferences about categories, researching this process is crucial to understanding how children navigate their social worlds.

The Importance of Labeling

Research conducted on racial and gender based categories has shown that at early age children mainly use gender and race based categories to make inferences about others. Research explains how children create and deploy representations of new social categories, the factors that contribute to social category formation.

Research addressing children's understanding of novel categories has demonstrated that category labels convey rich conceptual information (Gelman & Markman, 1986; Reyes-Jaquez & Echols, 2013; Waxman 2010) and that labeling a category plays an important role in guiding inferences (Moore & Parker, 1989; Gelman, 2003; Keates & Graham, 2008). When a novel social category is explicitly marked with a noun label, the category label can outweigh perceptual similarities when making inductive inferences (Gelman & Markman, 1986; Graham, Nayer, & Gelman, 2011; Diesendruk & haLevi, 2006). This suggests that a noun label can contribute to their acquisition of social categories (Waxman, 2010).

Providing a noun label for a category in which an individual is placed serves various purposes. First, the noun label highlights the individuals' membership status in that category and guides children's inferences about new and unnamed individuals (Gelman, 2003). Second, labels act as an attention spotlight that encourages children to examine the members of the category more closely.

Third, labeling a category increases children's attention to that category relative to a category that has not been labeled or a category that is indicated with a visual cue (Dunham et al., 2014). Powell & Spelke (2013) found that visual cues delineating group-membership were not required for infants to make inferences about categories. Having a single shared noun-label and evidence of engaging in similar behavior was sufficient for children to make conclusions about a category. Baron et al. (2014) supported the findings found by Powell and Spelke.

Baron et al. (2014) examined the principles that children use to create social-group categories. The experiment examined the isolated effects of noun labels and visual cues on social categorization. Test participants heard a story about two individual groups that participated in either antisocial behaviors or positive behaviors. The two groups had a visual cue to category membership. In the absence of a visual cue, children made much stronger inferences about labeled categories for groups that were defined by noun labels than groups that were defined by a visual-cue. Thus visual cues delineating similarity between group members were not required for categorization (Gelman, 2003; Graham, Nayer, & Gelman, 2011).

Similarity between individuals of the same category also guides how children form interpersonal bonds and friendships. At an early age children have already established stereotypes, and indicate preferences for members within the same category as themselves (Bigler & Liber, 2007; Baron et al., 2014). As children get older, and take part in more social interaction with other people, the tendency to prefer members of the same category determines with whom children form interpersonal bonds and friendships.

Children's Use of Similarity as a Basis of In-Group Preferences

When making new friends, children look for similarity markers like attitudes, preferences, and background (Johnson, 1989; Fawcett & Markson, 2010). School-age children experience an expansion in the social categories they are exposed to as they become older. As children develop, their social interaction extends to peers in a classroom setting. Bias can become a basis of interaction, leading to children preferring to interact with members of their own social category (Bigler & Liben, 2007; Baron et al., 2014).

Fawcett & Markson (2010) sought to understand if children aged 3 years old preferred to make friends with members that shared the same trait label. Researchers manipulated two dimensions: psychological similarity (i.e., shared food preference) and external physical similarity (i.e., appearance manipulated through t-shirt color). Children were presented with two gender identical puppets that matched the gender of the participant. The preferences of one of the puppets were similar to the child, while the other puppet's preference was dissimilar to the child. For example, for the dimension of internal psychological similarity, the similar puppet claimed to like the same food as the child, while the dissimilar puppet preferred the food the child disliked. After the puppets expressed their preferences, the child was asked to choose one puppet to interact with as a measure of liking. Children largely

preferred to interact and make friends with the puppet that shared their preferences. Shared preferences pose an important basis for friendship and bias as children seek out members who have similar interests and traits. The second part of the experiment sought to understand if this preference pattern would remain when using a subtle visual cue of stickers. Results indicated that when using colored stickers instead of labeling the group explicitly, children were less likely to choose the similar puppet.

As such, temporary labels like stickers have no basis for similarity as they are subject to change. Fawcett & Markson (2010) even concluded that children slightly preferred to interact with members that did not have the same colored sticker as them. Thus, when visual indicators are present children may respond more to permanent characteristics associated with the puppet rather than to temporary traits like a sticker.

Young children essentially make inferences about social categories to make predictions about different individuals of the world. As an extension of the research discussed above, researchers Reyes-Jaquez & Echols (2013) looked to address whether cues to similarity guide children's preferences for learning new labels for unknown objects. Three to five year old children were presented with informant puppets that were either similar or dissimilar in their preferences. After a familiarization period, children were presented with 3 unfamiliar objects and had the chance to request labels for the unfamiliar objects from the puppets. Children preferred to trust the object label given by the puppet that shared the same preferences and hair color as them as more reliable than the object label given by the puppet that was dissimilar.

Understanding Category Membership

The use of labels is aiding researchers in understanding the way children make inferences and learn new information about members within a category. Children have a basic understanding of broad gender, race, and age categories from a very young age and use their understanding of these categories to help make sense of the world. When introduced to novel categories children can rely on the category label to make inferences about group members. This study built on research conducted by Reyes-Jaquez & Echols (2013) and examined the role that labels have on children's willingness to learn new object names from others based solely on the cue of a shared label. The current experiment tested this broader question asking (a) if children use an informant's group category to guide their preferences for learning the label of an unfamiliar object and (b) if a social category label, on its own, can lead to a same-group learning preference.

Methods

Participants

A total of 16 typically developing 5 to 7 year old English speaking children participated in the study. The mean age of males was 6.1 years, and the mean age of females was 6.6 years. Children were randomly assigned to one of two noun label category groups. All children

were recruited from a participant database at The University of Texas at Austin's Children's Research Lab (CRL). Prospective participant's parents received invitations letters, followed by either a phone call or email.

Design

The goal of this study was to understand the effect of labeling on children's preference to learn from similar puppet informants (SI) versus dissimilar informants (DI). Similarity was determined by whether the child shared a group category label with the puppet informant or had a different label. The experiment used a between-subjects factorial design to compare two groups. The manipulation for this study was assigning the child and the similar or dissimilar informants with different category labels. The first novel category was labeled as the sun group. This group consisted of children that were randomly assigned the sun-group noun label. The second group was labeled as the moon group. This group consisted of children who were randomly assigned the moon-group noun label. The independent variable was the two categories that were given different label names. The dependent variable was the informant from which the child preferred to learn a new object label. Both groups took part in the same procedure. The children were tested in a laboratory setting. Each child took part in three trials. The presentation of object labels from the similar versus dissimilar informant was counterbalanced across trials and across children. The hypothesis was that children would indicate a preference to learn new object words from informants who had the same category label as the one given to them.

Materials and Stimuli

Participants were introduced to two realistic looking, gender identical puppets, and for each gender there were two puppet informants. The puppets' gender was matched to the child's. Both male and female gender identical puppets had brown or blond hair(see Figure 1). Each puppet informant introduced themselves to the child and then proceeded to identify their group category. Perception of similarity was manipulated by placing the child in the same category as one puppet while placing the other puppet in a category with a different label. To establish children's preference for the puppet informant's labels, the study made use of three novel objects (e.g., toys) that were child-friendly objects. The entire experiment was video-recorded so it could be used for data processing purposes.

Procedure

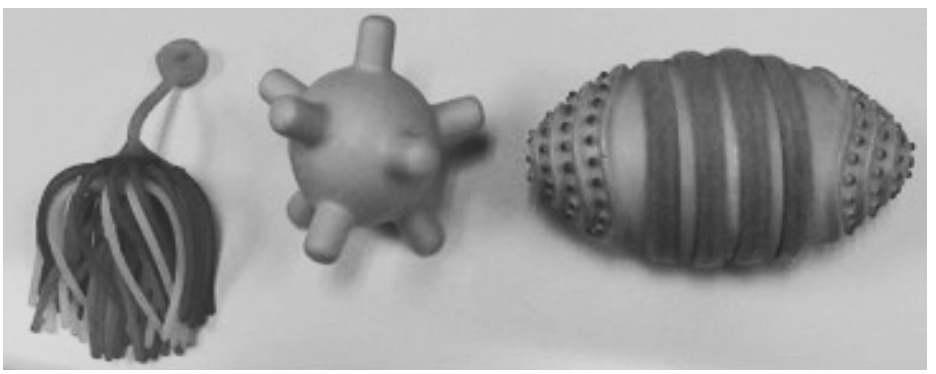
This study was adapted from the research conducted by Reyes-Jaquez & Echols (2013). After consent was obtained, the children were led to the data collection room at the CRL. In this study, children were introduced to two puppet informats and learned that one puppet shared a similar category label as the child and that the other puppet had a dissimilar category label.After the experimenter introduced the puppets to the participant, the ex-

Figure 1: a. Female Gender identical puppet informants; b. Male gender identical puppet informants with screen backdrop; c. Novel objects presented to children (in sequence).

a.

b.

c.



perimenter highlighted and reinforced the similarity of the child with one of the puppet informants and the disimilarity with the other informant (e.g. "Look! You and Amy are in the same group!") for the puppet informant that was in the same group as the child and (e.g. "Look! You and Lauren are in a different group!") for the dissimilar puppet informant. These statements were presented with equal enthusiasm for both the similar and dissimilar informant. The experimenter then asked the child a series of control questions to ensure that the child understood that one puppet from the pair was in a similar category and the other puppet from the pair was not.

Children were then presented with one of the three novel objects used in the experiment, and were asked "Do you know what this toy is?" After the children responded to the experimenter's question, the experimenter asked "Shall we ask Amy or Lauren what this toy is called? Maybe they know what this toy is called?" Once the child responded to the experimenter, the experimenter asked the puppet, "Can you tell my friend what this toy is called?" The puppet informant then labeled the object in front of the child. After the first puppet informant labeled the object, the second puppet informant was asked the same question, and labeled the object. Similar and dissimilar puppet informants both described the object with different names that were unfamiliar to the child (e.g. "this is a zom" or "this is a blick"). Children's memory was tested after each object had been labeled by asking, "Can you tell me what Amy thinks this is called?"

After each novel object had been labeled by each puppet informant, the experimenter questioned the child by asking what the child thought the object was called (e.g. "Can you tell me what this object is? Is this a zom or a blick?") For each participant, this procedure was replicated for the two other novel objects in the sequence. During the experiment parents were allowed to be present outside the room and observe their child complete the study but were asked to sit out of sight, preventing any inteference in the study. Upon completing the experiment, children were given a small toy for participation.

Measures

Choice. Our main measure of analysis was the puppet informant from which children chose to learn the new object label. Choice was defined as the informant that participants selected after the novel object had been labeled by both puppets and they were asked the question "Can you tell me what this object is? Is this a zom or a blick?" Choice was determined by either a verbal response (e.g., this object is called a blick) or pointing (e.g., pointing their finger to the puppet informant whose label they chose to endorse). We calculated the choice scores for each participant across trials by recording how often across the series of three trials the child endorsed the label given by the similar group informant versus how often the child endorsed the label given by the different group informant.

Control questions. Children were asked a series of control questions to ensure that they understood that one puppet informant was in a similar noun-label category as them while the other puppet informant was a dissimilar noun-label category. The control questions that were asked to the children have been adapted from the Fawcett & Markson (2010) study. Control questions that the child was asked were:

1. What group is Amy in? (Puppet 1 Female)
2. What group Lauren is in? (Puppet 2 Female)
3. What group are you in? (Child's Group)

*The same questions were used with different names, for the two male gender-identical puppet informants.

Results

The present study examined children's willingness to learn new object words. The independent variable for the study was the group label provided to the child by the similar and dissimilar puppet informants. The dependent variable was the child's preference for the novel object label provided by either the similar informant (SI) or the dissimilar informant (DI). The two research questions addressed were first, in the absence of any other information, can a novel category-label alone facilitate learning preferences? Second, are children more willing to learn new words for novel objects from informants that are in the same group category as themselves?

To examine if children preferred to learn new words from similar informants (SI) over dissimilar informants (DI), children received a score of 1 for every time they endorsed a word provided by the SI (and 0 otherwise; with a maximum score of 3 for the three novel objects). Across the 16 participants, children indicated a preference for words provided by the SI 62.5% of the time. Children's mean score for the SI was 1.875 (*SD*= 0.89) and for the DI was 1.125 (*SD*= 0.89), which also seemed to indicate a general preference for learning new words from similar over dissimilar group informants (see Figure 2). Children's preference for labels provided by the SI was tested by comparing the SI scores to a chance level of 1.5 via a one-sample *t* test. Children's preference for endorsing words provided by the SI approached significance, t (15) $=1.695$, $p=0.111$.

Analyses additionally examined the frequency of choice selection for both the SI and the DI. Across the three novel objects the modal number of response for the SI was 2 and for DI was 1. The pattern of responses across the 16 participants indicated that, children chose labels provided by the SI more frequently than those provided by the DI (See Figure 3).

Figure 2: Average number of times (maximum _ 3 per question type) that children endorsed labels provided by the similar informant (SI) or the dissimilar informant (DI). Error bars indicate standard errors.

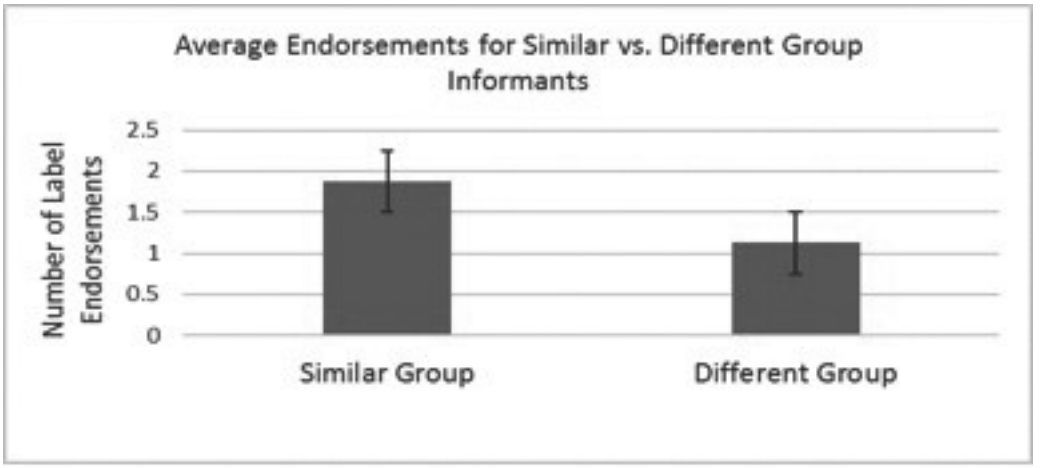
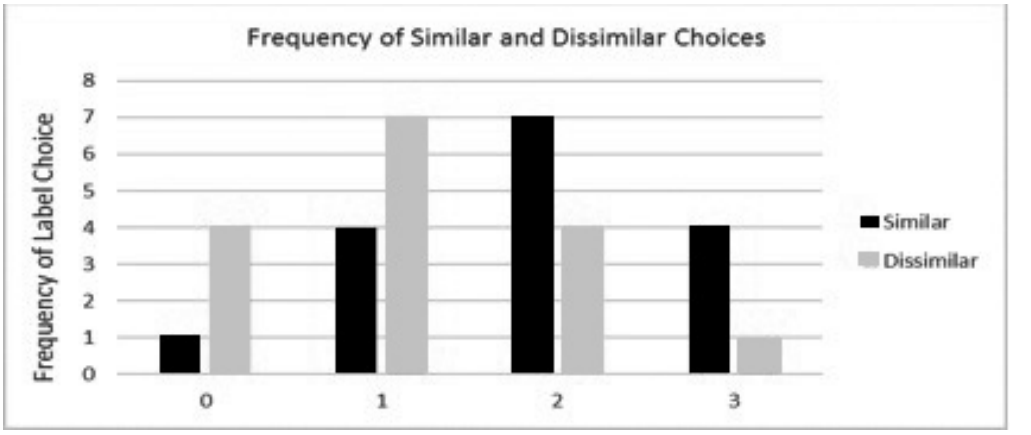


Figure 3: Frequency of response choices for the similar (SI) and dissimilar informant (DI) grouped across number of labels endorsed.



Discussion

The main goals of this study were to examine whether a group label alone would facilitate learning preferences and, specifically, whether children are more willing to learn new words from informants that are in the same group category as themselves versus informants who are in a different group category than themselves. Results showed support for the hypothesis. Simply sharing a category label with an individual tends to increase willingness to learn from that individual, and children differentially tended to prefer learning new words for novel objects from the similar group informant (SI) over the dissimilar group informant (DI).

The results of this study offer key insights into the conditions under which children rely on shared group labels to make inductive inferences. When the group label was explicitly marked and clearly identified, children made use of inductive inferences by relying on the shared group label to ascertain the objects name even when the labels were presented without other identifying markers signaling social category membership. This extends research conducted by Keates & Graham (2008), which documented that infants as young as 16-months are able to use noun-labels as markers of social category membership.

Children's preference for learning from similar informants may be due to them possessing early expectations that members of the same social group category should make similar choices. Simply sharing the same group category label as a similar informant (SI) may have been sufficient to prompt children to differentiate between words provided for the novel object by the SI versus the DI.

Previous studies that have looked at social categories have found that children use gender based categories to guide inferences about novel object preferences, as it is a highly potent social category. However, by making use of identical gender based puppets in this experiment, children were able to place special attention on the similarity of the labels instead of focusing on the gender. This experiment suggests that young children hold expectations of shared -behavior from group members and these expectations may motivate children to use similarity as a marker of group membership and as a means of selecting which sources to learn from.

Children indicated a preference for learning from similar informants over dissimilar informants. It is possible that children's willingness to learn from the similar informant was driven by in-group favoritism. In-group favoritism refers to a pattern of favoring members of one's own in-group over out-group members (Kinzler, Shutts, & Correll, 2010). Early representations of social group categories are organized dichotomously into "us" versus "them" (Gelman, 2003). Infants as young as 11-months have demonstrated an ability to segregate individuals on the basis of shared relationships (Powell & Spelke, 2013). By providing the same label to the child and one puppet informant, researchers highlighted the similarities and dissimilarities between the child and the informants. Children may have then associated themselves as being categorized in the same group as one of the puppet informants ("us") and the dissimilar puppet informant as being a member of another group to which they did not belong ("them").

Both Bigler and Liben (2007) and Baron et al. (2014) have documented that similarity-based indicators produce substantial discriminatory bias toward favoring members of our own group over members of another group. Participants in the present study may have indicated a similar in-group bias towards the similar puppet informant more often than the dissimilar puppet informant as a result of associating themselves with the similar puppet informant, thereby forming a favorable inter-group comparison (henceforth increasing their positive self-concept). To the extent that children use similarity cues as a maker of group membership, the experiment's findings are consistent with the idea that children as young as 5 years old are capable of identifying and relating to same group-members (Reyes-Jaquez & Echols, 2013).

In summary, these results make three important contributions to the study of the influence of labeling. To our knowledge, this is the first study to provide evidence indicating that children prefer learning new words for unfamiliar objects from similar over dissimilar informants when similarity is determined solely by labels. In the absence of other information, children made use of the group labels to indicate which puppet's label they preferred to endorse. Second, our study also suggested that children may display an in-group bias towards the similar informant (SI) that shares the same label as them. Finally, the findings indicate children's ability to comprehend and attend to differences between categories. Once similarity was explicitly marked with the label, children favored the similar over the dissimilar informant in regards to endorsement of labels.

These results are important not only in their contributions to understanding children's selection of learning methods but also because they have practical implications. When older preschoolers begin making friends in school, they may choose to interact with and potentially learn new information from peers that are in similar groups or organizations as them. Findings from this study provide us with knowledge that familiarity with a source is not necessary to elicit same-group biases, and simply interacting with an individual that shares the same-group label as them resulted in a majority of children choosing the word provided by the similar informant. However, it is possible that children may still choose to learn from dissimilar informants over similar informants when provided with more information about the source.

As a result of potential practical implications, it needs to be noted that the generalizability of the findings is limited. First, the lack of a robust difference between groups may be the result of a limited sample size. The study made use of only 16 participants resulting in a small effect size. Additional data are being collected to ascertain if children's preference patterns hold. Second, our participant pool was relatively homogenous in terms of socio-economic status and ethnicity limiting the extent to which the findings can be generalized to more heterogeneous groups.

Future research looking at the role of labeling in categorization can examine children's

learning preferences when they are presented with additional information about the source. For example, learning that information from a source is accurate or inaccurate might disrupt this trend and modify children's social group biases.

In conclusion the findings of the study suggest that given no other information, social category labels by themselves, influence children's willingness to learn new information about novel objects from different sources. Children displayed a tendency to favor informants that shared the same category label as the one given to them, and preferred to learn new information from them.

Acknowledgements

I would like to give special thanks to my thesis supervisor Dr. Catharine Echols for her guidance and support. Additionally, I would also like to thank Dr. Arthur Markman, Nicole Wen, and the Language Development Lab members for their assistance.

References

Allport, G. W. (1954). The nature of prejudice. Reading, MA: Addison-Wesley

Baron, A. S., Dunham, Y., Banaji, M., & Carey, S. (2014). Constraints on the acquisition of social category concepts. *Journal of Cognition and Development*, *15*(2), 238-268. doi:10.1080/15248372.2012.742902

Bigler, R. S., Jones, L. C., & Lobliner, D. B. (1997). Social categorization and the formation of intergroup attitudes in children. *Child development*, *68*(3), 530-543.

Bigler, R. S., & Liben, L. S. (2007). Developmental intergroup theory explaining and reducing children's social stereotyping and prejudice. *Current Directions in Psychological Science*, *16*(3), 162-166.

Birnbaum, D., Deeb, I., Segall, G., Ben-Eliyahu, A., & Diesendruck, G. (2010). The development of social essentialism: The case of Israeli children's inferences about Jews and Arabs. *Child Development*, *81*(3), 757-777.

Chen, M. L., & Waxman, S. R. (2013). "Shall we Blick?" Novel words highlight actors' underlying intentions for 14-month-old infants. *Developmental psychology*, *49*(3), 426.

Cimpian, A., & Markman, E. M. (2009). Information learned from generic language becomes central to children's biological concepts: Evidence from their open-ended explanations. *Cognition*, *113*(1), 14-25.

Diesendruck, G., & haLevi, H. (2006). The role of language, appearance, and culture in children's social category-based induction. Child Development, 77, 539–553. doi:10.1111/j.1467-8624.2006.00889.x

Diesendruck, G., & Deblinger-Tangi, R. (2014). The linguistic construction of social categories in toddlers. *Child development*, *85*(1), 114-123.

Dunham, Y., Stepanova, E. V., Dotsch, R., & Todorov, A. (2014). The development of race-based perceptual categorization: Skin color dominates early category judgments. *Developmental Science*, doi:10.1111/desc.12228

Fawcett, C. A., & Markson, L. (2010). Similarity predicts liking in 3-year-old children. *Journal of experimental child psychology*, *105*(4), 345-358.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition*, *23*(3), 183-209.

Gelman, S. A. (2003). The essential child: Origins of essentialism in everyday life. New York: Oxford University Press.

Gelman, S. A. (2004). Psychological essentialism in children. *Trends in cognitive sciences*, *8*(9), 404-409.

Gelman, S. A., & Meyer, M. (2011). Child categorization. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(1), 95-105.

Graham, S. A., Nayer, S. L., & Gelman, S. A. (2011). Two-Year-Olds Use the Generic/Non-generic Distinction to Guide Their Inferences About Novel Kinds. *Child Development*, *82*(2), 493-507.

Heyman, G. D., & Gelman, S. A. (2000). Preschool children's use of trait labels to make inductive inferences. *Journal of Experimental Child Psychology*, *77*(1), 1-19.

Hirschfeld, L. (1996). Race in the making. Cambridge, MA: MIT Press.

Hollander, M. A., Gelman, S. A., & Raman, L. (2009). Generic language and judgements about category membership: Can generics highlight properties as central?. *Language and Cognitive Processes*, *24*(4), 481-505.

Johnson, M. A. (1989). Variables associated with friendship in an adult population. Journal of Social Psychology, 129, 379–390

Keates, J., & Graham, S. A. (2008). Category Markers or Attributes Why Do Labels Guide Infants' Inductive Inferences?. *Psychological Science*, *19*(12), 1287-1293.

Kinzler, K. D., Shutts, K., DeJesus, J., & Spelke, E. S. (2009). Accent trumps race in guiding children's social preferences. *Social cognition*, *27*(4), 623.

Kinzler, K. D., Shutts, K., & Correll, J. (2010). Priorities in social categories. *European Journal of Social Psychology*, *40*(4), 581-592.

Martin, C. L., & Fabes, R. A. (2001). The stability and consequences of young children's same-sex peer interactions. *Developmental psychology*, *37*(3), 431.

Master, A., Markman, E. M., & Dweck, C. S. (2012). Thinking in categories or along a continuum: Consequences for children's social judgments. *Child development*, *83*(4), 1145-1163.

Moore, B. N., & Parker, R. (1989). *Critical thinking: Evaluating claims and arguments in everyday life*. Mayfield Pub. Co.

Over, H., & Carpenter, M. (2012). Putting the social into social learning: explaining both selectivity and fidelity in children's copying behavior. *Journal of Comparative Psychology*, *126*(2), 182.

Powell, L. J., & Spelke, E. S. (2013). Preverbal infants expect members of social groups to act alike. *Proceedings of the National Academy of Sciences*, *110*(41), E3965-E3972.

Rothbart, M., & Taylor, M. (1992). Category labels and social reality: Do we view social categories as natural kinds?

Reyes-Jaquez, B., & Echols, C. H. (2013). Developmental differences in the relative weighing of informants' social attributes. *Developmental psychology*, *49*(3), 602.

Shutts, K., Banaji, M. R., & Spelke, E. S. (2010). Social categories guide young children's preferences for novel objects. *Developmental science*, *13*(4), 599-610.

Shutts, K., Roben, C. K. P., & Spelke, E. S. (2013). Children's use of social categories in thinking about people and social relationships. *Journal of Cognition and Development*, *14*(1), 35-62.

Waxman, S. R. (1999). The dubbing ceremony revisited: Object naming and categorization in infancy and early childhood. *Folkbiology*, 233-284.

Waxman, S. R. (2010). Names will never hurt me? Naming and the development of racial and gender categories in preschool-aged children. *European Journal of Social Psychology*, *40*(4), 593-610.

Yamauchi, T., & Markman, A. B. (2000). Inference using categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(3), 776.

Transcending Autobiography: Simultaneity and Meaning in Elisabet Ney's Lady Macbeth

Kendall DeBoer

The University of Texas at Austin

Undergraduate Research Journal

Volume 15
Number 1
Spring 2016

Critics often interpret Elisabet Ney's *Lady Macbeth* (1903) through a sentimental, romantic lens. As a "heartsick, frustrated, embittered" woman artist who defied contemporary gender norms, Ney too easily becomes Lady Macbeth.[1] *Lady Macbeth* reads as a self-portrait encapsulating the emotional adversity of Ney's experience as an estranged mother, as a misunderstood artist, and as a deviant in patriarchal society. Shakespeare's tragic heroine likewise defies traditional notions of maternity and femininity, so it seems Ney must have related to her. This reading is appealing; it is enticing to claim the sculpture is a cathartic self-portrait. While Ney's life as an 'eccentric' must have

influenced her artistic perspective, reducing *Lady Macbeth* to a self-portrait reflecting Ney's emotions strips the work of its complexities. *Lady Macbeth* operates within larger artistic currents of classicism and Romanticism, responds to a contemporary German revival of Shakespeare, and illuminates nineteenth-century gender and performance conventions; thus, the work requires more than simple biographical analysis.

Writers make the false assumption that biographical information about Ney accounts for all (or most) of the meaning in *Lady Macbeth*. "Possibly Miss Ney's enduring interest in Lady Macbeth was due in part to the sculptor's inability to get along with her son," [2] remarks Baldwin, while the Smithsonian web-page for the work says it "suggests the remorse and guilt the artist felt about her relationship with her estranged son."[3] The page goes on: "Ney sympathized with Lady Macbeth because they were both strong women who did not submit to the conventions of their times."[4] Cutrer contends the sculpture "reveals the private woman, a side [Ney] usually zealously guarded" who is "emotionally frail."[5] Source after source proposes emotional narratives about Ney's identification with Lady Macbeth as if they account for the work's significance. The sources choose to look at Ney's personal experiences rather than analyze her art. These comments demonstrate the critical tendency to reduce women's works to an expression of their personal experiences. One cannot discount Ney's biography altogether when analyzing *Lady Macbeth*, but restricting the critical conversation to emotional parallels between Ney's and Lady Macbeth's lives is reductive.

I acknowledge the temptation to apply Ney's biography to the work is attractive. Ney—or at least the "Ney" perceived by many—shares qualities with Lady Macbeth. Described as "eccentric, flamboyant, and unpredictable," Ney defied societal norms. Rejecting "children, church, and kitchen to crash a man's profession," Ney possessed "overwhelming ambition" and "refus[ed] to conform to any standards."[6] Although Ney quit her job as a sculptor for twenty years to raise her only child, writers tend to fixate on her unconventionality. Other adjectives associated with Ney include "heartsick, frustrated, embittered," "immoral," and "crazy," and people often remark that she existed in "a field wholly dominated by men."[7] These terms are not different from those used to describe Lady Macbeth; her femininity is "categorized as corrupted" and critics commonly note her "ambition" and "lack of compassion" which render her masculine.[8] Both women appeared anti-feminine within the social constructs of their times and places.

Notably, Lady Macbeth uses "infanticidal"[9] language, and equally notably, Ney had an estranged relationship with her son as she "failed miserably" at being a mother.[10] McKenzie's article is not unique in its emphasis on Ney's lackluster mothering skills. Whether Ney was actually a bad mother is unclear, but many writers choose to portray her as such. The two exhibit anti-maternal behaviors, further distancing themselves from expected "feminine" qualities. Both ambitious women insert themselves into male-centric spheres, both are described in terms of insanity, and both defy maternal pressures. Thus, biographical readings of *Lady Macbeth* charm many critics with good reason.

But consider verisimilitude, the most intuitive basis of comparison, and the tradition of declaring *Lady Macbeth* a self-portrait falters. In Austin, Texas's Elisabet Ney Museum, one can stand in front of Ney's marble self-portrait bust in an angle that allows one to view simultaneously the bust and the plaster version of *Lady Macbeth*. Viewing the two works, there is not much of a resemblance between the women's faces. The self-portrait features a much softer, rounder face, whereas *Lady Macbeth* has a strong, square jaw with a pointed chin (see figs. 1 and 2). The self-portrait has a small nose with a more obtuse curvature, whereas *Lady Macbeth*'s more angular nose protrudes further into space. The strong-boned, lantern jawed *Lady Macbeth* has a rough power in her features, while Ney's self-portrait seems more matronly or dainty. I preempt the refutation, "*Lady Macbeth* is an idealized portrait!" by pointing out that claim would also apply to her self-portrait. Both are idealized, yet they include distinctive differences in facial features. These formal differences between the two sculptures demonstrate the need for a more complex reading of *Lady Macbeth*; it seems Ney chose specific facial features to serve an expressive purpose.

Moreover, biographical readings do not account for many compelling aspects of Ney's sculpture, such as influences of classicism and Romanticism. Starting in the eighteenth century and enduring in popularity throughout Ney's life, neoclassicism uplifted "idealized, perfect forms.[11] Neoclassicism embodied the order, symmetry, and rationalism perceived as characteristic of classical, ideal forms. Because "the antique was regarded as the most reliable model" to evoke idealized forms, neoclassicists emulated the forms of Ancient Greco-Roman art.[12] Neoclassicist tendencies especially manifested in sculpture. "Sculpture has served, from the Renaissance, as a synecdoche of antiquity itself" and "it was regarded as a distillation of all the ancients had learned about nature and beauty."[13] Artists wanted to recreate the still, calm dignity of antique sculpture in order to forward what they considered manifestations of ideal beauty and perfect form.[14] Ney uses neoclassicism as a tool in her works, such as her sculpture of *Saint Sebastian* (see fig. 3) from 1857, who "stands upright much in the same stance a Polyclitus's *Spear-Bearer*, clutching two arrows in his right hand."[15] She employs the idealized body and face characteristic of neoclassicism in this work. Ney quotes currents of neoclassicism promoting rational and orderly forms.

Ney's also demonstrates her classical training in *Lady Macbeth*'s form. "Copying ancient models was an important part of her training, as it was for all sculptors throughout the nineteenth century, and it instilled a knowledge of neoclassical canons of taste."[16] She used this knowledge of neoclassical canons of taste as modes informing her artistic choices. Neoclassicists "emphasize the ideal and downplay the human, flawed in both physical and moral senses."[17] In *Lady Macbeth*, Ney appropriates this mode by featuring idealized, smooth skin and a circular balance enforced by Lady Macbeth's counterpoised feet, hands, and head (see fig. 4). Walsh writes of the neoclassicists:

> The technical armoury of this school of sculpture included waxy, frosty or lard-like renderings in marble of human skin, which smoothed out (or 'generalized') the messy irregularities of real flesh; a generalized or idealized approach to facial features and expressions, especially in allegorical works; a clearly and correctly drawn, elegant outline or contour (reminiscent of graceful, antique low relief or decorative vase effects) designed normally with a fixed viewpoint in mind; a fine, smooth, surface finish; idealized, harmonious anatomical proportions; references to classical architecture; and the use of carefully ordered, chiton-like draperies.[18]

Ney's *Lady Macbeth* illustrates this smoothed out, marble flesh as well as an idealized facial expression, a "fine, smooth, surface finish," and harmonious proportions. *Lady Macbeth*, although anguished, is proportional and balanced. She wears a gown that recalls an antique chiton, which falls off her left shoulder.

Neoclassicism maintained its popularity throughout the nineteenth century, but a competing movement of artistic practice arose in Romanticism. Romantics espoused the "ever-becoming, creative power of nature at one with, rather than merely expressive of, a divine or spiritual presence" and their works often feature "a markedly non-static oneness."[19] Rather than incorporating the rigidity and rationality of classicism, Romantic works illustrate abstract, emotional qualities. Accordingly, "the Romantic imagination thrived increasingly on the loosely executed, with a suggestion of incompleteness."[20] Because of this emphasis on looseness in execution, Romanticism is more common in painting. Still, "Sculptors found a number of ways in which to accommodate a Romantic concern with non-material spirit, derived in part from earlier Neoplatonic values, within the physical matter and naturalistic possibilities of their art."[21] Oftentimes, these Romantic qualities would manifest themselves in sculptures' subjects, such as is in Ney's 1867 work, *Prometheus Bound* (see fig. 5). Cutrer notes how the sculpture's position "opens the form [...] making it less self-contained than any of Ney's earlier work," and the "lack of grace and control [...] adds to the tension and strain of the piece."[22] Ney's choice to emphasize Prometheus's defiance reflects Romantic writings of Byron and Shelley.

Ney incorporates a Romantic sensibility in many of her sculptures. At different points in her career, Ney "grafted some of the characteristic of French romantic sculpture onto academic German form."[23] In spite of her classical training, Ney followed "true romantic fashion" and came to believe "the life of the individual was worthwhile in and of itself."[24] This emphasis on the individual reflects Romantic ideology. Romantic sculptures "included relatively secular suggestions of individual or heroic character fitting the notion of self-determining subjects; the expression of sentiments of political or national feelings; pronounced expressivity that took the form of violence, animation and drama."[25] Ney incorporates these qualities in *Lady Macbeth*. Ney's portrait of the heroine highlights her individual character, who is ambitious and autonomous. Her character is political in that she provokes her husband's ascent to power. The sculpture shows a "pronounced expressivity" in her pained facial expression (see fig. 6). Lady Macbeth undergoes torment as she

tries to wash metaphorical blood from her hands. A character of violence born from staged drama, the subject of *Lady Macbeth* illustrates these Romantic tenets almost too perfectly.

Contemporary artistic currents of classicism and Romanticism also might explain why Ney chose to sculpt a character from Shakespeare. Ney's psyche saw a "tension [...] between the Romantic impulse and classical training," and this tension between the two schools makes sense in application to a Shakespearean subject.[26] Her simultaneous affinities for classicism and Romanticism meet in *Lady Macbeth*, because the sculpture exemplifies nineteenth-century German inclinations towards both tendencies, and artists from both schools admired Shakespeare. Both intellectual groups contributed to a German Shakespearean revival. "A Shakespearean theme thus seemed an appropriate subject for her to undertake."[27] These contemporary German intellectual movements praising Shakespeare might account for some of the inspiration behind *Lady Macbeth*.

The works Schiller and Goethe attest to classicists' esteem for Shakespeare's works. These two are forerunners of Weimar Classicism, which emphasizes "human capacity for reason, forbearance, [and] fortitude."[28] Schiller's work "abounds with characters similar to Shakespearean characters" including Lady Macbeth.[29] Goethe claims that "nothing is so much Nature as Shakespeare."[30] These classicists express the larger German cultural sentiment in favor of Shakespeare. Proponents of order "the classicists approved of his use of ancient forms and subjects" as many of Shakespeare's plays reference ancient history, or allude to ancient poetry.[31] Classicists such as Schiller and Goethe contributed to a larger cultural atmosphere imbued with 'Shakespearomanie.'

European Romantics also lauded Shakespeare. "Shakespeare gave the Romantics all they were craving for: a world that confronted great kings with fools and destitute beggars, characters that were as inconsistent as real human beings are, combining melancholia and obsession, madness and high intellectuality, sublime goodness and grotesque evil."[32] Thus, the Romantics also added to the German Shakespeare obsession. "The Romantics, like Schlegel, could appreciate both Shakespeare's sense of history and his exploration of the dark side of man's nature."[33] Ney's Romantic impulse could have naturally led her to her Shakespearean heroine.

In the nineteenth-century, German intellectuals formed "an enthusiastic Shakespeare cult" and "combined admiration with identification and appropriation."[34] In fact, "Shakespeare became more and more ingrained into Germany's national culture and performances that were true to the original were more frequent in Germany than in England."[35] It would not have been out of place for Ney to explore Shakespearean texts in her artwork. One did not need not be a crazed, flamboyant, rebellious, anti-maternal woman to feel attracted to Shakespeare within the framework of nineteenth-century German culture.

Not only did *Lady Macbeth* contribute to a larger cultural discussion regarding classicism, Romanticism, and Shakespeare; the sculpture resonated with Ney's social circle. During and after the process of sculpting *Lady Macbeth*, many of Ney's friends began to relate to the

sculpture. Numerous women claimed to be the models for the work as they "came to have a personal stake in the statue" and "identify with it."[36] From the work's conception, her friends Maggie Williams, Anna Penny Backer, and Laura Bryan Parker wrote Ney to ask about it while Ella Dibrell funded it.[37] Lily Haynie said she posed for the arms and shoulders; Alma Tips said she posed for the body; Emma Reinli said the face was her face. Nannie Huddle also saw herself in the sculpture.[38] *Lady Macbeth* must have appealed to Ney's friends if so many wanted to associate themselves with it.

Ney's friends' claims are ironic. "The members of Elisabet Ney's circle were all well-educated members of upper- and middle-class society who, except for an occasional mild interest in woman suffrage, subscribed to their culture's notions of women's proper sphere."[39]

For example, Ella Dibrell "was an ambitious resourceful wealthy club woman [...] whose husband, Joseph Burton Dibrell, was a member of the Texas senate."[40] Fannie Huddle was wealthy enough to have the leisure time to visit Ney daily, and take occasion to bring "a great bouquet of wild bluebells, which she had driven into the country to look for."[41] Huddle, a socialite, attended events like "the public high school commencement" and would discuss these events over tea.[42] A photograph (see fig. 7) at the Elisabet Ney Museum depicts Ney having tea with Sam Houston's granddaughter and great-granddaughter at Formosa, circa 1892. Sam Houston, president of the Republic of Texas and U.S. Senator, had a family with considerable social clout. That Ney would share tea with women of such a high social standing demonstrates the degree to which Ney's circle was elite. Records of Ney and her friends partaking in social tea illustrate aristocratic ties, as in the nineteenth-century, tea signified aristocratic homes, elegant trifles, and socializing over delicacies.[43] There is little to suggest that Ney's friends deviated much from socially acceptable codes of conduct.

As established, Lady Macbeth embodies characteristics that oppose societal norms. "Unlike the ideal nineteenth-century woman, the Shakespearean character was ambitious, aggressive, and as a consequence, evil" so "they were identifying with a personality that totally contradicted [...] ideas of piety and submissiveness."[44] Cutrer presumes autobiography, and claims this identification of Ney's friends with the "evil" character is a result of their desire to identify with the sculptress herself.[45] I posit their identification with *Lady Macbeth* is more indicative of their gender roles (or their desire to escape those roles) as women in nineteenth-century Texas.

Nineteenth-century Texan women were neither expected nor allowed to be independent individuals. Instead, women had to operate within specific, familial, restrained roles. "Whether widowed or unmarried, single women were in an awkward position" because they could not be financially independent under law.[46] Thus, women had to exist within the context of family (usually husbands or male relatives) in order to have access to financial resources. Women attempting to establish independence "were generally unsuccessful."[47] Further, "educational opportunities for girls and teaching positions for women were limited."[48] In general, Texans did not push for women's education. "Women were perhaps more

active in the establishment and nurture of churches than schools," so Protestantism and piousness took precedence over schooling.[49] Interestingly, religion provided one of the only opportunities for Texas women to for relationships with one another and influence events outside of their families, as they were otherwise barred from the public sphere.[50] Still, this platform for public expression had its limitations in that it was within the constraints of patriarchal church ideology. In Texas, women "were valued as wives, mothers, helpers, supporters, and slave laborers," and they had to live within the confines of those roles.[51] Ultimately, nineteenth-century Texan women had to adhere to strict gender norms of dependence, ignorance, piousness, and subservience.

Ney's *Lady Macbeth* is not an ideal wife, mother, helper, supporter, or slave laborer. In fact, the sculpture flouts these conventions of Texan womanhood. Textually, Lady Macbeth drives her husband to murder the king, and she uses infanticidal language. She ranks her ambition above the emotions and wellbeing of her husband, and becomes delusional and guilt-ridden because of her murderous actions. Before King Duncan's murder, Lady Macbeth calls upon spirits in a famous soliloquy that illustrates her status as antithesis to the nurturing, maternal ideal:

> The raven himself is hoarse
> That croaks the fatal entrance of Duncan
> Under my battlements. Come you spirits
> That tend on mortal thoughts, unsex me here,
> And fill be from the crown to the toe, top-full
> Of direst cruelty. Make thick my blood,
> Stop up th'access and passage to remorse,
> That no compunctious visitings of nature
> Shake my fell purpose, nor keep peace between
> Th'effect and it. Come to my woman's breasts,
> And take my milk for gall, you murdering ministers,
> Wherever, in your sightless substances,
> You wait on nature's mischief. Come thick night,
> And pall thee in the dunnest smoke of hell,
> That my keen knife see not the wound it makes,
> Not heaven peep through the blanket of the dark
> To cry, 'Hold, hold.' (I.v.38-54)

Calling upon spirits is suggestive of invoking demons or devils. She asks of them to "unsex" her, or strip her of her feminine qualities, so she might instead be full of "direst cruelty." She hopes her thickened blood will prevent the "compunctious visitings of nature" that might conjure feelings of remorse or penitence. She asks for gall, a bitter poison, to replace her breasts' milk. This passage illustrates Lady Macbeth's rejection of her prescribed feminine role so she can enact her ambition. Her independence, aggression, and lack of compassion defy both the gender conventions of her time and nineteenth-century Texan feminine ideals.

Yet Ney creates her sculpture from the scene in which Lady Macbeth has completely collapsed into madness and grief. Though Lady Macbeth has control over her aggression early in the plot, the "Out damned spot!" scene demonstrates a guilt-wracked, troubled woman who later commits suiccide. Though Lady Macbeth no longer possesses her powerful composure and cruel calculation from earlier acts, this scene continues to epitomize "un-womanly" qualities, because Lady Macbeth remains expressive. In Act V scene I of *Macbeth*, Lady Macbeth wrings her hands, washing them repeatedly in order to cleanse "a spot." She mumbles about the murder she provoked her husband to commit:

> Hell is murky. Fie, my lord, fie, a soldier and afeared? What need we fear? Who knows it when none can call our power to account? Yet who would have thought the old man to have had so much blood in him? (V.i.36-40)

These questions affirm her self-authenticating tyrannical rule brought about by her and her husband's actions. The last question indicates the violence and cruelty of the murder committed. In remarking on the murkiness of Hell, Lady Macbeth shows an awareness of the immorality of her murder, which certainly contrasts with the pious subservience expected of 19th-century Texan women. Lady Macbeth fixates on her eternal damnation throughout the scene, asking "What, will these hands ne'er be clean?" (V.i.43). She remarks "All the perfumes of Arabia will not sweeten this little hand" (V.i.51) and she states "What's done cannot be undone" (V.i.68), further iterating that her behavior is so damnable that she is beyond salvation. This scene typifies Lady Macbeth's madness and violence, which are two qualities in stark contrast to the docile, submissive feminine ideal.

If Lady Macbeth departs from nineteenth-century feminine norms, why would Ney's well-behaved, wealthy, socialite friends see themselves in her depiction? In her argumentation for biographical reading, Cutrer establishes that in Ney's friends' accounts, they "agreed that Ney was an unusual woman ill-suited to the community she lived in, one who offended most of her neighbors with her customs and habits," but they placed the blame on "those who were critical and suspicious."[52] This criticism levied by Ney's friends against the judgmental environment demonstrates that although they were members of the social elite, they may not have supported the social constraints placed on women. Cutrer believes Ney's friends "derived a certain vicarious fulfillment from identifying with [Ney,] a woman who did things they did not dare to do."[53] Just as easily, then, could Ney's friends derive that fulfillment from identifying with Lady Macbeth as an independent character. Lady Macbeth also "did things they did not dare to do," whether while sane or insane. It is possible that Lady Macbeth captured the imaginations of nineteenth-century aristocratic women because she is expressive. They could connect with both her ambition and her emotive qualities. Because the scene is an expression of despair, Ney's friends may have been able to relate because of their trapped positions in Texan high society. Additionally, theatrical iterations of madwomen were some of the only venues through which women could exhibit 'mascu-

line' strength and expression on stage. Lady Macbeth might have offered an emotional and psychological role unlike any Ney's friends' had previously known.

Ney's friends were able to impose their own ideas of themselves onto *Lady Macbeth* because the sculpture functions as a generic theatrical type. Late nineteenth-century theatrical studies involved learning universalizing gestures to communicate certain emotions or ideas. Francois Delsarte popularized his acting style that related interior emotions to systematic gestures. These gestures provided a vocabulary for emotional and artistic expression through the body. Anna Morgan explains the importance of gesture in her book on Delsarte's method:

> Gesture is the immediate revelation of the being. The most evanescent flashes of thought and temperament are first and immediately revealed in gesture. The quality of this gesture is the quality of the individual, and the touchstone to the character of the man. [...] Gesture in the language of nature, and is comprehensible to people of every tongue.[54]

These notions of gesture as revelatory of inner qualities were widespread and well-known. Morgan goes on to describe Gilbert Austin's *Chironomia* as "the most important work as yet produced on the subject of gesture."[55] Interestingly, *Chironomia* features an illustration of a complex, significant gesture Austin calls "agony of mind" (see fig. 8).[56] This gesture shows a woman wringing her hands to the right of her body, tossing her head to the left. The gesture is uncannily similar to that of Ney's *Lady Macbeth*. Considering Ney's subject is inherently theatrical (as it is a character from a Shakespearean tragedy), it is not farfetched to believe its pose would play into gestural conventions of its time. Also, "agony of mind" is an appropriate emotional state to attribute to Lady Macbeth in the depicted scene, as she attempts to scrub her 'bloodied' hands. In this view, *Lady Macbeth* becomes a theatrical type rather than a portrait of a specific model. And this type furthered Ney's friends' ability to inhabit a more complex role than the roles society had allotted them.

Although taking biography into consideration is important to understanding Ney's *Lady Macbeth*, this aspect only accounts for a part of the work's significance. *Lady Macbeth* is more than an autobiographical, inward-bent Ney epitaph, as the work reflects contemporary cultural and social trends in nineteenth-century Germany and Texas. In *Lady Macbeth*, I can trace Ney's incorporation of neoclassicism and Romanticism; I can see the nineteenth-century German revival of Shakespeare; I can note its relationship to contemporary gender and theatrical conventions. *Lady Macbeth* is more than a self-portrait, as Ney borrows from and participates in conversation with major cultural codes to create a complex work that transcends simplistic reductions.

Bibliography

Austin, Gilbert. *Chironomia; or, a Treatise on Rhetorical Delivery*. London: W. Bulmer, and company, 1806.

Baldwin, Florence. "Elisabet Ney: Texas' First Eminent Sculptor." *Austin American Statesman* (Austin), January 16, 1972, SMB1.

Cutrer, Emily Fourmy. *The Art of the Woman: The Life and Work of Elisabet Ney*. Lincoln: U of Nebraska P, 1988.

Downs, Fane. "'Tryels and Trubbles': Women in Early Nineteenth-Century Texas." *Southwestern Historical Quarterly* 90, no. 1 (1986): 35-56.

Grundmann, Heike. "Shakespeare and European Romanticism." In *Companion to European Romanticism*, edited by Michael Ferber, 29-48. N.p.: Wiley, 2008.

Loggins, Vernon. *Two Romantics and their Ideal Life: Elisabet Ney, sculptor; Edmund Montgomery, philosopher*. New York: Odyssey P, 1946.

McKenzie, John P. "Elisabet Ney: Texas Legend." *Austin American Statesman* (Austin), August 4, 1963, D1.

Morgan, Anna. *An Hour with Delsarte ; A Study of Expression*. N.p., 1890.

Phillips, Chelsea. "'Unsex Me Here': Bodies and Femininity in the Performance History of Lady Macbeth." *Testi e Linguaggi* 7 (2013): 353.

Pratt, James Norwood. "Tea (Meal)." In *Encyclopedia of Food and Culture*, edited by Solomon H. Katz, 394. Vol. 3. New York: Charles Scribner's Sons, 2003.

Shakespeare, William. *Macbeth*. Edited by Sandra Clark and Pamela Mason. 3rd ed. New York: Bloomsbury Arden Shakespeare, 2015.

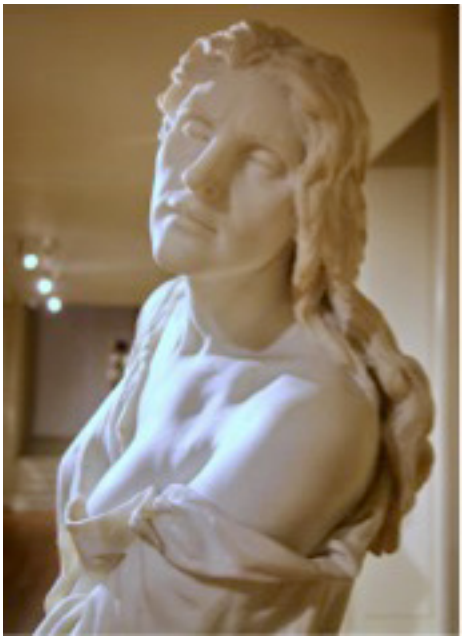
Smithsonian Institution. "Lady Macbeth." Smithsonian American Art Museum. http://americanart.si.edu/collections/search/artwork/?id=18677.

Walsh, Linda. "The 'Hard Form' of Sculpture: Marble, Matter and Spirit in European Sculpture from the Enlightenment through Romanticism." *Modern Intellectual History* 5, no. 3 (2008): 455-86.

"Weimar Classicism." In *The Oxford Encyclopedia of Theatre and Performance*, edited by David Kennedy, 1440. Vol. 2. New York: Oxford UP, 2015.

Figure 1: Elisabet Ney, *Self-Portrait* (late 19th century).
Figure 2: Elisabet Ney, detail from *Lady Macbeth* plaster (1902).
Figure 3: Elisabet Ney, *Saint Sebastian* (1857).

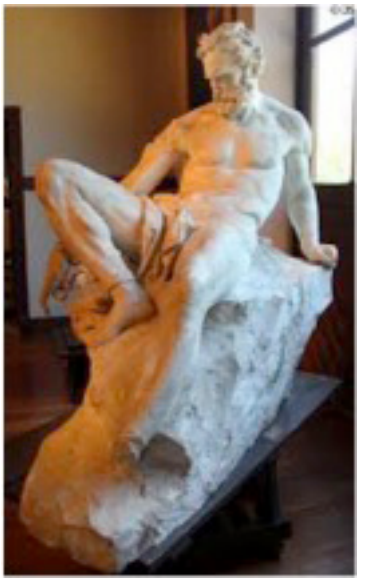
1.



2.

3.

Figure 4: Elisabet Ney, *Lady Macbeth* (1903).
Figure 5: Elisabet Ney, *Prometheus Bound* (1867).
Figure 6: Detail from *Lady Macbeth*.

4. 

5. 

6. 

Figure 7:	Photograph from the Elisabet Ney Museum showing tea-time with Sam Houston's granddaughter and great-granddaughter (ca. 1892).
Figure 8:	Illustration from Gilbert Austin's *Chironomia.* Complex significant gestures from left to right: Doubt first, Doubt second, Grief, Agony of mind (1806).

7.

8.

Endnotes

1 John P. McKenzie, "Elisabet Ney: Texas Legend," *Austin American Statesman* (Austin), August 4, 1963, D1.
2 Florence Baldwin, "Elisabet Ney: Texas' First Eminent Sculptor," *Austin American Statesman* (Austin), January 16, 1972, 5.
3 "Lady Macbeth," Smithsonian American Art Museum, http://americanart.si.edu/collections/search/artwork/?id=18677.
4 Ibid.
5 Emily Fourmy Cutrer, *The Art of the Woman: The Life and Work of Elisabet Ney* (Lincoln: U of Nebraska P, 1988), 213.
6 Baldwin, "Elisabet Ney: Texas' First," 4.
7 McKenzie, "Elisabet Ney: Texas Legend," D1.
8 Chelsea Phillips, "'Unsex Me Here': Bodies and Femininity in the Performance History of Lady Macbeth," *Testi e Linguaggi* 7 (2013): 1.
9 Ibid.
10 McKenzie, "Elisabet Ney: Texas Legend," D1.
11 Linda Walsh, "The 'Hard Form' of Sculpture: Marble, Matter and Spirit in European Sculpture from the Enlightenment through Romanticism," *Modern Intellectual History* 5, no. 3 (2008): 456.
12 Ibid., 465.
13 Ibid.
14 Ibid., 466.
15 Cutrer, *The Art of the Woman*, 15.
16 Ibid., 14.
17 Walsh, "The 'Hard Form' of Sculpture," 472.
18 Ibid.
19 Ibid., 476.
20 Ibid., 479.
21 Ibid., 481.
22 Cutrer, *The Art of the Woman*, 80.
23 Ibid., 54.
24 Ibid., 62.
25 Walsh, "The 'Hard Form' of Sculpture," 482.
26 Cutrer, *The Art of the Woman*, 212.
27 Ibid.
28 "Weimar Classicism," in *The Oxford Encyclopedia of Theatre and Performance*, ed. David Kennedy (New York: Oxford UP, 2015).
29 Heike Grundmann, "Shakespeare and European Romanticism," in *Companion to European Romanticism*, ed. Michael Ferber (n.p.: Wiley, 2008), 36.
30 Ibid.
31 Cutrer, *The Art of the Woman*, 212.
32 Grundmann, "Shakespeare and European Romanticism," in *Companion to European Romanticism*, 30.
33 Cutrer, *The Art of the Woman*, 212.
34 Grundmann, "Shakespeare and European Romanticism," in *Companion to European Romanticism*, 34.
35 Ibid., 39.

36 Cutrer, *The Art of the Woman*, 214.
37 Ibid.
38 Ibid., 215.
39 Ibid., 216.
40 Vernon Loggins, *Two Romantics and their Ideal Life: Elisabet Ney, sculptor; Edmund Montgomery, philosopher* (New York: Odyssey P, 1946), 295.
41 Ibid., 341.
42 Ibid., 342.
43 James Norwood Pratt, "Tea (Meal)," in *Encyclopedia of Food and Culture*, ed. Solomon H. Katz (New York: Charles Scribner's Sons, 2003), 3: 394.
44 Cutrer, *The Art of the Woman*, 216.
45 Ibid., 217: "Whether they realized it or not, they too had come to see *Lady Macbeth* not as a character from some remote drama, but as a representation of its creator. By identifying with the statue, they were also identifying with Elisabet Ney."
46 Fane Downs, "'Tryels and Trubbles': Women in Early Nineteenth-Century Texas," *Southwestern Historical Quarterly* 90, no. 1 (1986): 41.
47 Ibid.
48 Ibid.
49 Ibid., 43.
50 Ibid., 46.
51 Ibid., 55.
52 Cutrer, *The Art of the Woman*, 217.
53 Ibid., 218-219.
54 Anna Morgan, *An Hour with Delsarte ; A Study of Expression* (n.p., 1890), 58.
55 Ibid., 63.
56 Gilbert Austin, *Chironomia; or, a Treatise on Rhetorical Delivery* (London: W. Bulmer, and company, 1806), 621.

Comparison of Microbial Diversity in Local Wasps and Plant Surfaces

The University of Texas at Austin

Undergraduate Research Journal

Volume 15
Number 1
Spring 2016

Dylan Fall
Jo-anne Holley

Abstract

Symbiotic organisms can play important roles in the biology of their hosts, as they can assist in nutrition, detoxification, and health. The purpose of this study was to identify bacteria and yeast that have developed symbiotic relationships with local wasp species. This was done by comparing the bacteria and yeast diversity in the guts of 8 nectar-feeding wasps to the species found on the flowers they visit. The flowers were washed in PBS buffer and plated on YPD media, as were the gut contents of the wasps. Bacterial and yeast colonies were isolated and identified by using PCR to amplify DNA fragments 16S and

ITS respectively. These DNA sequences were then compared to pre-existing sequences in the NCBI BLAST database. While there was occasional overlap of species between the two sources, the majority of species found were exclusive to either the wasp gut flora or the plant surface. Of the 35 species grown in culture, 17 were found only on plant surfaces and 12 found strictly in the guts of wasps. Three species were found in both, they were: *Kosakonia cowanii*, *Rosenbergiella epipactidis*, and *Asaia bogorensis*. These findings indicate that wasps foster a microbial community different to the communities they encounter on plants, suggesting some selection in which bacteria establish populations in their guts.

Introduction

The microbiota within animals, meaning the community of unicellular organisms living inside an animal, is integral to the host's function and health. These microorganisms and the animal mutually benefit each other; the organisms provide numerous benefits to the host while the host provides a safe place for the organism to thrive and receive nourishment. In humans, the complex communities of helpful bacteria within the digestive system aid in breaking down of food matter, nutrient uptake, and fighting off pathogens. Microbiomes are different and diverse among all clades of animal life, as these microbial communities adapt and evolve alongside their host (Ruth et al., 2008). Insects in particular, due to the high amount of speciation and specialization within the class, have a wealth of potential in terms of discovery of diverse microbial communities (Krishnan et al., 2014). By studying these complex, species-specific biomes, we can identify how these organisms help each other and interact (Brucker and Bordenstein, 2011).

The microbiota of insects has been found to have diverse reasons for investigation and study. Not only can it help us understand the different symbiotic relationships that occur between the insects and the diverse members of their gut flora, the discovery of these novel species may lead to the development of new antibiotics (Poulsen *et al.* 2011). In addition, knowledge of how the microbiome functions in pollinators such as bees can help us move forward in a thoughtful manner in fields such as agriculture. We have recently seen how certain pesticides can affect pollinator function, such as the effect of neonicotinoid pesticides on honey bee brains (Williams et al., 2015). Further research may protect pollinators from similar unforeseen damaging effects of agricultural products.

Most microbiomes tend to stay the same throughout an insect's life cycle due to the balance of gut flora being crucial for functions such as digestion, nitrogen fixation, and nutrient synthesis (Fukatsu 2012). However, the way these communities are formed can differ based on the species. There are two types of microbial transmission, vertical and horizontal, and both methods play important parts in the formation of insect microbiota.

Vertical transmission is the transmission of microbial communities from one generation to the next. This has a large impact on the gut environment, as the adults with advantageous gut flora transfer their microbiome to their offspring in order to protect them from various

bacterial and fungal pathogens. This transfer can serve several important functions for the species, such as enabling them to digest and receive nutrients from the local food supply (Kaltenpoth *et al.* 2005) or helping to ensure that they are able to successfully reproduce through increased fitness and color change (Douglas, 1998). Horizontal transfer of environmental factors, such as the insect's diet and habitat, can also have an impact on microbiome assembly. For nymphs of the bean bug *Riptortus pedestris, Burkholderia* bacterium that is picked up from the soil protects them from an otherwise deadly pesticide (Kikuchi *et al.* 2012). Often, these two methods of transmission combine, such as in social bees. The combination of bees transferring helpful microbes to their larvae and the transfer of microbes through social interaction results in a more diverse microbiome (*Martison et al.* 2010).

One aspect of the microbiome that has gone largely uninvestigated is the fungal inhabitants of the gut. Although they have been generally unexplored, yeast mutualists have symbiotic abilities comparable to their bacterial counterparts. The strains of yeast found in different species of the fly genus *Drosophila* were impacted greatly by the flies' environment and diet, suggesting that yeasts are picked up from the environment rather than vertically transmitted (Chandler *et al*. 2012). An example of a known insect-yeast mutualisms is the relationship between the ambrosia beetle and basidiomycotan fungus, which utilize the fungus' enzymatic function to colonize and consume rotting wood (You et al., 2015). Additionally, a common disease in the insect *Megachile rotundata* known as chalkbrood kills the host by suppressing the growth of other fungal and bacterial communities, suggesting that these other microbial communities have a beneficial effect on the host (McFredrick et al., 2014).

There is some research done on how yeast relies on insect species such as wasps, not a lot is known about the relationship between yeast and insects. Fungi, including yeasts, can live within living organisms, but are also found in substances such as nectar and on plant surfaces. Plants, however, go dormant during the winter months, which is a problem for the continued survival of the yeast that rely on them. Therefore, the yeast species are harbored within insects such as bumblebees and wasps until they can be transferred to new plant life in the spring (Stefanini *et al.* 2012). Yeasts are ubiquitous on most plant surfaces, and factors such as nectar pH and sucrose concentration have little effect on yeasts abundance (Brysch-Herzberg 2004). This suggests that the insect emergence and pollination habits may influence the presence, abundance, and diversity of yeasts. Although yeasts can survive in other environments other than the wasp digestive tract, the yeasts become a part of the gut flora to not only survive the winter, but thrive by being transported to numerous attractive flowers where other insects will pick the strain up as well (Brysch-Herzberg 2004). For example, a relationship has been found between the dominant strains of yeast found in a certain species of wasps and the yeasts found in the wine and beer of that area, indicating that the yeasts are able to survive through the dormant season inside the wasp microbiome (Stefanini *et al.* 2012).

Not a lot is known about the yeast species diversity in wasp gut flora, as well as the benefits received by the wasp through carrying the yeast. There is some evidence that bumblebees pollinating and gathering nectar prefer flowers with large concentrations of yeast (Herrera *et al.* 2013). Through monitoring the behavior of bumble bees after receiving nectar with yeast, it was discovered that yeast has the ability to "modify pollinator foraging patterns, pollination success, and the quantity and quality of seeds produced by insect-pollinated plants" (Herrera *et al.* 2013). While there is clear benefit for the plant as well as the yeast, further study is needed to identify benefits, if any, the yeast provide wasps and other pollinators. Identifying the diversity of yeast species, as well as where they are located in the microbiome, will reveal important details as to what the yeast's function might be.

While many cases of yeast/plant mutualisms are known, other microbial mutualisms exist in plants. A prominent one is the relationship between plants and *Bradyrhizobium* bacteria, which aids the plant in nitrogen fixation (Kulkarni et al., 2015). Knowing that plants have the potential for both fungal and bacterial mutualisms, we can see that insects such as wasps can pick up a variety of microbes from their environment, which could potentially affect their microbiota.

Objectives

The main research objective is to survey the diversity and abundance of yeast and bacteria in the wasp microbiota. In addition, the microbial diversity in the wasp gut flora will be compared with the microbial communities found on the flowers of local plant fauna. The yeast communities on the local plant fauna are expected to be equally or more diverse compared to the bacterial community in wasp gut flora. Furthermore, the yeast species diversity is expected to resemble the yeast communities of the wasps' habitat. Through extraction protocols such as phenol-chloroform extractions, DNA samples will be obtained from plated out microbe colonies obtained from the digestive tract of wasps. This process will be done several times with wasps of various species in the local area, in an effort to get a complete picture of the diversity of the vespid gut flora.

Methods

Specimens were collected at Hornsby Bend, located in Travis County, Texas (Lat: 30.228, Lon: -97.633) on July 7, 2015. Gut samples from were removed from eight wasps of various species using flame-sterilized forceps and plated out on both a Yeast Extract-Pepto-Dextrose Medium (YPD) as well as an YPD / antibiotic mix medium (consisting of Ampicillin 100µg/ml and Chloramphenicol 150 µg/ml), in an attempt to view the full diversity of bacterial as well as fungal communities. Plant specimens, composed of six samples of *Monarda citriodora* (lemon bee balm) and four samples of *Rudbeckia hirta* (black-eyed susans), were washed with 100 µl of phosphate buffered saline, and then the buffer was plated in a similar manner. After a growth period of approximately 20 hours in a standard aerobic incubator, colonies were

isolated and the DNA extracted using a Phenol-chloroform protocol (see Xiao 2006), where samples were suspended in 200µl of lysis buffer, underwent 2 cycles of -80˚C for 2 minutes and 95˚C water bath for 1 minute, vortexed for 30 seconds, and then combined with 200µl of phenol-chloroform. Then, samples were vortexed for 2 minutes, centrifuged for 3 minutes at 13,200 rpm and 4˚C. The aqueous layer of solution was then pipetted into 400µl of ice-cold ethanol and allowed to precipitate for 5 minutes. Samples were then centrifuged for 5 minutes at 13,200 rpm and room temperature. Supernatant was poured out and samples were washed with 1ml 75% ethanol, allowed to air dry, then resuspended in 20 µl ddH2O. Standard PCR protocol with Taq DNA Polymerase (see protocol by NEB) was used for both primers, with an initial denaturation phase of 95˚C for 2 minutes, followed by 35 cycles of 95˚C for 10 seconds, 54˚C for 20 seconds, 68˚C for 1 minutes and 10 seconds, and a final extension phase of 68˚C for 5 minutes. For bacterial identification, the 16S region was targeted using 27F (5□-AGAGTTTGATCATGGCTCAG) and 1492R (5□-TACCTTGTTACGACTT) primers. For fungal identification, the ITS region was targeted using ITS1 (5'- TCCGTAG-GTGAACCTGCGG) and ITS4 (5'-TCCTCCGCTTATTGATATGC) primers. PCR product quality was assessed using Agarose Gel Electrophoresis and UV Imaging (see Lee *et al.,* 2012), and viable samples were sent to the UT Austin DNA Sequencing Facility on an Applied Biosystems 3730 DNA Analyzer. Sequences were identified by comparison to the NCBI BLAST Database (see Madden 2002).

Results

After sequencing of the targeted DNA regions 16S and ITS, 20 members of the microbial community on plant surfaces and 15 members of the microbial community in the wasp guts were successfully identified . There was not a great deal of overlap in the two communities, but 3 species were found both in wasp gut flora and plant surface.

The microbial species that were found to occupy the explored plant surfaces included bacterial samples such as: 3 species of *Pseudomonas,* 3 species of *Pantoea*, 2 species of *Rosenbergiella,* 2 species of *Microbacterium,* 2 species of *Cronobacter* and 1 species each of the following genera: *Asaia, Erwinia, Neokomagatae, Sphingomonas,* and *Xanthomonas*. In addition, two yeast species was found, found to be a species of the genus *Fusarium* and *Metschnikowia*.

The microbial composition found within the microbiome of wasps were found to differ greatly with that of plant surfaces, though there was some overlap. The bacterial samples were identified as: 2 species of *Bacillus,* 2 species of *Staphylococcus*, 2 species of *Pantoea* (distinct from *Pantoea* found on plant surfaces), and 1 species from each of the following genera: *Fructobacillus, Kosakonia, Stenotrophomonas, Swingsia, Rosenbergiella, Asaia,* and *Aneurinibacillus.* Two yeast species were found as well, from genera *Saccharomyces* and *Candida.*

Discussion

The finding that wasp microbial communities share a select amount of members with the surrounding plant environment is thought-provoking, and leaves room for a great deal of additional research to be done in this area. As we have seen, 3 species were found to exist in both the wasp microbiome and the plant surface: *Asaia bogorensis, Rosenbergiella epipactidis,* and *Kosakonia cowanii*. *Rosenbergiella* is known to be commonly found in nectar (Lenaerts *et al.*, 2014) and *Kosakonia* strains has been found to exist in some plants as an endophytic growth-promoting bacteria (Ming *et al. 2015)*. Since no evidence of these microbes interacting with insects is present within the current literature, it is not likely (though also not impossible) that these microbes play a large role in the wasp microbiota. *Asaia bogorensis*, however, has been found as a stable associate within the gut flora of the Asian mosquito (Fatvia *et al.,* 2007). Knowing that this bacterial strain is a symbiont within the gut flora of another insect, continued discovery of *Asia borgorensis* may lead to more investigation in the future as to the role and nature of *Asaia bogorensis* within the wasp microbiota.

In this study, the few yeasts that were found were diffuse and did not match up with what previous literature had shown to be expected from the wasp microbiota (Stefanini *et al.*, 2012). In addition, the microbial communities were extremely varied between the wasp microbiota and the plant surface, excluding the 3 shared species. One reason this may have occurred would be the anaerobic conditions of some sections of the wasp gut where microbes might populate. While the air composition within the wasp gut have not been extensively characterized, it is known that within the insect gut, there are varied levels of O_2 and CO_2 found within the various sections (Engel and Moran, 2013). Because all the microbes in the study were grown in O_2-prevalent environments, it is possible that a sizable fraction of the microbiota was excluded. It could also signify that the wasp immune system, or the bacteria that lives within the gut, is suppressing the growth of environmental bacteria (Eleftherianos *et al.*, 2013). Regardless of the cause, further study is needed to identify the role of these microbes within the wasp microbiome. Along with the absence of anaerobic plating in this study, the sample size of wasps was on the lower end, resulting in decreased representation of the wasp microbiome when compared to the environmental microbe communities. In future studies, these areas will be of utmost importance to ensure that the study succeeds in obtaining a full and comprehensive picture of the wasp microbiota.

References

Brucker, R.M. and Bordenstein, S.R. (2011).The roles of host evolutionary relationships (genus: *Nasonia)* and development in structuring microbial communities. *Evolution.* 66, 349-362.

Brysch-Herzberg, M. (2004) Ecology of yeasts in plant-bumblebee mutualism in Central Europe. *FEMS Microbiology Ecology*. 50, 87-100.

Chandler, J. A., Eisen, J. A. & Kopp, A. (2012). Yeast communities of diverse Drosophila species: comparison of two symbiont groups in the same hosts. *Applied Environmental Microbiology*. 78, 7327–7336.

Douglas, A. E. (1998) Nutritional Interactions in Insect-Microbial Symbioses: Aphids and Their Symbiotic Bacteria *Buchnera. Annual Review of Entomology.* 43: 17-37

Eleftherianos I, Atri J, Accetta J, Castillo JC. Endosymbiotic bacteria in insects: guardians of the immune system? *Front Physiol.* 4: 46.

Engel P, Moran N A. (2013) The gut microbiota of insects – diversity in structure and function. *FEMS Microbiology Reviews.* 37 (5).

Favia G, Ricci I, Damian C, Raddadi N, Crotti E, Marzorati M, Rizzi A, Urso A, Brusetti L, Borin L, Mora D, Scuppa P, Pasqualini L, Clementi C, Genchi M, Corona S, Negri I, Grandi G, Alma A, Kramer L, Esposito F, Bandi C, Sacchi L, Daffonchio D. (2007) Bacteria of the genus *Asaia* stably associate with *Anopheles stephensi*, an Asian malarial mosquito vector. *Proc Natl Acad Sci U S A*. 104 (21): 9047-51.

Fukatsu, T. (2012) Next-generation sequencing sheds light on intricate regulation of insect gut microbiota. *Molecular Biology*. 21, 5908-5910.

Herrera, C.M., Pozo, M.I., Medrano, M. (2013) Years in nectar of an early-blooming herb: sought by bumble bees, detrimental to plant fecundity. *Ecology*. 94, 273-279.

Krishnan M, Bharathiraja C, Pandiarajan J, Prasanna A. V, Rajendhran J, Gunasekaran P. (2014) Insect gut microbiome - An unexploited reserve for biotechnological application. *Asian Pac J Trop Biomed.* 4: 16-21.

Kaltenpoth, M., Gottler, W., Herzner, G., Strohm, E. (2005) Symbiotic Becteria Protect Wasp Larvae from Fungal Infestation. *Current Biology*. 15, 475-479.

Kulkarni G, Busset N, Molinaro A, Gargani D, Chaintreuil C, Silipo A, Giraud E, Newman D.K. (2015) Specific Hopanoid Classes Differentially Affect Free-Living and Symbiotic States of *Bradyrhizobium diazoefficiens*. *MBio.* 6: 5.

Lenaerts M, Alvarez-Pérez S, de Vega C, Van Assche A, Johnson SD, Willems KA, Herrera CM, Jacquemyn H, Lievens B. Rosenbergiella australoborealis sp. nov., Rosenbergiella collisarenosi sp. nov. and Rosenbergiella epipactidis sp. nov., three novel bacterial species isolated from floral nectar. *Syst Appl Microbiol*. 37 (6): 402-11.

Lee PY, Costumbrado J, Hsu CY, Kim YH. (2012) Agarose gel electrophoresis for the separation of DNA fragments. *J Vis Exp.* 62.

Madden, Tom. (2002) The BLAST Sequence Analysis Tool. *The NCBI Handbook*.

Martinson, V.G., Moy, J. and Moran, N. (2012) Establishment of characteristic gut bacteria during development of the honeybee worker. *Applied and Environmental Microbiology*. 78, 2830–2840.

McFrederick QS, Mueller UG,James RR. (2014) Interactions between fungi and bacteria influence microbial community structure in the *Megachile rotundata* larval gut. Proc. R. Soc. B 281: 20132653.

Meng X, Bertani I, Abbruscato P, Piffanelli P, Licastro D, Wang C, Venturi V. (2015) Draft Genome Sequence of Rice Endophyte-Associated Isolate *Kosakonia oryzae* KO348. *Genome Announc.* 3(3); 594-15.

Poulsen, M., Oh, D., Clardy, J., Currie, C.R. (2011) Chemical Anayses of Wasp-Associated *Streptomyces* Bacteria Reveal a Prolific Potential for Natural Products Discovery. *PLoS ONE*. 6.

Roulis M, Bongers G, Armaka M, Salviano T, He Z, Singh A, Seidler U, Becker C, Demengeot J, Furtado GC, Lira SA, Kollias G. (2015) Host and microbiota interactions are critical for development of murine Crohn's-like ileitis. *Mucosal Immunol.*

Ruth E. L, Micah H, Catherine L, Peter T Rob Roy R, J. Stephen B, Michael L. S, Tammy A. T, Mark D. S, Rob K, Jeffrey I. G. (2008) Evolution of mammals and their gut microbes. *Science.* 320(5883): 1647–1651.

Stephanini, I., Dapporto, L., Legras, J., Calabretta, A., Di Paola, M., De Filippo, C. Viola, R. Capretti, P., Polsinelli, M., Turillazzi, S., Cavalieri, D. (2012) Role of social wasps in Saccharomyces cerevisiae ecology and evolution. *Proceedings of the National Academy of Sciences*. 109, 13398–13403.

Xiao, Wei. (2006) Yeast Protocols. *Methods in Molecular Biology*. 17-18.

You L, Simmons D. R, Bateman C. C, Short D. P, Kasson M. T, Rabaglia R. J, Hulcr J. (2015) New Fungus-Insect Symbiosis: Culturing, Molecular, and Histological Methods Determine Saprophytic Polyporales Mutualists of Ambrosiodmus Ambrosia Beetles. *PLoS ONE.* 10: 9.

Appendix

Specimen ID	Medium used	Organism	Microbial Species Identity	Microbe type
1-3	YPD	Plant	*Pseudomonas parafulv*a	Bacteria
2-1	YPD	Plant	*Kosakonia cowanii*	Bacteria
2-2	YPD	Plant	*Pseudomonas fulva*	Bacteria
2-3	YPD	Plant	*Rosenbergiella nectarea*	Bacteria
3-1	YPD	Plant	*Rosenbergiella nectarea*	Bacteria
3-2	YPD	Plant	*Pseudomonas mosselii*	Bacteria
4-1	YPD	Plant	*Kosakonia cowanii*	Bacteria
4-2	YPD	Plant	*Pantoea anthophila*	Bacteria
4-3	YPD	Plant	*Rosenbergiella epipactidis*	Bacteria
5-1	YPD	Plant	*Sphingomonas sanguinis*	Bacteria
5-2	YPD	Plant	*Microbacterium testaceum*	Bacteria
5-3	YPD	Plant	*Microbacterium trichothecenolyticum*	Bacteria
5-5	YPD	Plant	*Xanthomonas campestris*	Bacteria
6-1	YPD	Plant	*Pantoea stewartii*	Bacteria
7-2	YPD	Plant	*Rosenbergiella epipactidis*	Bacteria
7-3	YPD	Plant	*Pantoea eucrina*	Bacteria
8-1	YPD	Plant	*Rosenbergiella nectarea*	Bacteria
8-2	YPD	Plant	*Pantoea eucrina*	Bacteria
8-3	YPD	Plant	*Cronobacter sakazakii*	Bacteria

815-1YAC	**YPD/antibiotic mix**	**Wasp**	***Candida pseudohaemulonii***	**Yeast**
815-2	YPD	Wasp	*Staphylococcus caprae*	Bacteria
815-3	YPD	Wasp	*Staphylococcus warneri*	Bacteria
816-2	YPD	Wasp	*Pantoea agglomerans*	Bacteria
816-3	YPD	Wasp	*Pantoea ananatis*	Bacteria
817-1	YPD	Wasp	*Asaia bogorensis*	Bacteria
817-3	YPD	Wasp	*Aneurinibacillus migulanus*	Bacteria
818-1YAC	YPD/antibiotic mix	Wasp	*Saccharomyces paradoxus*	Yeast
7-1	YPD	Plant	*Metschnikowia koreensis*	Yeast
8-5	YPD	Plant	*Cronobacter dublinensis*	Bacteria

815-1YAC	**YPD/antibiotic mix**	**Wasp**	***Candida pseudohaemulonii***	**Yeast**
815-2	YPD	Wasp	*Staphylococcus caprae*	Bacteria
815-3	YPD	Wasp	*Staphylococcus warneri*	Bacteria
816-2	YPD	Wasp	*Pantoea agglomerans*	Bacteria
816-3	YPD	Wasp	*Pantoea ananatis*	Bacteria
817-1	YPD	Wasp	*Asaia bogorensis*	Bacteria
817-3	YPD	Wasp	*Aneurinibacillus migulanus*	Bacteria
818-1YAC	YPD/antibiotic mix	Wasp	*Saccharomyces paradoxus*	Yeast
7-1	YPD	Plant	*Metschnikowia koreensis*	Yeast
8-5	YPD	Plant	*Cronobacter dublinensis*	Bacteria

Plants	**Found in Wasps**
Pseudomonas parafulva	*Fructobacillus tropaeoli*
Pseudomonas fulva	*Stenotrophomonas chelatiphaga*
Rosenbergiella nectarea	*Bacillus pumilus*
Pseudomonas mosselii	*Bacillus aryabhattai*
Pantoea anthophila	*Swingsia samuiensis*
Sphingomonas sanguinis	*Candida quercitrusa*
Microbacterium testaceum	*Staphylococcus caprae*
Microbacterium trichothecenolyticum	*Staphylococcus warneri*
Xanthomonas campestris	*Pantoea agglomerans*
Pantoea stewartii	*Pantoea ananatis*
Pantoea eucrina	*Aneurinibacillus migulanus*
Cronobacter sakazakii	*Saccharomyces paradoxus*
Erwinia persicina	*Asaia bogorensis*
Neokomagataea thailandica	*Kosakonia cowanii*
Fusarium chlamydosporum	*Rosenbergiella epipactidis*
Metschnikowia koreensis	
Cronobacter dublinensis	
Asaia bogorensis	
Kosakonia cowanii	
Rosenbergiella epipactidis	

Ambiguity in Biotechnology Patent Policy: Lessons from Association for Molecular Pathology v. Myriad Genetics, Inc.

The University of Texas at Austin

Undergraduate Research Journal

Volume 15
Number 1
Spring 2016

Chari Noddings

I. Introduction

In 2010, the Southern District Court of New York was faced with a clear question: Are human genes patentable? While the question may seem straightforward, the answer is complicated and controversial. In the United States - before an answer to this question was constructed - researchers, biotechnology firms, clinicians, the United States Patent and Trademark Office (USPTO), the Obama Administration, and three courts all weighed in on the issue of genetic patents. Unsurprisingly, the final Supreme Court ruling was far from perfect and has left the biotechnology industry moving forward toward an uncertain future for

innovation. The ruling invalidated genomic DNA patents, but upheld complementary DNA (cDNA) patents. While the implications of this recent ruling are still unknown, a study of the events leading to the Supreme Court decision sheds light on the need to reform the current patent system. Specifically, the system needs to be re-evaluated in order to ensure that patents will continue to "promote the Progress of Science and the useful Arts"[1] in the modern age. This evaluation will be made by studying the litigation against Myriad Genetics (Myriad), a private biotechnology firm that has been at the center of the genetic patent controversy since its founding in 1991.[2]

An analysis of the saga of Myriad Genetics reveals three main failures in the current patent system in relation to the biotech field. First, the patent system fails to provide distinct means for researchers to freely use patented information to perform non-commercial research. This gray area in patent law caused unnecessary confusion over patent use among researchers and patent holders, which in the case of Myriad Genetics, generated hostilities between Myriad and public researchers, impeding the progress of research through wasted resources spent on patent protection measures. Second, the patent system fails to have clearly defined subject-matter eligibility guidelines for patents, creating uncertainty over what can and should be patentable. In litigation against Myriad, the courts failed to adequately and coherently address subject-matter eligibility for patents due to their inability to address patentability in light of scientific, political, social, economic, and ethical concerns. This considerably limits the effectiveness of any court exceptions to patent law. The third failure is the current absence of policy changes to patent law. Policy makers have not yet addressed the current concerns regarding patent policy for the biotech field. The future of this sector depends on the advent of swift yet carefully crafted patent policy that takes into account the multitude of potential implications of policy reform. At present, the lack of policy reform suggests the need for novel solutions to create and implement new patent law proposals.

II. The Controversial Beginning

The case of *Association for Molecular Pathology v. Myriad Genetics, Inc.* (569 U.S. 12-398 (2013)) originates from litigation surrounding Myriad Genetics' patent claims on isolated gene sequences of the *Brca1* and *Brca2* genes as well as diagnostic methods to test for the presence of mutations within these genes and methods to test for possible therapeutics against the negative effects of mutated forms of these genes. Myriad Genetics adopted a fairly aggressive strategy to protect its patent rights in an effort to be the sole provider of diagnostic tests for the *Brca1* and *Brca2* genes. This approach caused hostility toward Myriad from the scientific community who believed Myriad was inhibiting innovation in the field. The differing opinions over the scope of patent rights reveal the current failure of patent law to strike an appropriate balance between private and public research interests.

The controversy began with research in the early 1990s that sought to find a genetic basis for high incidences of certain cancers. Researchers eventually discovered that two

genes, known as breast cancer genes *Brca1* and *Brca2*, produce tumor suppressor proteins, which help reduce the risk of associated cancers.[3] Further research presented evidence that mutations in either of these genes can greatly increase the risk of certain types of cancer. For example, research has shown that a mutation in the *Brca1* gene increases the risk of breast cancer in women in the general population from 12% to 57% and the risk of ovarian cancer increases from 1.4% to 40%.[4, 5] The drastic increase in cancer rates associated with mutations in either of the *Brca* genes made diagnostic testing for the presence of mutations in these genes a high priority for health care providers and the public with the hope that early detection could lead to life-saving preventative measures.[6]

In 1994, Myriad published a paper in *Science* detailing the results of the *Brca1* sequence mutations which had been linked to increases in the incidence of breast and ovarian cancers.[7] In 1997, Myriad was also granted patents for *Brca1* gene mutation sequences,[8] diagnostic testing for the *Brca1* mutations,[9] and finally a patent on the entire *Brca1* consensus sequence,[10] all uses of the gene, and methods to develop therapeutics.[11] Myriad then moved quickly to elucidate the sequence and relevant mutations of the *Brca2* gene. Myriad was granted a patent on the *Brca2* consensus sequence, relevant mutations of the genes, and diagnostic methods in 1998[12] and a patent for *Brca2* mutations in 2000.[13] The nine patents held by Myriad effectively gave the company complete control over diagnostic testing for the *Brca1* and *Brca2* genes because their patents severely limited the ability of other commercial firms to work with isolated versions of the genes and limited the use of methods to detect the presence of mutations in a patient's DNA sample. Among Myriad's gene patents were exclusive rights to both the DNA sequence and the complementary DNA (cDNA) sequence of the genes. cDNA differs from isolated, purified DNA because it lacks introns[14], or intervening sequences, that are removed from the genetic information by the cell's machinery before it is used to synthesize a protein. While neither the DNA nor cDNA sequence of the genes are withheld from public knowledge, as both are accessible to the public via GenBank,[15] commercial use of the isolated DNA or cDNA was strictly limited, thus giving Myriad significant control over the commercialization of diagnostic services and also the potential for future therapeutics.

Myriad began marketing three diagnostic tests in 1996 and collaborated with insurance companies (such as Kaiser Permanente[16]), laboratories, physicians, and clinicians to give patients access to their diagnostic services. The comprehensive BRACAnalysis provided full sequence testing of the *Brca* 1 and 2 genes at a cost of $2,400 and Myriad offered their diagnostic testing services through physicians, not direct to the consumer. [17] In 1998, after being granted patent rights to the *Brca* gene sequences, Myriad sent cease-and-desist letters to several laboratories that were offering diagnostic testing for the genes, including a private clinic, Genetics and In Vitro Fertilization Institute (GIVF); the University of Pennsylvania's Genetic Diagnostic Laboratory (GDL); and Yale DNA Diagnostic Laboratory. GIVF and Yale complied with the letter, but GDL refused, claiming they had a research exemption

because they were providing diagnostic tests for patients for the purposes of research. GDL also ensured the cease-and-desist letter received public attention and claimed that Myriad was attempting to impede clinical research.[18] Myriad claimed that GDL did not fall under the Hatch-Waxman research exemption because GDL was providing patients with testing results and receiving payments for the results from health care professionals associated with the National Cancer Institute's (NCI) Cancer Genetics Network.

Myriad was eventually able to negotiate with GDL, likely due to GDL's unwillingness to engage in a prolonged legal battle with Myriad in light of the extensive patents and legal knowledge that Myriad possessed. GDL eventually scaled down its diagnostic testing and opted to only provide diagnostic testing for its own research purposes without offering testing results to patients involved in the NCI studies.[19] Myriad also agreed to a Memorandum of Understanding (MOU) with the NCI that allowed Myriad to obtain rights to provide diagnostic testing at or below market cost for NCI related projects[20] in order to deter future patent infringement. This conflict represents a blatant problem with current U.S. patent policy due to the lack of explicit policies relating to potential patent infringement by clinical research pursuits. Currently, the Price Competition and Patent Term Restoration Act of 1984 (the "Hatch-Waxman Act") narrowly defines research exemption in regard to the pharmaceutical industry.[21] Efforts to expand the scope of exemption outside the realm of pharmaceuticals have been recommended[22] and are discussed in more detail in Section IV of this paper.

Myriad came to be perceived by the scientific community as a hostile entity that threatened litigation against research groups as well as other commercial firms. Although Myriad did appear to take an aggressive line of patent enforcement, public perception may have been unnecessarily critical of the company. There is little evidence to show that Myriad directly pressured scientists and clinicians into halting research involving the *Brca* genes (with the exception of GDL). Furthermore, Myriad only sent two universities cease-and-desist letters and there is no evidence that Myriad sued researchers over patent infringement on the *Brca* genes.[23] One indication that Myriad did not significantly inhibit research on *Brca* 1 or 2 is the number of articles published on either gene during this time. In 2010, there were 6,785 articles published on the *Brca* genes in the National Center for Biotech Information (NCBI) public data base, PubMed.[24] Myriad claimed that by 2014 more than 8,600 research papers had been published on the *Brca* genes with roughly 18,000 total researchers involved in the research.[25] One could argue that more research may have been done had Myriad not been granted patents on the sequences of the two genes, but it provides evidence that Myriad did not restrict all related research. Despite a lack of evidence to show Myriad restricted research, these claims greatly damaged Myriad's reputation with the public. One study published in *Genetics in Medicine* found that 77.6% of articles published in the media about Myriad and its *Brca* patents had a "negative overall tenor".[26]

Myriad's reputation, rather than its actual policies regarding information sharing, may

have deterred scientists from exploring research interests relating to the *Brca* genes and thus limited potential innovation in the area. The restriction that arose was not necessarily endemic to the patent rights exercised by Myriad but seemed to be caused by a lack of information that lead to misunderstandings between Myriad and the scientific community about use of Myriad's patents. Although Myriad was not the first group to be granted a gene patent,[27] the scientific community was critical of Myriad's patents and many researchers spoke out against the company. Many researchers were concerned that the patents blocked further research into genetic mutations of the genes and thus inhibited the ability to design more effective and less expensive testing.

Myriad appears to have also entered into the market at a time when hostility toward gene patents from the scientific community was becoming widespread. In 1998, Heller and Eisenberg's article in *Science* caused concern among biotech researchers about the increasing fragmentation of intellectual property rights, especially in regard to gene patents.[28] In 2001, the first draft of the human genome was publicly released in line with the Human Genome Project's (HGP) commitment to free data access as drawn up in the "Bermuda Principles".[29] The scientific community viewed the HGP, not as an avenue for various groups to profit from, but as a cooperative effort to find new knowledge.[30] Many researchers felt that gene patents inhibited their ability to use the findings of the Human Genome Project efficiently.[31] The timing of these events along with Myriad's patent grants, monopoly on diagnostic services, and distribution of cease-and-desist letters may have made Myriad a prime target for the growing hostility within the scientific community. Despite this, Myriad seemed to operate well within their legal patent rights, although admittedly the company was uncharacteristically aggressive in their patent protection. While the research community directed its hostility toward Myriad, the real controversy stems from varying interpretations of the purpose and scope of patents, thus pointing to a failure, not by Myriad, but by the current patent law, which does not provide clear stipulations on patent-eligible subject matter or the extent of research exemption for use of patents. These incoherencies in the current law have led public and private interests to interpret patent protections differently, which has caused significant conflict and a failure to strike a balance of power between the two groups, thus creating barriers to research.

III. The Litigation Phase

The building hostilities between the research community and Myriad eventually resulted in litigation brought against Myriad by groups claiming Myriad's patents impeded research relating to the *Brca* genes and created a monopoly over diagnostic testing services. The courts were thus tasked with the goal of sorting out the incoherencies and limitations of patent law in relation to the biotech sector. Three courts, the Southern District Court of New York, the Federal Circuit Court of Appeals, and, finally, the U.S. Supreme Court, all weighed in on the matter, with each court providing largely different rulings that stem

from conflicting interpretations of the current subject-matter eligibility law. While the U.S. Supreme Court had the final word, the results of these rulings reveals confusion over subject-matter eligibility interpretation within the judiciary, pointing to a need for policy reform to bring clarity to the current law. Furthermore, the limitations of the courts, which cannot provide rulings based on scientific, political, social, economic, or ethical considerations, also suggests policy reform must come from an agency with the ability to consider these far-reaching implications.

A. The Southern District Court of New York

In 2009, a group in the United States sued Myriad over fifteen claims of composition and method in seven of Myriad's patents. The plaintiffs consisted of nonprofits, university-affiliated researchers, the American Civil Liberties Union (ACLU), and individuals claiming Myriad's patents impeded research relating to the *Brca* genes and created a monopoly over diagnostic testing services. The defendants included Myriad Genetics, the University of Utah Research Foundation (UURF), and the USPTO. Judge Sweet of the Southern District Court of New York presided over the lawsuit and granted summary judgment to the plaintiffs.

The opinion of the court first presents this question to be answered: "Are isolated human genes and the comparison of their sequences patentable?"[32] Based on 35 U.S.C. §101, novelty and utility are the two requirements necessary for patentability and US §103 further expands patentability to also require non-obviousness.[33, 34] *Diamond v. Chakrabarty* (1980) presented limits to patent eligibility: "This is not to suggest that §101 has no limits or that it embraces every discovery. The laws of nature, physical phenomena, and abstract ideas have been held not patentable... ."[35] The USPTO justified granting patents on genetic sequences as "isolated DNA", asserting that purification of DNA from outside an organism renders it a new artifact. The court, however, found that "DNA's existence in an 'isolated' form alters neither [the] fundamental quality of DNA as it exists in the body nor the information it encodes" and ruled Myriad's composition claims patent ineligible under §101. Furthermore, the Court found Myriad's method claims to be ineligible for patenting under §101, claiming "comparison[s] of DNA sequences are abstract mental processes."[36]

Many of the arguments at issue in the Myriad case are enumerated in the various amici briefs submitted to the Federal District Court for consideration. The American Medical Association argued to invalidate the patents-in-suit because the patents claim natural phenomena (which is subject-matter ineligible), the patents are unnecessary to promote innovation for genetic research, and the patents violate scientific and medical ethics.[37] The amicus brief submitted by the Breast Cancer Coalition and other non-profit groups also add that the patents-in-suit restrict future research and scientific progress.[38] In opposition, the Biotech Industry Organization (BIO) argues the patents are subject-matter eligible because they differ from naturally-occurring, non-isolated DNA and that the patents provide incentives for investments in biotech needed to promote the advancement of science.[39]

With experts in the field on both sides of the issue, it is difficult for the District Court, with a limited technical background, to make an informed and reasonable decision in regard to the case. Furthermore, as the opinion of the court mentions,[40] the court is not qualified and cannot address whether patents promote or impede scientific advancements, which was one of the main disputes set forth on both sides of the argument in relation to gene patents. The court also chose not to consider any ethical arguments in the written opinion. In Judge Sweet's decision, the court struck down all of Myriad's composition claims based on the assertion that isolated DNA (both genomic DNA and cDNA) is not man-made and is not "markedly different" from natural phenomena. The opinion, however, is unclear because as Myriad and the USPTO argue, isolated DNA sequences are purified out of the cells in which they originated and the phosphodiester bonds[41] on either side of the sequence are severed, making the isolated DNA chemically different from that found within a human. Thus, the court seemed to artificially draw a line between natural and man-made without adequate or clear reasoning.

The court also invalidated all of Myriad's method claims-in-suit. The claims comparing the patient's sample DNA with known sequences to analyze whether the patient has mutations in the relevant genes were invalidated under the "machine or transformation" test adopted by the Federal Circuit in 2008 to implement §101.[42] The Court explained that the methods described mental processes, not any physical transformations. The therapeutic method claim for comparing the growth rates of cells was also invalidated on grounds that the claim patents a basic scientific principle, the scientific method itself.[43] These interpretations, however, are also not well substantiated, as the Supreme Court has recently declared that there is no definitive test for patentable subject matter. According to the Court, the machine-or-transformation test is no more than "a useful and important clue, an investigative tool, for determining whether some claimed inventions are processes under § 101."[44]

B. The Federal Circuit Court of Appeals

Myriad appealed to the Federal Circuit on the rulings that invalidated both the composition and method claims.[45] Myriad argued that the District Court misinterpreted the Supreme Court's precedent on the patent ineligibility of products of nature. Myriad claimed that the court focused solely on the similarities between isolated and natural DNA instead of considering the differences between the two substances. Myriad also argued that the invalidation of the composition claims through a "products of nature" argument is against the USPTO *Utility Examination Guidelines* (2001) and is "unworkable" because all matter is composed of natural material.[46]

The basic problem at hand is how to interpret "naturally occurring" in a clear, concise manner. The Federal Circuit took into an account an amicus brief from the Obama Administration that did not side with the position of the USPTO, but instead argued that isolated genomic DNA was patent-ineligible, while isolated cDNA was patent-eligible. In

oral arguments, the Administration used a "magic microscope" test to decide if a substance is naturally occurring. If a theoretical microscope could reveal a molecule as it exists in the human body, then the molecule is naturally occurring and is patent-ineligible. Under this test, isolated genomic DNA is patent-ineligible, while cDNA is patent-eligible because it cannot be revealed *in vivo*, but is rather engineered in a lab.[47] Despite this argument, the circuit court ruled both the DNA and cDNA claims subject matter eligible based on the argument that isolated, purified DNA is markedly different from native DNA.

The opinion was based on the argument that isolated DNA does not include the associated histone proteins and three-dimensional structure that native DNA undertakes when organized in the cell as chromatin. Furthermore, in order to isolate DNA, two phosphodiester bonds must be broken, rendering the molecule chemically distinct from native DNA. Finally, the opinion also claims that genomic DNA segments can be synthetically created in the lab without the need to purify the molecule from a cell, thus making it "man-made."[48] The court therefore upheld the USPTO's longstanding policy on genomic patents and discounted the Obama Administration's "magic microscope" test to determine what is "naturally occurring". The opinion also notes that at that time the USPTO had issued 2,645 patents on isolated DNA and over 40,000 patents relating to the human genome.[49] Furthermore, Congress never acted to indicate that the USPTO's policy on genomic patents was incompatible with §101 (however, Congress never endorsed the USPTO's policy either). The court also noted that a decision to invalidate genomic patents should be made by Congress, not the judiciary, because of the implications concerning scientific innovation, an implication the court could not adequately nor properly address when issuing a ruling.[50]

The Federal Circuit also considered Myriad's method claims, the first for analyzing and comparing DNA sequences and the second for screening potential cancer therapeutics. The Federal Circuit upheld the District Court's decision ruling Myriad's first method claims patent-ineligible, citing *Mayo v. Prometheus*, and arguing the claims cover abstract mental processes. The court ruled the second method claims patent-eligible based on the argument that methods of screening for potential therapeutics do not cover a basic scientific principle, as the lower court had claimed, but rather cover a process of evaluation that applies to a specific type of cell transformed with specific genes and grown with a specific therapeutic.[51] Both the rulings on the composition claims and the method claims further reveal the incoherencies associated with subject-matter eligibility, as the Federal Circuit arrived at different rulings than the District Court by implementing strikingly different interpretations of the same law.

The Federal Circuit, however, appears more adept than the District Court at understanding the arguments surrounding the *Myriad* case for two reasons. First, the majority of the Federal Circuit's jurisdiction consists of three types of cases: administrative law cases, cases involving money damages against the U.S. government and intellectual property cases, the latter of which makes up 31% of the Court's cases.[52] Furthermore, nearly all of the intellectual

property cases involve patent rights. Therefore, the Circuit Court has more experience and expertise in dealing with patent cases. Second, one of the presiding judges on the *Myriad* case, Justice Lourie, who wrote the opinion of the court, is a classically trained chemist with a PhD in chemistry, which perhaps afforded him the scientific knowledge needed to more adequately assess the merits of the case.[53]

Despite these considerations, the Federal Circuit Court has been criticized for its reasoning set forth in the *Myriad* case. First, the Federal Circuit, like the District Court, asserted that DNA was the only carrier of biological information, which is not consistent with current molecular biology principles that regard many biological molecules as carriers of information.[54] Second, in Justice Bryson's dissenting opinion he stated, "Biochemists extract the target genes along lines defined by nature so as to preserve the structure and function that the gene possessed... ."[55] This opinion represents a fundamental misunderstanding of the concept of genes and thus points to the lack of technical knowledge necessary to address this case, even at the Federal Circuit level.[56] In fact, it has been argued that there "is no sense acting as if the category of genes, or for that matter any given gene, is in a fundamental sense ever *anything* but the product of human invention. The concept of a gene is entirely a human construct... ."[57] Despite this argument, none of the judiciary opinions of the District Court, circuit court, or later the Supreme Court, attempted to address or even mention what constitutes a gene at the molecular level.

The intricacies of molecular biology and the ever expanding discoveries associated with this area of research continue to redefine how genetic material is viewed. While every court involved with the *Myriad* case began the court opinion with an overview of basic molecular biology concepts, the limited understanding of the court severely diminished the coherency and validity of court interpretations. This is not to be seen as a suggestion for the court to necessarily delve further into the technical issues of each case, but, as Judge Lourie's opinion suggests, to allow Congress or a more informed governing body formulate more decisive patent policy.

C. The Supreme Court

After the Federal Circuit's decision on remand from the Supreme Court, the Supreme Court granted certiorari to review the *Myriad* decision in regard to the composition claims in an effort to establish whether gene patents are subject-matter eligible. Justice Thomas delivered the unanimous opinion of the Court, which affirmed in part and reversed in part the Federal Circuit judgment. The Court held that, generally speaking, patents on cDNA are patent eligible, while isolated genomic DNA is patent-ineligible. The Court argued that isolated DNA was not "markedly different" from anything found in nature, citing the judicially created exception to subject-matter eligibility under §101 previously articulated by the Court in *Funk Bros.*[58] and *Chakrabarty*.[59] The Court rejected the Federal Circuit's argument that the chemical changes to the DNA involved in purification and isolation constituted a

distinctive change. The Court pointed out that Myriad's patent claims did not focus attention on these chemical changes, but rather the information contained within the isolated DNA sequences, which were unaltered from the native DNA sequences. The opinion argued that cDNA was patent eligible because it is not a naturally occurring substance, but one that is synthesized within the lab.[60]

Thus, the Court followed the opinion set forth in the Obama Administration's amicus brief, which took a middle line in regard to patent eligibility. While the amicus was dedicated mostly to interpretations of the law, the last section entitled "field-specific concern about encouraging innovation..." looked at the implications of the proposed decision to restrict patents on genomic DNA but allow patents on cDNA. The authors argued that restricting patents on isolated genomic DNA would not deter progress within the biotech sector because other related innovations may be patented, such as "novel methods of identifying, isolating, and using DNA molecule... [and] any new and useful alteration of those molecules through human intervention." Furthermore, the brief argues that "an overbroad conception of patent-eligibility under §101 can impose significant social costs by requiring the public to pay to study and exploit..." and the "interest of the public has consistently been given precedence."[61] The Supreme Court seemed to agree with the opinion set forth by the U.S. amicus brief and chose not to defer to the USPTO whose policies were challenged by the Department of Justice and were not endorsed by Congress.

The Supreme Court's decision has generated significant controversy both for the reasoning behind the Court's judgment and for the implications that the judgment carries. First, the reasoning has been criticized as touting "genetic exceptionalism," which is the view that information contained in genes is qualitatively different from other forms of information and therefore is subject to different regulation. This view arguably led the court to rule DNA patent-ineligible but cDNA patent-eligible despite the fact that the sequence of cDNA contains analogous information to DNA, which is directed by nature.[62] Furthermore, this view perhaps incorrectly separates genetic material from other biological substances that have historically been patent-eligible. For example, *Parke-Davis & Co. v H.K. Mulford Co.* effectively held that a purified form of the molecule adrenaline was patent eligible because once isolated it constituted "a new thing commercially and therapeutically."[63]

The Court's decision in *Myriad* now sets genomic DNA apart from other biological material in terms of patent policy. The Court's ruling has also been criticized for its decision not to review the method claims in question at the petition stage before the Federal Circuit. Thus, the Federal Circuit's rulings on those claims are left standing without any interpretation or comment by the Supreme Court, leaving unresolved questions about the "law of nature" exception to subject matter eligibility. The Federal Circuit argued that Myriad's "comparing" and "analyzing" claims cover an "abstract, mental step" and later that it constituted a "law of nature" and is thus, in light of *Mayo*, patent-ineligible. The argument does not clearly state whether the claims are ineligible because they are abstract ideas or

laws of nature. Furthermore, the idea that collecting and analyzing data is a "law of nature" is controversial and the Supreme Court's omission of interpretation of this argument suggests that perhaps the grounds for this judge-made exception to §101 are not well substantiated.[64]

IV. Repairing the System

More time is needed to determine how the case of Myriad Genetics and the recent Supreme Court rulings will affect innovation in the biotech sector. The saga of Myriad points to numerous failures in the current system of research and development in this quickly growing field and the courts seem ill-suited to address and mitigate these problems. In fact, the courts have only further muddled the laws concerning subject-matter eligibility and have shrouded the future of biotech research and innovation in uncertainty. While many factors contribute to this murky landscape, including the initial failure of Myriad to work and communicate with the scientific community and the failure of the courts to dispel the confusion surrounding patent law, a third failure continues to perpetuate this uncertainty. This is the failure of policy makers to construct new legislation that could strike a more appropriate balance between the goals of public research groups and private research firms.

The scarcity of new legislation in this area is not due to a lack of concern over the issues uncovered by Myriad. As a direct response to Myriad's patents on the *Brca1*/2 genes, a bill, the Genomic Research and Diagnostic Accessibility Act of 2002,[65] was introduced to permit physicians to use genetic testing without obtaining license agreements from patent holders and to require research groups supported by federal funds to publicly disclose gene sequences after filing for a patent. Another bill, the Genomic Science and Technology Innovation Act of 2002,[66] called for studies to be conducted to assess the extent to which researchers and clinicians were inhibited in collecting sufficient information in order to conduct research. Neither of these bills passed.[67] Finally, a third bill, the Patent Reform Act of 2005,[68] was introduced to overhaul the patent system, which led to the Leahy-Smith America Invents Act (AIA) of 2011.[69] However, the AIA did not address patent policy issues specifically related to the biotechnology sector.[70] Thus, Congress failed to pass crucial legislation that could help ensure innovation in biotech is still adequately encouraged within the patent system. In contrast, the European Union (EU) enacted legislation concerning the specifics of patent eligible subject matter in biotechnology. The EU Directive 98/44/EC states that isolated biological material is patent-eligible, while diagnostic methods are ineligible.[71] Unlike the EU, Congress has failed to provide clear subject matter eligibility guidelines for biotechnology innovations. Due to the quickly evolving nature of the biotech field, policy makers need to create tailored and clear legislation in a timely manner to address issues as they arise. The actions of Myriad and the decisions set forth by the courts reveal a maladjusted patent system which prompts the need for novel solutions.

While there is reason to believe that the patent system may work itself out as the biotech industry settles into more predictable business norms, this marks a key opportunity to craft

policy to guide this relatively new industry towards finding a balance between the interests of public researchers and private firms. During the litigation, the courts, the Justice Department, and the USPTO all had differing opinions on interpretations of subject matter eligibility, while the legislature remained mostly quiet on the issue. The incoherencies between these various groups strongly point to the need for a directed effort to establish coherent policy regarding patent eligibility and the scope of patent protections. First, policy regarding patent eligibility could be set and regulated by a new governing agency or these tasks could be absorbed by an existing agency. Second, the scope of patent protection can be limited to promote basic research[72] by clearly defining and enforcing research exemption. These suggestions should not be regarded as a comprehensive solution to overhaul patent policy, but rather as insight into two potential avenues for improving the current patent system.

A. Governing Agency

The current uncertainty with respect to patent subject-matter eligibility impedes further progress in the biotech sector because researchers, corporations, and investors cannot clearly assess whether patent rights will be granted. Without the assurance of intellectual property rights in the future, actors will not invest upfront costs into research and development.[73] In order to effectively reform patent policy, an agency with various perspectives is needed to assess the impact of possible patent policies. This agency would need to have the appropriate technical knowledge to understand the types of patents at issue, an ability to analyze economic trade-offs and incentives to innovate within the field, an ability to draft coherent policy, the means to enforce this policy, and the foresight to determine social and ethical implications of any policy reforms. In principle, Congress may seem well-positioned to take over these responsibilities; however, Congress' current inability to pass legislation in a timely manner as well as its susceptibility to lobbying efforts makes it unable to adequately carry out the proposed responsibilities. Instead, a governing agency must either be created or an existing agency must absorb the task of crafting and implementing new patent policy.

As John M. Golden has proposed, the USPTO may be best suited to address questions concerning patentable subject matter. The USPTO is most familiar with the patent application and review process and already must draft guidelines to interpret current patent law when reviewing patent applications.[74] The USPTO appears to have adequate technical knowledge of the science involved to better assess the innovative implications of changes to patent law while also having the ability to draft and enforce patent policy. However, the other criteria (i.e. analysis of economic trade-offs and incentives to innovate as well as considerations of social and ethical implications) may present more unfamiliar aspects for the USPTO to adequately address at this time.

B. Research Exemption

Much of the concern regarding gene patents relates to the possibility that these patents infringe on further research and innovation. While the Supreme Court has struck down isolated genomic DNA patents, cDNA patents are still subject-matter eligible. While Myriad claims they did not inhibit basic research concerning the *Brca1*/2 genes, the dissemination of cease-and-desist letters to at least three academic research groups may have deterred scientists from pursuing research in this area. While a private company like Myriad has legal rights to their genetic patents, the progress of scientific research should not be unnecessarily inhibited by private interests. Had Myriad waived their rights to collect licensing fees from academic researchers, like other biotech companies have opted to do,[75] Myriad may not have been viewed as such a hostile entity toward the research community. Furthermore, clear and open communication between Myriad and public research groups could have led to collaborative efforts to continue research on these genes, instead of resorting to cease-and-desist threats, the withholding of scientific information concerning the genes, and costly litigation, all of which impede the progress of the science.

It is not patents that cause these impediments to research, but rather the current inability for the patent system to accommodate a larger scope for research exemption. While the understanding of research exemption in many cases is generally and informally followed by patent holders, a formal agreement would further protect the interests of public researchers and allow them to freely pursue research interests without the fear of litigation, as is suggested by Donna M. Gitter. Furthermore, this would promote the general principle in science that discoveries and inventions should be tested and scrutinized by the scientific community in the form of peer review, which can only be done with the availability of free access to information. Strengthening the experimental exemption in a legal framework would incentivize public, non-commercial research while also enabling private firms to secure patents for profitability.[76] This strategy does have limitations as it is sometimes difficult to define public, non-commercial research, especially in the biotech sector where basic and applied research is often blurred. While not a perfect solution, the strengthening of experimental exemption could hold an important place in patent law, similar to the "fair use" defense that is applicable to copyright law.[77] A clear, well-crafted research exemption is needed to specifically allow free use of patented biological materials for non-commercial research, including diagnostics. This research exemption will help strike a balance between research from public, noncommercial groups and private, commercial patent-holders, diminishing hostilities and thus promoting continued innovation.

V. Conclusion

An analysis of the Myriad Genetics case, which starts from the initial research and development phase, then evaluates the litigation phase, and finally considers the absence of policy reforms, reveals three main failures with the current patent system in the biotech

field. First, the failure of patent law to adequately and clearly provide reasonable access to and use of patented information by public research groups. This failure to articulate a clear legal understanding of subject-matter eligibility and research exemption generated hostilities between Myriad and public researchers, impeding the progress of research through wasted resources spent on patent protection and subsequent litigation. Second, there was a significant failure by the courts to adequately and coherently address subject-matter eligibility in the ruling on *Association for Molecular Pathology (AMP) v. Myriad Genetics*. This failure has led to an incoherent and inadequate Supreme Court decision that has resulted in uncertainties for future biotech innovation. Furthermore, the Court lacked the ability to consider the scientific, political, social, economic, and ethical implications of the subsequent changes in patent policy, thus potentially resulting in negative effects for continued innovation in the biotech industry. Finally, policy makers have failed to address the current concerns regarding the inadequate patent policy in relation to the quickly changing field of biotech.

The future of the increasingly important and dynamic biotech field depends on the advent of swift yet carefully crafted patent policy that takes into account the multitude of potential implications of policy reform. Novel solutions, such as policy creation and implementation by a governing agency like the USPTO, as well as clarification and expansion of research exemption, are needed to dispel the current ambiguities and promote continued innovation. For individuals working in the biotech sector, the current patent protection system is flawed and has failed to provide consistent and coherent policy, leaving the future of innovation in this industry in an uncertain predicament. However, with significant attention and effort placed on resolving the current issues at this crucial junction in patent policy, the field of biotechnology can be encouraged to reach its full potential for innovation.

Endnotes

1 U.S. Const. art. I § 8, cl.8.

2 *See* Inside Myriad: History (visited Dec. 4, 2014), http://www.myriad.com/about-myriad/inside-myriad/history/

3 *See* Y. Wang et al., *BASC, a super complex of Brca1-associated proteins involved in the recognition and repair of aberrant DNA structures*, 14 Genes Dev., 927, 927-39 (2000); *See* K.J. Patel et al., *Involvement of Brca2 in DNA repair*, 1 Molec. Cell 347, 347-57 (1998) ("These findings define a function of *Brca2* in DNA repair, whose loss precipitates replicative failure, mutagen sensitivity, and genetic instability.").

4 *See* S. Chen S, G. Parmigiani, *Meta-analysis of Brca1 and Brca2 penetrance*, 25 Journal of Clinical Oncology 1329, 1329–1333 (2007) ("Meta-analytic mean cumulative cancer risks for mutation carriers at age 70 years were as follows: breast cancer risk of 57% (95% CI, 47% to 66%) for *Brca1* and 49% (95% CI, 40% to 57%) for *Brca2* mutation carriers; and ovarian cancer risk of 40% (95% CI, 35% to 46%) for *Brca1* and 18% (95% CI, 13% to 23%) for *Brca2* mutation carriers.").

5 *See* N. Howlader et al., *SEER Cancer Statistics Review, 1975-2010,* National Cancer Institute (visited Dec. 4, 2014), http://seer.cancer.gov/csr/1975_2010/ (See *Table* 4.1 of Section on "Cancer of the Female Breast (Invasive)).

6 See *Brca1 and Brca2: Cancer Risk and Genetic Testing,* National Cancer Institute (visited Dec. 4, 2014), <http://www.cancer.gov/cancertopics/factsheet/Risk/*Brca*>

7 *See* Y. Miki et al., *A strong candidate for the breast and ovarian cancer susceptibility gene Brca1,*266 Science 66, 66-71 (1994).

8 U.S. Patent No. 5,693,473 (issued Dec. 2, 1997).

9 U.S. Patent No. 5,709,999 (issued Jan. 20, 1998); U.S. Patent No. 5,710,001 (issued Jan. 20, 1998); U.S. Patent No. 5,747,282 (issued May 5, 1998); U.S. Patent No. 5,753,441 (issued May 19, 1998); and U.S. Patent No. 6,162,897 (issued Dec. 19, 2000).

10 A representative nucleotide sequence in which each nucleotide is the one which occurs most frequently at that site in the different sequences which occur in nature. See *Medical Subject Headings (MeSH) Browser,* U.S. National Library of Medicine (visited Dec.4, 2014), http://www.nlm.nih.gov/mesh/meshhome.html.

11 U.S. 5,747,282 (issued May 5, 1998).

12 U.S. Patent No. 5,837,492 (issued Nov. 17, 1998).

13 U.S. Patent No. 6,124,104 (issued Sep. 26, 2000).

14 Introns are sequences of DNA that are removed from the primary gene transcript. *See* U.S. National Library of Medicine, *supra* note 10.

15 *See* D.A. Benson et al., *GenBank,* 33 Nucleic Acids Research D34, D34-D38, ("GenBank® is a comprehensive database that contains publicly available DNA sequences for more than 165000 named organisms...").

16 *See* E. Gold and J. Carbone, *Myriad Genetics: In the Eye of the Policy Storm*, 12 Genet Med. S45 (2010) at S45 ("In 2000, Myriad signed an agreement to provide its BRACAnalysis breast and ovarian cancer susceptibility test to Kaiser Permanente patients.").

17 *See id.* At S44-S45 ("[Myriad] began marketing three principal diagnostic tests...").

18 *See id.* At S46 ("Myriad sent letters to the Genetics and IVF Institute (GIVF) and the University of Pennsylvania's Genetic Diagnostic Laboratory (GDL) in 1998. GIVF acquiesced to Myriad's demand to stop testing but GDL initially refused to do so, claiming that a "research exemption" covered its activities.").

19 *See* S. Parthasarathy, *Architectures of genetic medicine: comparing genetic testing for breast cancer in the USA and the UK*, 35 Soc Stud Sci 5, 24 (2005) ("After a series of negotiations, Myriad forced GDL to shut down its *Brca* testing laboratory...This resolution, which arose through a combination of Myriad's patent rights and legal resources and GDL's reluctance to engage in a prolonged legal battle with the company, allowed Myriad to control the definition of research and impose its definition of the boundary between research and commercial service.").

20 *See* T. Reynolds, *NCI-Myriad agreement offers Brca testing at reduced cost*, 92 J. Natl Cancer Inst.596, 596 (2000). ("Greg Critchfield, M.D., president of Myriad's diagnostic laboratory subsidiary, said it is in the company's interest as well as the public's to promote and facilitate research on the genes. This fact led Myriad to offer researchers within or funded by the National Institutes of Health full *Brca* gene sequencing for $1,200 per person, just less than half the commercial cost, netting Myriad no profit.").

21 *See* Madey v. Duke University, 307 F.3d 1351 (Fed. Cir. 2002) where the Federal Circuit upheld the narrow scope of the experimental use research exemption in stating the exemption "does not immunize any conduct that is in keeping with the alleged infringer's legitimate business, regardless of commercial implications," which extends to university research. Also *See* A. Baldwin and R. Deegan-Cook, *Constructing narratives of heroism and villainy: case study of Myriad's BRACAnalysis and Genentech's Herceptin*, 5 Genome Medicine 8, 22 (2013) ("The US research exemption, however, does not include 'research using' an invention...").

22 *See* K.J. Strandburg, *The Research Exemption to Patent Infringement: The Delicate Balance between Current and Future Technical Progress*, Intellectual Property and Informational Wealth 8, 52 (2006) ("...the traditional research exemption has been reduced to a mere *de minimis* exception that bears little relation to the implications of a particular experimental use for follow-on innovation...").

23 *Association for Molecular Pathology v. USPTO*, 702 F.Supp.2d 181, 240-42 (S.D.N.Y. 2010) ("On June 10, 1999, Myriad's general counsel, Christopher Wright, sent a letter to the University of Pennsylvania seeking written assurances that Dr. Kazazian and the University of Pennsylvania had ceased *Brca1/2* clinical testing... In December 2000, the director of the Yale DNA Diagnostics Lab received a cease and desist letter concerning *Brca1/2* genetic testing being conducted by the lab.").

24 *See* Gold, *supra* note 16 at S50 ("(A search of PubMed for articles containing *Brca1* or *Brca2* reveals 6785 articles. Note that this number includes articles on ethics and clinical care.)").

25 *See* M. Capone, *Setting the Record Straight: Comments on Recent Media Reports Regarding Brca1/2 Patents,* Myriad 1, 1 (viewed on Dec. 4, 2014), https://www.myriad.com/lib/speakerportal/Setting%20the%20Record%20Straight.pdf ("Since the discoveries of *Brca1* and *Brca2* more than 18,000 scientists have researched the *Brca* genes...").

26 *See* T. Caulfield et al., *Myriad and the mass media,*9 Genet Med. 850, 850-55 (2007) ("Myriad's patents were largely portrayed as a negative story... and as an example of the problems associated with gene patents.").

27 U.S. Patent No. 4,447,538 (issued on May 8, 1984).

28 *See* M.A. Heller and R.S. Eisenberg, *Can Patents Deter Innovation? The Anticommons in Biomedical Research*, 280 Science 5364 (1998) ("...a resource is prone to underuse in a 'tragedy of the anticommons' when multiple owners each have a right to exclude others from a scarce resource and no one has an effective privilege of use.").

29 See *Statement on genome data release,* Wellcome Trust (viewed on Dec. 4, 2014) http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTD002751.htm ("It was agreed that all human genomic sequence information, generated by centres funded for large-scale human sequencing, should be freely available and in the public domain in order to encourage research and development...").

30 *See* D. Butler & D. Gershon, *Breast cancer discovery sparks new debate on patenting human genes,* 371 Nature 271, 272 (1994) ("...many in the research community remain uncomfortable about patenting genes. Much of this discomfort stems for their belief that the human genome project should be a cooperative search for new knowledge, rather than a self-interested search for profits.").

31 J.F. Merz et al., *Diagnostic testing fails the test*, 415 Nature 577, 578 (2002) ("Setting aside the debate about whether it is wise to allow patenting of human gene sequences at all, many are concerned about the ramifications of gene patents for biomedical research and clinical medicine. Unfortunately, few empirical data exist about the effects of patents on the translation of genomic discoveries into medical advances...").

32 *Association for Molecular Pathology v. USPTO*, 702 F.Supp.2d 181, 182 (S.D.N.Y. 2010).

33 35 U.S.C. §101 ("Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor...").

34 35 U.S.C. §§102 and 103.

35 *Diamond v. Chakrabarty*, 444 U.S. 303 (1980).

36 *Association for Molecular Pathology v. USPTO*, 702 F.Supp.2d 181, 185 (S.D.N.Y. 2010).

37 Brief for the AMA as Amicus Curiae, p.17, *Association for Molecular Pathology v. USPTO*, 702 F.Supp.2d 181, 197 (S.D.N.Y. 2010) ("[The AMA] amici contend that the patents-in-suit are directed to unpatentable natural phenomena in violation of Article I, Section 8, Clause 8 of the Constitution...").

38 Brief for the BCC as Amicus Curiae, p.18, *Association for Molecular Pathology v. USPTO*, 702 F.Supp.2d 181, 198 (S.D.N.Y. 2010) ("[The BCC] amici contend that Myriad's patents represent patents on natural phenomena and laws of nature, thereby restricting future research and scientific progress.").

39 Brief for the BIO as Amicus Curiae, p.20, *Association for Molecular Pathology v. USPTO*, 702 F.Supp.2d 181, 200 (S.D.N.Y. 2010) ("BIO contends that patents direct to isolated DNA fall within the categories of patent-eligible subject matter because they differ 'in kind' from naturally-occurring DNA...patents such as the ones in dispute here provide incentives for investment in biotechnology that promotes the advancement of science.").

40 *Association for Molecular Pathology v. USPTO*, 702 F.Supp.2d 181, 259 (S.D.N.Y. 2010) ("...there exists a sharp dispute concerning the impact of patents directed to isolated DNA on genetic research and consequently the health of society...the resolution of these disputes of fact and policy are not possible within the context of these motions.").

41 Phosphodiester bonds are the links between two sugar molecules using a phosphate molecule. These sugar linkages join nucleotides together to create a DNA sequence. *See* Robert F. Weaver, Molecular Biology, 5th Edition, 17 (2011).

42 *Association for Molecular Pathology v. USPTO*, 702 F.Supp.2d 181, 316 (S.D.N.Y. 2010). ("In *Bilski* [545 F.3d at 953], the Federal Circuit set forth 'the definitive test'... Under this 'machine or transformation' test, '[a] claimed process is surely patent-eligible under 101 if: (1) it is tired to a particular machine or apparatus, or (2) it transforms a particular article into a different state or thing.'...")

43 *Association for Molecular Pathology v. USPTO*, 702 F.Supp.2d 181, 329 (S.D.N.Y. 2010) ("[The] claimed "process" is, in fact, the scientific method itself, and claim 20 seeks to patent a basic scientific principle...").

44 *See* J. Golden, *Patentable Subject Matter and Institutional Choice,* 89 Texas L. Rev 1041, 1083 (2011); Bilski v. Kappos, 130 S. Ct. 3218, 3227 (2010).

45 It should be noted that after the original Federal Circuit decision in *Myriad,* a petition for certiorari was filed with the Supreme Court and granted. The Supreme Court vacated the Federal Circuit's decision and remanded the case for further analysis in light of the recent *Mayo v. Prometheus* decision ruled by the Supreme Court. The Federal Circuit's judgment was not altered in its re-consideration of the case, with Judge Lourie writing the opinion, Judge Moore concurring in part, and Judge Bryson concurring in part and dissenting in part. The subsequent analysis of the Federal Circuit judgment is regarding the second opinion of the court. *See* CRG Staff, *Supreme Court Orders New Review of Myriad Gene Patents*, Council for Responsible Genetics (viewed on Dec. 4, 2014), http://www.councilforresponsiblegenetics.org/genewatch/GeneWatchPage.aspx?pageId=425

46 *Association for Molecular Pathology v. USPTO,* 689 F.3d 1303, 1341-42 (Cic. 2012) ("Moreover, Myriad asserts that an ultimately-derived-from 'products of nature' exception not only would be unworkable, as every composition of matter is, at some level, composed of natural materials, but also would be contrary to this court's precedents, the PTO's 2001 *Utility Examination Guidelines*, and Congress's role in enacting patent laws.").

47 Brief for the Department of Justice as Amicus Curiae, p.37, *Association for Molecular Pathology v. USPTO,* 689 F.3d 1303 (Cic. 2012) ("For the foregoing reasons, the Court should reverse the District Court's invalidation of the composition claims that are limited to cDNAs and similar man-made constructs, but affirm the District Court's conclusion that the claims encompassing isolated human genomic DNA are invalid.").

48 *See Association for Molecular Pathology v. Myriad Genetics,* 569 U.S. 12-398, 1340 (2013) ("The isolated DNA molecules before us are not found in nature. They are obtained in the laboratory and are man-made, the product of human ingenuity. While they are prepared from products of nature, so is every other composition of matter.").

49 *See* E. Rogers, *Can you Patent Genes? Yes and No,*93.1 Journal of the Patent and Trademark Office Society 19, 19 (2011) ("An estimate in 2005 suggested that the

United States had granted 40,000 DNA–related patents, covering about 20 percent of the genes in the human genome...").

50 *See* Association for Molecular Pathology *supra* note 49 at 1340-41 ("We would further note, in the context of discussing what this case is not about, that patents on life-saving material and processes, involving large amounts of risky investment, would seem to be precisely the types of subject matter that should be subject to the incentives of exclusive rights. But disapproving of patents on medical methods and novel biological molecules are policy questions best left to Congress, and other general questions relating to patentability and use of patents are issues not before us.").

51 *See* Association for Molecular Pathology *supra* note 49 at 1363 ("Furthermore, the claim does not cover all cells, all compounds, or all methods of determining the therapeutic effect of a compound. Rather, it is tied to specific host cells transformed with specific genes and grown in the presence or absence of a specific type of therapeutic.").

52 See *Court Jurisdiction,* UNITED STATES COURT OF APPEALS OF THE FEDERAL CIRCUIT (visited on Dec. 4, 2014), <http://www.cafc.uscourts.gov/the-court/court-jurisdiction.html> ("The court's jurisdiction consists of administrative law cases (55%), intellectual property cases (31%)...").

53 See *Alan D. Lourie, Circuit Judge,* UNITED STATES COURT OF APPEALS OF THE FEDERAL CIRCUIT (visited on Dec. 4, 2014), < http://www.cafc.uscourts.gov/judges/alan-d-lourie-circuit-judge.html> ("Judge Lourie received his Bachelor's degree from Harvard University (1956), his Master's degree in organic chemistry from the University of Wisconsin (1958), and his Ph.D. in chemistry from the University of Pennsylvania (1965).").

54 *See* A. Laakmann, *The New Genomic Semicommons*, UC Irvine Law Review 1, 5 (2015 Forthcoming) (viewed on Dec. 4, 2014). Available at SSRN: http://ssrn.com/abstract=2509571 ("DNA undoubtedly is an essential molecule that contains the requisite molecular code for intracellular protein formation. But it is hardly unique in its capacity to embody information...Indeed, all human functions stem from complex cascades of signaling events between biological molecules within and across cells.").

55 *See* Association for Molecular Pathology *supra* note 49 at 1314.

56 *See* LAAKMANN *supra* note 55 at 5 ("In addition to being logically incoherent, the Supreme Court's Myriad opinion reinforced the lower courts' apparent misunderstanding of fundamental principles of biochemistry.").

57 *See* D. Burk, *Edifying Thoughts of a patent watcher: The Nature of DNA,* 60 UCLA L. REV. DISC 92, 95 (2013).

58 *Funk Bros. Seed Co. v. Kalo Inoculant Co.* 333 U.S. 127 (1948).

59 *See* Diamond *supra* note 36.

60 *Association for Molecular Pathology v. Myriad Genetics,* 569 U.S. 12-398, 414 (2013) ("...the lab technician unquestionably creates something new when cDNA is made. cDNA retains the naturally occurring exons of DNA, but it is distinct from the DNA from which it was derived. As a result, cDNA is not a "product of nature" and is patent eligible under §101...").

61 Brief for the United States as Amicus Curiae, p. 33, *Association for Molecular Pathology v. Myriad Genetics,* 569 U.S. 12-398, 414 (2013).

62 *See* Laakmann *supra* note 55 at 5 ("The Court reasoned that Myriad's claims on isolated DNA did not satisfy § 101 requirements because Myriad's proprietary interest was in naturally determined genetic information, not chemical compositions. Nonetheless, the Court held that cDNA is patent-eligible because such molecules are man-made laboratory creations, even though the nucleotide sequence of cDNA is also dictated by nature.").

63 *Parke-Davis & Co. v H.K. Mulford Co.* 189 F.95, 103 (C.C.S.D.N.Y. 1911).

64 *See* Laakmann *supra* note 55 at 8 ("The Myriad Court thus left unresolved questions about the contours of the law of nature exception to patent eligibility and its relationship to the product of nature exception. Perhaps the Court was wary of probing Mayo's shaky foundations – like the product of nature doctrine, the law of nature doctrine has no solid scientific underpinning.").

65 H.R.3967, 107th Cong., (2002).

66 H.R.2966, 107th Cong., (2002).

67 *See* Gold, *supra* note 16 at S43 ("The first real opposition within government came from two bills introduced by Representative Lynn Rivers. Rivers was inspired to introduce this legislation by the growing controversy surrounding Myriad's patents on the *Brca1* and *Brca2* genes and diagnostic test...").

68 H.R.2795, 109th Cong., (2006).

69 Pub. L. No. 112-29, 125 Stat. 284 (2011). Also *see See* Joe Matal, A Guide to the Legislative History of the America Invents Act: Part I of II, 21 FED. CIR. BAR J. 435, 438 (2012) ("The first version of what became the AIA was introduced on June 8, 2005 ....").

70 *See* Joe Matal, A Guide to the Legislative History of the America Invents Act: Part II of II, 21 FED. CIR. BAR J. 435, 438 (2012) ("The first version of what became the AIA was introduced on June 8, 2005 ....").

71 *See* Morelli Gradi, G., Patenting biotechnologies: the European Union Directive 98/44/EC of the European parliament and of the council of 6th July 1998 on the legal protection of biotechnological inventions. Forum (Genova) 9, 25–36. (1999).

72 "Basic" or "fundamental" research describes research undertaken primarily for the advancement of scientific knowledge, without the specific goal of practical applica-

tion. *See* DONALD STOKES, PASTEUR'S QUADRANT: BASIC SCIENCE AND TECHNOLOGICAL INNOVATION 64 (1997).

73 *See* LAAKMANN *supra* note 55 at 29 ("...a tragedy of the commons can occur with respect to information resource production where individuals have insufficient incentives to invest their privately owned labor and tools into R&D and to disclose their results...the fear that others will defect by withholding meaningful information could compel players to defect in order to avoid the sucker's payoff."). While the story of Myriad Genetics points to a need for the patent system to ensure patented material can be freely used for non-commercial research in the biotech industry specifically, this feature of the patent system need not necessarily apply to other industries. In the case of the biotech industry, patents are critical to incentivize companies to undertake the risky and expensive research projects inherent to biotechnology (*See* M.A. Heller and R.S. Eisenberg, *Can Patents Deter Innovation? The Anticommons in Biomedical Research*, 280 SCIENCE 5364 (1998) ("Patents and other forms of intellectual property protection for upstream discoveries may fortify incentives to undertake risky research projects...")). In contrast, industries that rely less on the issuance of patents to secure value for innovation may not benefit from explicit free use policies. *See* John M. Golden & Hannah J. Wiseman, The Fracking Revolution: Shale Gas as a Case Study in Innovation Policy, 64 EMORY L.J. 954, 1000 (2015) ("[A] common part of the origin story of the shale boom is that its beginnings were fundamentally patent free.").

74 *See* J. GOLDEN, *supra* note 45 at 1095 ("With Congress apparently unable to do an adequate job of developing legal doctrine on subject-matter eligibility, we are back to the debate over whether responsibility for that task should lie primarily with the courts or an administrative agency.").

75 *See* D. Gitter, *International Conflicts over Patenting Human DNA Sequences in the United States and the European Union: An Argument for Compulsory Licensing and a Fair-Use Exemption,* 76 N.Y.U. L. Rev. 1623, 1685 (2001) ("For example, Celera has stated that academic researchers can use its sequence data free of charge...").

76 *See id* at 1684-85 ("In addition to implementing a compulsory-licensing system for human DNA sequences, which would preserve incentives for private investment in research and development, Congress also should codify an experimental-use exemption for public-sector researchers at the federal level and nonprofit researchers.").

77 *See id* at 1687 ("Essentially, fair use in copyright law turns on the degree to which the infringer 'has added substantial value to the original work and "transformed" it in some way. Such an exemption is appropriate for biotechnology research, especially research relating to homologous DNA sequences, since later inventors often contribute significant information about a particular sequence's function, thereby transforming scientists' understanding of that sequence.").

About the uncountability of the number of irrational powers of irrational numbers evaluated as rationals and solutions' estimation for $x^x=y$ and $x^{x^x}=y$

The University of Texas at Austin

Undergraduate Research Journal

Volume 15
Number 1
Spring 2016

Anca Andrei

Abstract

It is well known that numbers are often used to define numbers. For example, two integers are used to define a rational number and two real numbers are used to define a complex number. It might be expected that an irrational power of an irrational number would be an irrational number. Despite this expectation, it is possible for an irrational power of an irrational number to be a rational number. For many generations, this circulated as a non-constructive proof by contradiction in logic for discrete mathematics textbooks and college courses [5]. Since the '80s, a constructive proof circulated orally

such as $(\sqrt{2})^{log\ \sqrt{2}3}$ equals to 3. A written proof was published in 2008 by Lord [3], which immediately implies that there is a countable number of such pairs of irrational numbers with this property.

The first contribution of this paper is to show that there is actually an uncountable number of such pairs of irrational numbers such that the power of one to the other is a rational number. This is an improvement of the previous believed result.

Marshall and Tan answered the question of whether there is a single irrational number a such that a^a is rational 4]. They proved that given $I=((\frac{1}{e})^{\frac{1}{e}},\infty)$, then every rational number in I is either of the form a^a for an irrational a or is in the very thin set {1, 4, 27, 256, ..., n^n, ... }. It seems a challenging task to analytically solve the equation $x^x=y$ for any real y. To the best of our knowledge, there is no work on finding x from a given y. We will prove that $ln(ln(y))<x$ for $y>e$ and $x<ln(y)$, for $y>e^e$.

Hence, the second contribution of this paper is to estimate the real x such that either one of the equations $x^x=y$ or $x^{x^x}=y$ holds, for a given y.

MSC classification: 03F03 (Mathematical logic and foundations - Proof theory and constructive mathematics, general)

1. Introduction

This work establishes that there exists an uncountable number of irrational numbers a and b such that a^b is a rational number.

In 1900, David Hilbert posed a list of challenging problems, including a general problem referring to whether $2^{\sqrt{2}}$ and $\sqrt{2}^{\sqrt{2}}$ are transcendental numbers, among other interesting powers. A number is *algebraic* if and only if it is the root of a polynomial equation with rational coefficients. On the contrary, a real number is *transcendental* if and only if the number is not algebraic. In 1934, Gelfond and Schneider proved, independently [1; 2; 6; 7], that if α and β are algebraic numbers with $\alpha\neq0$, $\alpha\neq1$, and $\beta\notin Q$, then α^β is transcendental. Following Gelfond-Schneider Theorem, both numbers $2^{\sqrt{2}}$ and $\sqrt{2}^{\sqrt{2}}$ are transcendental.

In many discrete mathematics books, such as [5], a non-constructive proof for the existence of irrational powers of irrational numbers that are rational is provided. For example, the proof described in [5] does not indicate any such numbers, but only a non-constructive proof based on the fact that $\left((\sqrt{2})^{\sqrt{2}}\right)^{\sqrt{2}}=2$. In more detail, the following non-constructive proof circulated for many decades [Ros2011]: "To show that there exist irrational numbers x and y such that x^y is rational, consider the number $\sqrt{2}^{\sqrt{2}}$. If it is rational, we have two irrational numbers x and y with x^y rational, namely, $x=\sqrt{2}$ and $y=\sqrt{2}$. On the other hand if $\sqrt{2}^{\sqrt{2}}$ is irrational, then we can let $x=\sqrt{2}^{\sqrt{2}}$ and $y=\sqrt{2}$ so that $x^y=(\sqrt{2}^{\sqrt{2}})^{\sqrt{2}}=\sqrt{2}^{(\sqrt{2}\cdot\sqrt{2})}=\sqrt{2}^2=2$. " One quick note is that according to Gelfond-Schneider Theorem, $\sqrt{2}^{\sqrt{2}}$ is transcendental, hence irrational. So, the first case in the above non-constructive proof cannot happen. Hence, only the second case holds, which now becomes a constructive proof.

We continue these works by stating a few results about the irrational powers of irrational numbers which are evaluated as rationals. It is easy to prove that $\left((\sqrt{n})^{\sqrt{n}}\right)^{\sqrt{n}}$ is rational if and only if n is an

even positive number, since $\left((\sqrt{n})^{\sqrt{n}}\right)^{\sqrt{n}}=n^{\frac{n}{2}}$. For example, $\left((\sqrt{4})^{\sqrt{4}}\right)^{\sqrt{4}}=4^{\frac{4}{2}}=16$, hence rational, while $\left((\sqrt{5})^{\sqrt{5}}\right)^{\sqrt{5}}=25^{\sqrt{5}}$, which is irrational.

In 2008, Nick Lord [3] published a constructive proof based on the fact that $(\sqrt{2})^{log_{\sqrt{2}}3}=3$. Both $\sqrt{2}$ and $log_{\sqrt{2}}3$ are irrational numbers, but their power composition $(\sqrt{2})^{log_{\sqrt{2}}3}$ is a rational number. Lord generalizes this problem to more than just one number. He proved that if $\frac{m}{n}$ is any positive rational with $m\neq n$, and p is any prime which is neither a factor of m nor of n, then $\sqrt{p}$ and $log_{\sqrt{p}}\frac{m}{n}$ are both irrational. Furthermore, their power composition $(\sqrt{p})^{log_{\sqrt{p}}\frac{m}{n}}$ is the rational $\frac{m}{n}$.

An interesting question is how many irrationals α and β exist such that α^{β} is rational. Lord's result implies that there are (at least) a countable set of numbers α and β such that α^{β} is a rational. A *countable* set is a set with the the same cardinality as the set of natural numbers. A set that is not countable is called *uncountable*. Cantor proved that the rational set of numbers and the algebraic set of numbers are both countable sets. He also proved that the set of real numbers is uncountable.

Since the set of algebraic numbers is countable and the set of real numbers is uncountable, it follows that the set of transcendental numbers is uncountable.

Marshall and Tan answered the question of whether there is a single irrational number a such that a^a is rational [4]. They proved that given $I=((\frac{1}{e})^{\frac{1}{e}},\infty)$, then every rational number in I is either of the form a^a for an irrational a or is in the very thin set {1, 4, 27, 256, ..., n^n, ... }. It seems a daunting task to analytically solve the equation $x^x=y$ for any real y. To the best of our knowledge, there is no work on finding x from a given y. We will prove that $ln(ln(y))<x$ for $y>e$ and $x<ln(y)$, for $y>e^e$.

The paper's contribution is two-fold:

1. to show that there is an uncountable number of pairs of irrational numbers such that the power of one to the other is a rational number;
2. to estimate the real x such that either one of the equations $x^x=y$ or $x^{x^x}=y$ holds, for a given y.

2. The result about uncountability of irrational powers of irrational that are rational

Lemma 2.1 *If a is a positive transcendental number and c is a positive integer, then $log_a c$ is an irrational number.*

Proof By contradiction, we assume that $log_a c$ is a rational number. Hence, there exists a positive integer p and a non-zero integer q such that $log_a c=\frac{p}{q}$. This means $a^{pq}=c$ which implies $a=(c)^{qp}$. The number $(c)^{qp}$ is algebraic because it is the root of the polynomial $X^p-c^q=0$ with rational coefficients. On the other hand, a is a transcendental number, so the equality $a=(c)^{qp}$ represents a contradiction. As a consequence, $log_a c$ is an irrational number and $a^{log_a c}$ is rational.

Lemma 2.2 *If a is a positive transcendental number, then there exists an irrational number b such that a^b is a rational number.*

Proof Let a be a positive transcendental number. Let $b=log_a c$, where c is a positive rational integer. Clearly, $log_a c$ is well defined because a and c are positive numbers. Since a is transcendental, $a\neq 1$ so the logarithmic function log_a is well defined.

By Lemma 2.1, $log_a c$ is an irrational number. Hence $a^b=a^{log_a c}$ is rational.

We are ready to prove the main result of this section.

Theorem 2.1 *There exists an uncountable set of irrational numbers a and b such that a^b is a rational number.*

Proof We choose a to be a positive transcendental number and $b=log_a c$, where c is an arbitrary positive rational number. According to Lemma 2.1, it follows that b is an irrational number. Based on Cantor's Theorem, the set of positive transcendental numbers is uncountable. Combining these aforementioned statements, this theorem is proved.

3 An estimation of the solutions $x^x = y$ and $x^{x^x} = y$, for given y

To obtain an estimation of the solution of $x^x=y$, we need an intermediate support lemma.

Lemma 3.1 *If z is a positive real, then $e^z>z+ln(z)$.*

Proof If $z\in(0,1]$, then $e^z>1>z+ln(z)$ and the conclusion holds.

If $z\in(1,\infty)$, then we define f: $R\rightarrow R$ given by $f(z)=e^z-z-ln(z)$. Obviously, f is a continuous and differentiable function on $(1,\infty)$, with $f'(z)=\frac{z\cdot e^z-z-1}{z}$. Defining g: $R\rightarrow R$ by $g(z)=z\cdot e^z-z-1$, we see that function g is continuous and has its first continuous derivatives. Its first derivative is $g'(z)=e^z+z\cdot e^z-1$ and its second derivative is $g''(z)=(z+2)\cdot e^z$. Clearly, $g''(z)>0$, for $z>1$. Hence g' is a monotone increasing function, so $g'(z)>g'(1)$. Since $g'(1)=2\cdot e-1$ is a positive real, g is also a monotone increasing function. Hence $g(z)>g(1)$. Since $g(1)=e-2$ is a positive real, $g(z)>0$. Thus $f'(z)>0$, $\forall z>1$. This implies that f is a monotone increasing function. Hence $f(z)>f(1)=e-1$. Since this quantity is positive, the inequality of this lemma holds for $z>1$, too.

Considering the equation $x^x=y$ for any real given y, Theorem 3.1 estimates the solution x.

Theorem 3.1 *Let us consider the equation $x^x=y$, for a given real y. Then, the following estimations of solution x hold:*

(a) if $y>e^e$, then $x<ln(y)$.

(b) if $y>e$, then $x>ln(ln(y))$.

Proof (a) Since $y=x^x$ and $y>e^e$, the inequality $x<ln(y)$ is equivalent to $x<ln(x^x)$, which is equivalent in turn to $x<x\cdot ln(x)$. Since $x^x=y>e^e$, it follows that $x>e$. Hence it follows that $ln(x)>1$, so $x<x\cdot ln(x)$ holds.

(b) By substituting $y=x^{x^x}$, we get an equivalent inequality $x>ln(ln(x^x))$, where $x\cdot ln(x)>1$. This inequality is equivalent to $x>ln(x\cdot ln(x))$, which is in turn equivalent to $x>ln(x)+ln(ln(x))$ for any real x such that $x\cdot ln(x)>1$. Since $x\cdot ln(x)>1$, it means that $x>1$. Thus, there exists $z>0$ such that $x=e^z$. The inequality to be proved becomes: $e^z>z+ln(z)$, Since this is the inequality from Lemma 3.1, we conclude the proof.

We are now ready to handle the exponentiated version of the previous equation, that is, $x^{x^x}=y$, for a given y.

Lemma 3.2 *If x>e, then the following statements hold:*

(a) $x<x\cdot ln(x)+ln(ln(x))$;

(b) $x>ln(x\cdot ln(x)+ln(ln(x)))$.

Proof

(a) Since $x>e$, it means $ln(x)>1$, so $ln(ln(x))>0$. Hence $x<x\cdot ln(x)+ln(ln(x))$;

(b) We prove a stronger inequality, that is, $x>ln(x+1)+ln(ln(x))$. Since $x>e$, it means $\exists z>1$ such that $x=e^z$. We need to prove $e^z>ln(z\cdot(e^z+1))$. We define f: $R\rightarrow R$ given by $f(z)=e^z-ln(z\cdot(e^z+1))$. Obviously, we observe that f is continuous and differentiable with first derivative $f'(z)=\frac{z\cdot(e^z)^2-e^z-1}{z\cdot(e^z+1)}$. So, $f(z)>f(1)>0$. Hence $x>ln(x+1)+ln(ln(x))$. The right hand side can be rewritten as:

$$ln(x+1)+ln(ln(x))=ln((x+1)\cdot ln(x))=ln(x\cdot ln(x)+ln(x)).$$

But $ln(x)<x$, so $ln(x)>ln(ln(x))$ by substituting x with $ln(x)$. Hence

$$ln(x\cdot ln(x)+ln(x))>ln(x\cdot ln(x)+ln(ln(x))).$$

Therefore $x>ln(x\cdot ln(x)+ln(ln(x)))$. The lemma is proved.

Theorem 3.2 *Let us consider the equation* $x^{x^x}=y$, *for a given* $y>e^{e^e}$. *Then the following estimations of solution x hold:*

(a) $x<ln(ln(y))$;

(b) $x>ln(ln(ln(y)))$.

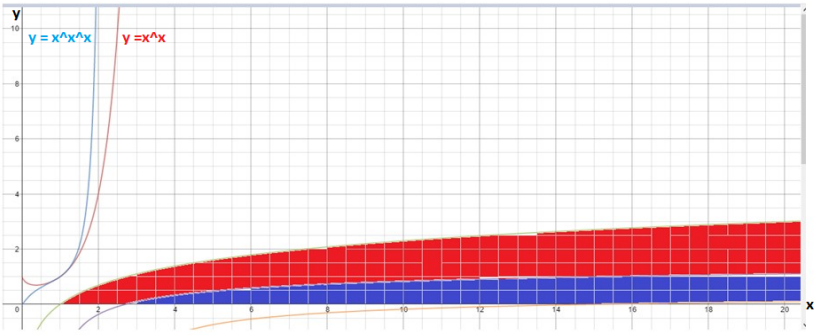
Proof Since $x^{x^x}=y$ and $y>e^{e^e}$, it means that $x>e$.

(a) The inequality $x<ln(ln(y))$ is equivalent to $x<ln(x^x\cdot ln(x))$ or $x<x\cdot ln(x)+ln(ln(x))$. By item (a) of Lemma 3.2, this inequality holds.

(b) Similarly, $x>ln(ln(ln(y)))$ is equivalent to $x>ln(x\cdot ln(x)+ln(ln(x)))$. According to item (b) of Lemma 3.2, this inequality holds.

A graphical representation for the approximation of the inverse of the two functions (these are, $y=x^x$ and $y=x^{x^x}$, respectively) is shown in Figure 1. The red line is $y=x^x$ and the red region represents the region of the inverse of that function. Similarly, the blue line is the graph of $y=x^{x^x}$ and the blue region shows the region of the inverse of that function.

Figure 1: The approximation of the two inverse functions



A graphical representation for the approximation of the inverse of the two functions discussed in the paper. The red line is $y = x^x$ and the red region represents the region of the inverse of that function. Similarly, the blue line is the graph of $y = x^{x^x}$ and the blue region shows the region of the inverse of that function.

Acknowledgments

We thank Dr. Ted W. Mahavier from Lamar University and to Dr. Ronny Hadani from University of Texas at Austin for their valuable comments while proofreading this work.

Conclusions

This paper proved that there exists an uncountable number of irrational numbers *a* and *b* such that *a* is a rational number. We estimated the solutions x for both equations *x=y* and Our results represent extensions of the works done by Lord in [3] and Marshall and Tan in [4].

References

A.O. Gelfond. Sur la septieme Probleme de D. Hilbert. *Comptes Rendus Acad. Sci., URSS Moscou*, vol. 2, pp. 1-6, 1934.

A.O. Gelfond. Sur la septieme Probleme de D. Hilbert, Bull. *Acad. Sci. URSS Lemingrade*, vol. 7, pp. 623-634, 1934.

Nick Lord. Math Bites: irrational powers of irrational numbers can be rational. *Mathematical Gazette*, vol. 92, pp. 534-534, 2008.

Ash Marshall, Yiren Tan. A rational number of the form *a* with *a* irrational. *Mathematical Gazette*, vol. 96, pp. 106-109, 2012.

K.H. Rosen. *Discrete Mathematics and Its Applications*. McGraw-Hill, Seventh edition, 2011, ISBN: 978-0-07-338309-5

T. Schneider. Transzendenzuntersuchungend periodischer Funktionen I. *J. reine angew. Math.*, vol. 172, pp. 65-69, 1934.

T. Schneider. Transzendenzuntersuchungend periodischer Funktionen II. *J. reine angew. Math.*, vol. 172, pp. 70-74, 1934.