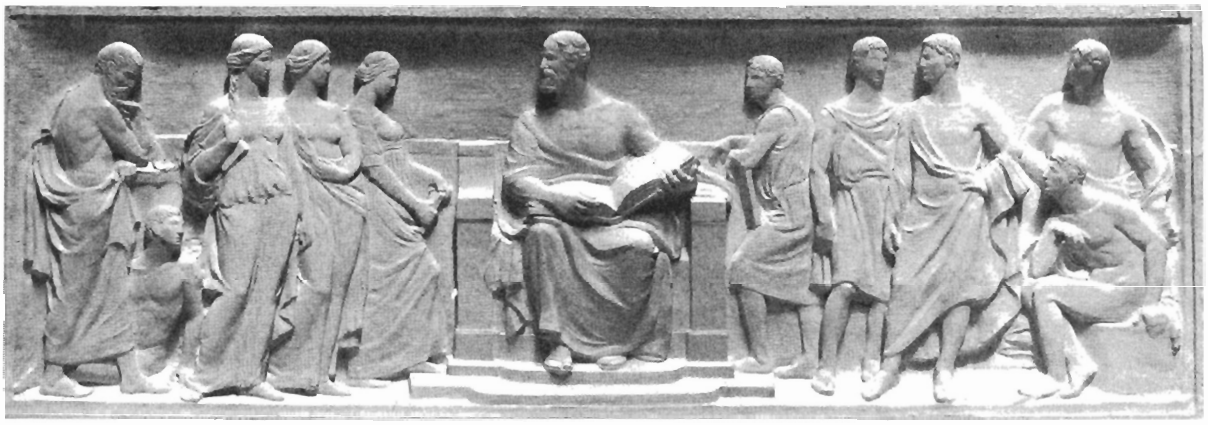
THE COLLEGE OF NEW JERSEY



JOURNAL OF STUDENT SCHOLARSHIP

VOLUME VIII
APRIL 2006

THE COLLEGE OF NEW JERSEY

Journal of Student Scholarship

VOLUME VIII
APRIL 2006

THE COLLEGE OF NEW JERSEY

Journal of Student Scholarship

VOLUME VIII

The College of New Jersey Journal of Student Scholarship is funded by the Office of Academic Affairs. It is published annually.

Preface

There are many ways to gauge academic standing, but what better measures beyond the quality, output, and vibrancy of student and faculty scholarship and creativity? As The College of New Jersey has been recognized by Barron's Guide to be among the 75 most competitive colleges and universities in the United States, there is continued cause to celebrate student scholarship. This volume presents but a fraction of the work done by students at TCNJ.

Although I have come to expect high quality work from students at The College of New Jersey, I continue to be surprised at how very good it really is. Indeed, the fostering of collaborative efforts between students and professors in search of original questions and thoughtful, researched responses speaks to the very heart of the intellectual enterprise. It unbinds students from the routine of "regular" classroom activities and allows their minds to soar. At the same time it frees faculty members to follow their own spirits.

Liberal learning, in its broadest sense, is at the heart of TCNJ's transformed curriculum. It calls all students to achieve intellectual and scholarly growth, to be mindful of and proactive toward civic responsibilities, and to engage themselves in the broad sectors of human inquiry. These principles are suffused into the entire curriculum.

The papers in this volume are the result of extraordinary efforts by students and professors; it is work that stretches their minds and expands their intellectual and scholarly horizons. At the same time, these papers cut across the broad sectors of human inquiry and collectively represent a multidisciplinary approach—the whole is not containable in any particular academic field. Even as specific issues of community, race, gender, and global perspectives may not be directly addressed, the fact that students share their work strengthens the community itself.

This Journal is its own demonstration of TCNJ's coming of age as an academic institution of the highest rank.

Robert J. Anderson
Director of Liberal Learning

A Message from the Editor

During the 2005–2006 academic year, The College of New Jersey concludes the sesquicentennial celebration of its inception in 1855. Over the intervening 150 years, the College has grown far beyond the dreams of its founders, becoming a comprehensive institution built on an impressive foundation in the liberal arts and sciences. Under its present administration, TCNJ has become increasingly recognized as a model undergraduate institution, providing its students with the rich and varied curricula and accomplished and dedicated faculty of a distinguished liberal arts college combined with the accessibility and affordability of a state school. Since its founding in 1998, *TCNJ Journal of Student Scholarship* has grown with the College. The quality of the wide range of essays published in this volume reflects the firm commitment of students, faculty, and administration to premiership among the nation's undergraduate colleges.

I wish to thank the many people who have contributed to Volume VIII of *TCNJ Journal of Student Scholarship* beginning with the students whose papers are contained herein and the professors who encouraged and mentored their work. The administration of President R. Barbara Gitenstein and Provost Stephen R. Briggs graciously provided moral and financial support and release time, without which this volume would not have been possible. The staff of the Office of Academic Affairs, including especially Nancy Freudenthal and Ellie Fogarty, has been kind, patient, and helpful with a variety of concerns and inquiries. I also extend thanks to Anthony Marchetti and Cindy Friedman of the Office of Public Affairs. For consultation on production matters, I thank Lisa Angeloni, Director of Admissions. I offer special thanks to the members of the editorial board of the *Journal* for lending their time and expertise to the reviewing of papers in the midst of many other responsibilities. Paulette LaBar of the Department of English kindly and expertly assisted in the editing of the essays printed in this volume and periodically updated the Web site of the *Journal*. I would also like to thank my wife, Jeanne Conerly, and our daughter, Katherine, for their support and understanding during an extraordinarily busy year. To Professor Robert Anderson, whose many responsibilities include administering the new Liberal Learning program, I owe thanks for his contributing the preface to Volume VIII. Finally, I would like to thank Associate Editor Romulo Ochoa for his wise advice and careful editing of the science essays. As these acknowledgments indicate, this volume is the product of a community-wide effort.

For information about the *Journal*, including submission procedures, format requirements, and application forms, please telephone the editor at 609.771.2155 or contact him by e-mail at dventuro@tcnj.edu; or, visit the *Journal* Web site at: http://sjournal.intrasun.tcnj.edu.

Now, I invite you to turn to the essays published in this volume and to read and enjoy them.

David Venturo
Editor
Department of English

Contents

VOLUME VIII

Gwrthryfel: Welsh Rebellion in *1 Henry IV* and in British History

Francine Roche,
The College of New Jersey
Faculty Sponsor:
Professor Jo Carney,
Department of English

Abstract

Throughout English history, the Welsh people have often been depicted as boorish, violent miscreants and rebels, steeped in mysticism and superstition and living in a wild, inhospitable, and ungodly land. By studying historical and cultural records, the reasons behind repeated Welsh rebellions become clear: The Welsh rebelled against the English as a result of centuries-long attempts at territorial, economic, cultural, and even linguistic subjugation. Even after the 1543 annexation of Wales, by England, Anglo-Welsh relations remained strained, widening the cultural gap that already existed between the English and their "foreign" neighbors. When William Shakespeare wrote *1 Henry IV,* many Elizabethans still viewed their Welsh neighbors with much distrust and suspicion despite the Tudors' Welsh lineage, and Shakespeare inscribed these cultural tensions and societal biases into his play. It is almost certain that one of Shakespeare's historical sources for *1 Henry IV* was Raphael Holinshed's *Chronicles of England, Scotland, and Ireland,* and in accordance with Holinshed's representation of the Welsh, Shakespeare depicted Henry IV's Welsh adversary, Owen Glendower, as an arrogant rebel and superstitious dissenter who had to be stopped at all costs. While his portrayal of Glendower was stereotypically negative, Shakespeare ingeniously managed to highlight many positive Welsh attributes, particularly through the character of Glendower's daughter, Lady Mortimer. In addition, by portraying Edmund Mortimer, an Englishman, and Lady Mortimer in a loving relationship, Shakespeare underscored the positive aspects of a strong Anglo-Welsh union.

Introduction

There are few noticeable differences visible when crossing into Wales from England—so few that many people—tourists and British alike—don't even realize that they've gone from one country into another. After all, the grass is as green in Wales as it is in England, the roads are similarly marked and paved, the country cottages are equally quaint, and the sky, dotted with heavy, cottony clouds, is the same blue-gray. The one conspicuous difference: the language suddenly changes from the familiar English to an incomprehensible Celtic dialect. Road signs bearing greetings such as "Welcome to London" or "Welcome to Bath" give way to those displaying the Welsh greetings, "Croeso i Gymru" or "Croesawiad y Gwalia," both of which mean "Welcome to the land of the Welsh" or "Welcome to Wales." Moving deeper into Welsh territories, commonplace English towns such as Exeter, Dorset, and Shrewsbury give way to place names as alien and unpronounceable as Aberystwyth, Llandrindod, Caernaervon, and Ystradgynlais. To the non-Welsh, these unfamiliar words are foreign-sounding and strange, denoting not just their entry into a different territory but a territory inhabited by a people who for over a thousand years managed to hold on to their culture, language, and identity.

The people themselves are welcoming and pleasant; they look no different from their English or Scottish counterparts. Some don't even have the thick brogues or burrs that one might find in the Scottish highlands, in industrial English towns like Manchester or Liverpool, or secluded havens such as Cornwall or the Cotswolds. Yet something is different about these people—these Welsh who still like to call themselves "Cymry" even in modern times—something that is intangible and indefinable. This difference is what has plagued them as "the other" for as long as anyone can remember and it is what has set them apart from the neighboring English for centuries.

It is this very difference that sets the characters and the land apart, both linguistically and culturally, in Shakespeare's *The First Part of Henry IV.*[1] Unquestionably, the primary narrative in *1 Henry IV* deals with Prince Hal and his transformation from princely rogue to the venerable and worthy son of Henry IV. Throughout the play, the presence of the irrepressible Welshman, Glendower, and the underlying Welsh rebellion makes for an intriguing subplot, chiefly because the Glendower rebellion lasted 10 years of Henry's 14-year reign. Henry himself invaded Wales five times during that period in an effort to stanch the Welsh rebellion (Norwich 155). While it is important from a historical perspective to understand why Henry found it imperative to contain the Welsh uprisings, it is just as vital to understand the reasons behind the Welsh rebellion. What made the English and Welsh—people who not only shared the same Celtic background but the land as well—so violently antagonistic toward each other? What caused them to go their separate ways? What was it that made the Welsh so resistant to English rule? What made the English want to conquer the Welsh so desperately?

When Shakespeare wrote *1 Henry IV,* many Welshmen already lived in London, the seat of the kingdom and one of the most populated cities of the time. Despite living alongside each other, the English still viewed the Welsh with distrust and suspicion; this circumspection only widened the cultural gap that already existed between the English and their "foreign" neighbors. Living within London's politically charged atmosphere, it is likely that Shakespeare knew some Welshmen and, being a topical writer, he would not have been blind to English prejudice against the Welsh. But did Shakespeare view the Welsh with the same disdain most English did, or was he open to their cultural and linguistic differences? Was *1 Henry IV* written with a cultural bias? Were the Welsh portrayed irreverently to please English audiences, or was it written to reinforce Anglo-Welsh relations, especially since the Tudors had strong Welsh ties?

How did Shakespeare's reading of history, especially within the context of the works of Raphael Holinshed and Edward Hall, affect the way he wrote *1 Henry IV,* particularly in his depiction of the Welsh? After all, Shakespeare portrayed Owen Glendower as the "typical" Welshman: arrogant, rebellious, mystical, unruly. While his treatment of Glendower—and in essence, all the Welsh—is quite tame in comparison to Raphael Holinshed's view of Wales and its people, there is still a palpably negative connotation to his depiction. It is important to understand why Shakespeare portrayed the Welsh the way he did, how much of his depiction was accurate, and how Elizabethan culture, flavored by history, by personal and social biases, and by past rulers, influenced his writing and, conversely, how his writing affected his audience.

Finding answers to these questions will require an understanding of the Welsh character. It will also be necessary to revisit the historical beginnings of Britain when England, Wales, and Scotland were one country, before the Anglo-Saxon and Norman conquests. Apprehending how, when, and why the Welsh rebellions started and the reasons the Welsh opposed English rule, annexation, and occupation is essential. Conversely,

a look at why the English repeatedly strove to conquer Wales will explain why Henry IV spent a good portion of his time, effort, and money attacking the Welsh, with Owen Glendower, Henry's sworn enemy, at its head. Altogether, a more thorough awareness of Anglo-Welsh relations will help explain the Welsh rebellion in *1 Henry IV* as well as the treatment of the Welsh by Elizabethans, including Shakespeare.

The Isle of Prydein

"Britain, the best of the islands, is situated in the Western Ocean, between France and Ireland" (Geoffrey 53). So wrote Geoffrey of Monmouth, who in the 12th century was commissioned by Geoffrey Plantagenet, the father of Henry II, a Norman king, to compose and record the history of the Normans' newly conquered land. In this text, titled *Historia regum Britanniae* or *The History of the Kings of Britain,* Geoffrey recorded the island's history from ancient times to just about the end of the seventh century. He related how the island and the people were named, then detailed the lives of all British kings, from Brutus to Cadwalader, including the histories of King Lear and Cymbeline, two other kings of whom Shakespeare wrote.

Geoffrey claimed that the island was named after Brutus, great-grandson of Aeneas, who called the island "Britain" and the people "Britons" after himself (72). Before Brutus died, he divided his kingdom among his three sons: Locrinus, Kamber, and Albanactus. The first-born, Locrinus, inherited the part of the island called Loegria, named after him; Kamber, the middle son, inherited the western district, which became known as Cambria. The northern part of the island went to the youngest, Albanactus, and was named Alba (75). Loegria constitutes most of current England; Cambria is now Wales, while Alba is Scotland.

In reality, however, the inhabitants of the island were Celts who shared the same lineage as those who had migrated from Eastern Europe and westward into Gaul.[2] After warring with the Romans, many of these Gaelic Celts traveled north and populated Ireland and Britain, which was known to the Greeks as Ρρεττανοζ or Prydein (Snyder 67). Her people were called the Prittani or Brittani, the Latin *Britannus.* These Celts became the Britons; they divided their lands by tribes, and named themselves accordingly. Conceptions of English or Welsh were nonexistent. They were just Celts; they were all Britons. They had a name for themselves: Cymry, or "fellow countrymen" (Snyder 72).

Sub-Roman Britain and the "Adventus Saxonum"

The Roman occupation of Britain lasted for 400 years.[3] After the Romans withdrew from the island in the beginning of the fifth century, the Britons fell into a series of civil wars.[4] Many early British kingdoms grew out of old tribal systems, but with no Romans to police them, these chieftains and kings warred with each other over land disputes. While these kingdoms battled each other, the marauding Irish, Scotti[5], and Picts[6] began raiding and pillaging Briton towns and villages. The Britons asked Rome for help, and at first the Romans complied, sending troops to push back the invaders. But the Britons proved incapable of protecting themselves, so after repeated incursions and numerous calls for help, the Romans told the Britons they were on their own and had to deal with their problems themselves.

Around this time, a strong Briton leader emerged and managed to curtail Irish and Pictish incursions into Britain. Many historians refer to him as Vortigern, either a name or a title which means "superior ruler" (Morris 2). But Gildas, a seventh-century monk, referred to him as *superbus tyrannus,* proud tyrant. Why did Gildas refer to Vortigern as a tyrant? Because he and his council of Briton leaders were behind the collapse of Briton society. As Gildas wrote in his seminal *De Excidio Brittaniae,* or *The Ruin of Britain:*

> [T]hey convened a council to decide the best and soundest way to counter the brutal and repeated invasions and plunderings by the peoples I have mentioned. ... Then all the members of the

> council, together with the proud tyrant ... devised for our land ... that the ferocious Saxons (name not to be spoken!), hated by man and God, should be let into the island ... to beat back the peoples of the north. ... Of their own free will they invited under the same roof a people whom they feared worse than death. (26)

This was the beginning of the period known as the *Adventus Saxonum*, or the Anglo-Saxon invasion (Snyder 249). Bede, a ninth-century monk, who relied heavily on Gildas's accounts, explained in the *Ecclesiastical History of the English People:*

> [T]he Angles or Saxons came to Britain at the invitation of King Vortigern in three longships, and were granted lands in the eastern part of the island on condition that they protected the country: nevertheless, their real intention was to subdue it. They engaged the enemy from the north, and having defeated them, sent back news of their success to their homeland, adding that the country was fertile and the Britons cowardly. Whereupon a larger fleet quickly came over with a great body of warriors. (62)

These Anglo-Saxons, culled from three very powerful Germanic tribes called the Angles, Saxons, and Jutes,[7] were drawn to their western neighbor. Plagued by overpopulation and barely arable land, the Saxons welcomed any opportunity to move westward into Britain. The first Saxon settlements were in Kent and along the southeastern coast. John Blair, in his essay on the Anglo-Saxon period, observes, "With the natural fertility of lowland Britain, and the evidence that its inhabitants deliberately imported mercenaries, this flooding of coastal settlements helps to provide an explanation for the Migration" (54). Additionally, Charles Thomas writes: "The Anglo-Saxons had through a variety of ways come to know this once rich, agriculturally still-desirable part of the empire: it is abundantly clear that their overwhelming interest lay in the countryside, often well farmed and an arena ripe for expanding settlement" (40). The Saxons had come to settle Britain and make her their own.

The mercenaries—known to the Romans as *Saxones* and to the Britons as *Saeson* (Thomas 39) or English (Morris 2)—were notoriously brutal killers. Vortigern was aware of their cruel reputation, yet he brought them to Britain willingly, out of desperation. Pictish and Irish raids became more brazen as time went on, and he needed ruthless warriors to counter his enemies. Gildas remarked, "The island of Britain was once occupied by Britons, who ruled it all, from sea to sea, before the foreigners arrived and scattered the *cives*, who, if they could win back God's favor, would expel the foreigners and win back the land of the Britons" (quoted in Snyder 79). After promising to stave off the Picts and Irish, the mercenaries—the Angles, in particular—made an alliance with the Picts and turned their arms against the Britons (Bede 62). The Anglo-Saxons' duplicity was not surprising. By their agreement, each spring Vortigern had to relinquish more land to the ever-increasing Saxons. He did this for years until nearly all the land on the east coast belonged to the mercenaries. Still, these lands were insufficient for the burgeoning Saxon population, and as a result, the Saxons began their harassment of the Britons.

Their attacks began in the eastern coastal villages; they made their way westward, burning, massacring, and pillaging as they went along. By the end of the sixth century, the invaders controlled half of Britain (Williams 3). And what of the Britons? What became of them? Many were either killed or forced into slavery. A large number, however, chose to leave the eastern part of the island, moving west and south instead. Those who moved south either settled in Dumnonia, the area now composed of Cornwall, Dorset, and Exeter, or sailed across the channel towards Armorica. Those who moved west settled in Cambria. As Bede laments:

> A few wretched survivors captured in the hills were butchered wholesale, and others, desperate with hunger, came out and surrendered to the enemy for food, although they were doomed to lifelong slavery. ... Some fled overseas in their misery; others, clinging to their homeland, eked out a wretched

> and fearful existence among the mountains, forests and crags, ever on the alert for danger. (64)

The main Briton enclave now became Cambria, and Blair explains that "Refugees from the east had doubtless swelled its population. Christianity survived, and with it some distinct traces of Roman culture" (57). Archaeological records confirm that drastic depopulation occurred in Britain in the fifth and sixth centuries, and this decline is the clearest sign that Briton society was shattered (59). Only in Armorica and Cambria did substantial Briton cultural elements, such as their traditions, histories, and language, survive. In fact, by the middle of the fifth century, the term Brittani, what the Britons called themselves, took on a different meaning: Not only did Brittani refer to the people, but it also took on a geographic meaning. In Gallic texts, Brittani referred to those Briton refugees who settled in the Armorican peninsula (Snyder 67), hence the modern-day name for Armorica: Brittany (Thomas 47).

The Saxons tried to follow the Britons into Cambria but were unsuccessful. The rough terrain and harsh weather made incursions into Cambria all but impossible. As a result, a few early Briton kingdoms survived in Dumnonia and Cambria (Williams 3), while the Cambrian kingdoms of Gwynedd, Powys, Dyfed, and Gwent were solidified by the middle of the sixth century (Blair 57). The Dumnonian kingdoms held out against the Saxons until the middle of the ninth century, and thanks to this relatively late conquest, many archaeological Briton remains survived in Cornwall, Exeter, Dorchester, and Ilchester (58). The Cymry (Britons and Cambrians alike) employed guerilla tactics to protect themselves, using the terrain to their advantage every time the Saxons attacked. The Cymry were so successful at repelling Saxon conquests that in the late eighth century, Offa's Dyke was constructed by the Saxons to protect themselves from Cymry counterattacks (Blair 74, Thomas 112).

The rebellious Cambrian Britons called themselves Cymry or Cymbrogi, meaning "brother countrymen" (Thomas 47), but the Saxons referred to them as the *Wealh*, meaning slave, stranger, or foreigner (Blair 59). The Saxons never referred to *Cambria* but called the lands west of Saxon territory *Wallenses*, which eventually became Wales (Snyder 72). So the Britons not only lost their land and their homes to the Saxons, but they lost their cultural identity as well. No longer were they Britons or Cymry; they were the Welsh. Britain and Cambria no longer existed, but *Angelcynn* and Wales did.

While the Anglo-Saxons never physically conquered Cambria or Alba, neither place existed by the middle of the Anglo-Saxon period. Brutus's divisions of Britain—Loegria, Cambria, and Alba—were no more. In their place were *Englaland, Wales,* and *Scotland.*

The Norman Invasion and the Breton Hope

The Anglo-Saxon period came to an end in 1066 with the arrival of William the Conqueror and his defeat of the Anglo-Saxon king Harold Godwinsson. By this time, divisions between the Anglo-Saxons—or the English, as they now called themselves—and the Welsh were sharply delineated. The English were united culturally and ethnically, and the Welsh, also united by a common language and faith, considered themselves superior to, and separate from, the English. The British Isles alone emerged from the Middle Ages using vernacular languages (English, Welsh, Pictish, and old Irish) not derived from Latin or Greek (Snyder 251).[8]

Even though William defeated the English, his reign was not an easy one initially. The English resented becoming an oppressed majority in their adopted homeland, with a new royal family and ruling class as well as a new culture and language. In the first three years of William's rule, numerous uprisings occurred against the Normans, including rebellions in the Welsh Marches (Gillingham 104). It took five years before William considered the Norman conquest complete, and only then did he turn his attention towards Wales.

Wales in the 11th century was a collection of small kingdoms with unstable borders in mountainous country. These kingdoms expanded and contracted with the politics of the time, and although English kings claimed overlordship, they were never able to conquer Wales. The grueling topography and weather, as well as the lack of roads, made it difficult to subdue the Welsh, and as a result, any colonizing efforts were confined to the lowlands and river valleys of the south (136). For close to 200 years, Wales was at war with the English. The border country, however, was still occupied by English "Marcher" barons such as the Mortimers, whose duty was to hold the Welsh frontier. In return for keeping the Welsh at bay, the Marcher lords were masters of their own lordships, unbound by royal law (Schama 148).

Still, the Normans were persistent. When Henry II was crowned king, he quickly assessed the political advantage to be gained from Geoffrey's work (Berthelot 37).[9] Both Henry and his father believed there was an advantage to be gained from legitimizing the position of the indigenous Welsh. By linking the old Briton heritage to that of the continent (especially France), they could prove that the Normans were destined to be rulers of England. Henry knew he needed the support of the Welsh against the Saxons, who continued to balk at Norman rule. Instead of trying to unite these two dissenting and quarrelling factions, he would gain the advantage by setting them against each other. But as much as the Welsh hated the Saxons, they were slow to rally to the Plantagenet king's aid (38) and Henry knew exactly how to exploit this tentativeness.

Part of the Welsh's hesitation was their belief in the Breton Hope, a sufficiently widespread belief among the Britons that in their time of greatest need, the Welsh king, Arthur, would return to set his people free from their subjugators. It was certainly no coincidence that Geoffrey devoted a large part of the *Historia* to chronicling the life of Arthur.[10] By recording the exploits of this legendary leader, Geoffrey fanned the Breton Hope's flames.[11] Henry, cognizant of this, fueled it even more by commissioning the Anglo-Norman cleric Wace to rewrite the *Historia* as the prose romance *Brut*.[12] By making Arthur a worthy model for both the Welsh and Normans, Henry emphasized what they had in common: namely their enemy, the Saxons. Henry wanted to show the Welsh that their future was inextricably linked to the Norman cause and that by legitimizing the legend, he was presenting them with Arthur's legitimate heir: himself. With the resurgence of Arthurian lore, however, the Welsh believed that salvation was close at hand, and this constituted an obstacle to Henry's political ambitions (Berthelot 39).

Angered by the lack of Welsh support, one of the last things Henry did before his death was destroy the Breton Hope. Both the *Historia* and the *Brut* ended with Arthur sailing off to the Isle of Avalon to heal after his last battle, and Henry set about finding his "Avalon."[13] The monks of Glastonbury Abbey, near Wales, were happy to oblige Henry with a location for the fabled isle. The Abbey, which had suffered serious fire damage, was not particularly prosperous. The monks quickly recognized the advantages to be gained by being a part of the Welsh Arthurian mythos. Indeed, shortly after Henry's death in 1189:

> [T]he monks of Glastonbury Abbey put the final touches to the revised version of the myth by 'discovering' the tomb of Arthur and Guinevere in the cemetery attached to the Abbey. It is not clear whether this piece of theatre had been planned with Henry's knowledge: what we do know is that the king had done much to further the abbey's fortunes during the final years of his reign. The discovery of the two skeletons ... guaranteed the authenticity of the legend and turned Arthur into a figure of undeniable historical reality, at the same time enabling ... the Plantagenets ... to bask in his reflected glory. (Berthelot 42)

What is clear, however, is that the discovery of these graves[14] dealt a fatal blow to the Breton Hope. The Welsh, once again, had a

piece of themselves brutally taken away, although this time, at the hands of the Normans.

The English Conquest of the Welsh

Every generation or so, a powerful Welsh king tried to unite the Welsh kingdoms (Schama 149). In Henry III's time, this king was Llewellyn ap Gruffydd of Gwynedd who, with his grandfather, Llewellyn ab Iorweth, had standardized Welsh law and had clearly demarcated boundaries in Wales. By the time Henry III and the King of the Scots recognized Llewellyn as Prince of Wales, Llewellyn was master of two-thirds of the Welsh territories. His court was not that of a primitive tribal king; if anything, it rivaled Norman and continental courts, with a full complement of falconists, harpists, and bards. In 1267, the Treaty of Montgomery recognized Llewellyn as "Prince of Wales" while he, in turn, acknowledged Henry III as his feudal overlord. While Wales was not recognized as a completely independent state, within the Welsh territories only a Welsh prince—and not the Marcher lords or any other colonist—was recognized as supreme ruler of that area (Schama 189). However, with the death of Henry III came a new English ruler, Edward I. Edward would leave an enduring legacy of bitterness, dotting the landscape of his conquered territories with massive fortresses in Scotland and Wales. Edward's goal was to conquer the old Celtic lands—Ireland, Wales, and Scotland—and he wanted an unconditional admission from their leaders that he was their feudal overlord. And this admission included their obligation to provide him with men and money where and when he chose (185).

In many respects, England, Scotland, and Wales were becoming more—not less—alike: All three were subject to the demanding policies of aggressive, highly intelligent, and dynamic princes who wanted to impose a single law and a single rule on disparate portions of their jurisdiction. However, the more alike these regions—and rulers—became, the more determined they were to remain autonomous (189).

When Edward succeeded to the throne, Llewellyn failed to perform the basic obligation incumbent on a feudal vassal: He declined to kneel at the feet of the new king, place his hands in the hands of his liege lord, and pledge fealty (189). Edward summoned Llewellyn five times for his pledge, but Llewellyn failed to answer any of the summonses; after the fifth time, Edward pronounced Llewellyn a rebel and declared war. Edward marched into Gwynedd and laid siege to Snowdonia, Llewellyn's home base, with nearly 16,000 men. Nearing winter, with the threat of starvation, Llewellyn yielded and paid homage to Edward. But Edward, still incensed over Llewellyn's behavior, stripped away all of Llewellyn's territories, leaving him only Gwynedd. To add insult to injury, the English built castles on Welsh soil as garrisons. When clashes over jurisdiction occurred, English justices would hear the cases only in English courts and routinely overruled Welsh laws, further reducing Llewellyn's dignity and power. Llewellyn and his younger brother Daffydd suspected that the reduction of their state was just a prelude to what, sooner or later, Edward meant to be a complete annexation. The only choice left to him seemed to be submission to the English king or resistance. And so in 1282, Llewellyn chose resistance, and the Welsh and English, once again, went to war.

During a truce in the fighting, John Pecham, Archbishop of Canterbury, attempted a sort of arbitration: Would Llewellyn consider a lordship in England worth £1,000 a year in return for handing over Gwynedd to the king? Llewellyn replied that the country was not his to give:

> Let it be known that since Snowdonia[15] is something that pertains to the principality of Wales which he[16] and his ancestors have held since the time of Brutus, his council does not permit him to renounce that land and take in its place a land in England to which he has less claim. (Schama 192)

It was this same proposal of Pecham's that provoked the Welshmen's stirring declaration

which insisted that even if Llewellyn or any other Welsh prince "should give overlordship to the King, they themselves would refuse to do homage to any foreigner of whose language, customs and laws they were ignorant" (166).

The war was over less than a year after it started, punctuated by Llewellyn's death. His head was sent to Edward, and with it went the genius that had sustained the Welsh rebellion. Daffydd, however, continued to wage war against the English until he was betrayed by some of his own men who were afraid of Edward's reprisals. Early in 1284, finishing the process begun by Henry II, Geoffrey Plantagenet's great-great-great-grandson Edward I finally presided over an Arthurian-style court in north Wales to formalize the annexation of that country.

As Simon Schama notes, "This was the beginning of a comprehensive exercise in cultural demolition and imperial control. ... [Edward] knew exactly what to rip out from the Welsh tradition to maximize their demoralization" (193). Edward systematically began the annexation: He demolished monasteries, including the one housing Llewellyn's grandfather's remains; Welsh relics and royal jewels, including Llewellyn's coronet, were taken away to England. As for the natives, they were now second-class citizens in their own land. They were forbidden to carry arms or have strangers stay overnight without English permission. They were allowed to keep some of their more esoteric laws as long as they understood that the king's laws could supercede theirs at any time (196). As for Welsh laws and customs, Edward announced "certain of them we have abolished; some we have allowed, some we have corrected, others we have added" (Gillingham 138). But the worst thing Edward did was take away the one thing the Welsh held sacred: their sovereignty. Edward's son, Edward II, was born in Caernarvon in 1284, and the baby was bestowed the title "Prince of Wales." This title had defined Welsh autonomy from the Anglo-Saxon period on, and Edward took that away. From that day forward, the title "Prince of Wales" was conferred on the heir apparent to the *English* crown; no true Welshman has held that title since Llewellyn died.

Glendower and Henry IV

Despite Edward I's conquests of Wales, the Welshmen were still distrusted by the English in the late 14th and the early 15th centuries. Celtic prejudice against the English increased with all the bitterness and resentment that an oppressed people were capable of: "The tyranny and cruelty of the English...are notorious throughout the world, as manifestly appears in their usurpations against the French, the Scots, Welsh, Irish and neighbouring lands" (Griffiths 166). Demoralized and subjugated, the Welsh once again turned to war; Welsh uprisings of the early 15th century culminated in the Owen Glendower rebellion of 1400–1410, which recalled prophecies of English expulsion from Wales. The English believed this latest upheaval vindicated their centuries-old suspicions and misgivings about the Welsh.

What caused the Glendower rebellion? Economics, mostly. England's most important industry at the time was textiles, and its success was based on the land. The finest wool in Europe was produced from very large flocks of sheep, and the wealthiest regions were the Welsh lowlands and the gently rolling hill-countries of the midland and southern shires (Griffiths 181). When the black plague hit in the mid-1300s, Wales was strongly affected. In an effort to conserve their incomes, some Marcher lords—the Mortimers included, who had vast estates in Wales—exploited and pushed the still-recovering Welshmen to continue producing wool for the English (189). The distress experienced by this plague-ridden people, oppression at the hands of English landowners, maybe even resentment at the removal of Richard II,[17] all combined to throw the country into revolt.

The Welsh rebellion started early in the reign of Henry IV. Returning from a

Scottish campaign in 1400, Henry had gone no further than Northampton when news of an uprising in Wales reached him. He was told that a long-standing quarrel between the Marcher Lord Grey of Rutheyn and Owen Glendower, an influential Welsh landowner, had erupted into open war. Glendower ravaged the border lands of Lord Grey, sacking several large English towns and terrorizing its inhabitants (Griffiths 222). Henry had no choice but to pursue Glendower, but Glendower and his men took to the woods, eluding Henry and his men (Norwich 133).

Glendower continued attacking the garrison castles and any Anglicized towns he came across. He and his guerilla forces exploited the mountainous terrain whenever the English army attacked. Glendower harassed and exhausted Henry and his men, often disappearing among rocks and caves. The Welsh rebellion was the most serious threat Henry IV had to face, and it became the most expensive to suppress. As a result, Henry often had trouble raising men and money. Parliament was very tightfisted with the king, and it refused to spend more than what it deemed necessary, including financing his wars. As for men, Henry sought to send armies of at least 4,000 men into Wales; this was a considerable number since his French incursions rarely exceeded 5,000 to 6,000 men. Glendower, for his part, occasionally mustered a warhost of 8,000; he also sought and received aid from France as well as from fellow Celts in Scotland and Ireland. He attended a few of Henry's parliaments, producing grand schemes for an independent Wales. His alliance with the Percies, though tenuous, was intended as a prelude to the dismemberment of Henry's realm (Griffiths 198).

In 1402, two years into Glendower's rebellion, the Welsh captured Edmund Mortimer, uncle of the young Earl of March, Richard II's heir to the throne. On hearing the news, Henry decided to put down the Welsh once and for all and planned a major campaign. But this expedition was no more successful than the one he had organized two years earlier. Once again, the Welsh refused to be drawn into battle; the weather was atrocious, and as described by both Shakespeare and Holinshed, this bad luck was attributed to magic spells. Henry returned to England once more, but this would not be the last time he tried to kill or capture Glendower. All told, Henry marched on Wales five times during Glendower's 10-year rebellion; he never won against the Welshman.

Prince Hal, on the other hand, had been posted to Wales since Henry's initial encounter with Glendower in 1400 when Hal was just 15; Hal had met Glendower many times in battle and was starting to make headway against the Welsh rebels by the time the Battle of Shrewsbury occurred. In 1403, Hal was appointed by the council to represent his father in Wales, and in May of that year, he invaded the country once more, destroying two of Glendower's castles.

At the age of 21, Hal finally achieved his first major victory against the Welsh: One of Glendower's sons was killed in battle (Norwich 142, 156). Two years later, Glendower's son-in-law and most important English ally, Edmund Mortimer, was killed in one of Hal's siege attacks. Glendower's wife, two daughters, and three Mortimer grandchildren were taken prisoner and sent to London, and Glendower was left without a stronghold. In 1410, Glendower launched his final attack against Hal, but three of his captains were captured and he barely escaped alive.[18] Glendower retreated to the mountains of Wales, alone and starving, and no longer a threat.

Knowing what we now know regarding the history of the Welsh and the abuses they sustained at the hands of the English—both Anglo-Saxon and Anglo-Norman—for close to 1,000 years, it is no wonder that the Welsh rebelled repeatedly against their oppressors. With the defeat of Glendower, the last pre-modern Welsh rebel leader, the conquest of the Welsh by the English was complete. Even though Henry IV never succeeded in defeating Glendower, it was during

his reign that the Welsh rebels were finally put in their place. The Welsh rebellion had finally come to an end.

The Elizabethans, Shakespeare, and the Welsh

Few English are taught the history of Wales in any detail. As depicted by Englishmen, the Welsh sing; they mine coal or tin or make wool; most of them are small and dark and their language contains many unpronounceable phonemes; much of Wales is wild. Most of these generalizations are nonsense (Thomas 105). To the early modern English, the Welsh stood as that "which is both familiar and strange, both a part of England and an alien land on the other side of the Severn" (Sullivan 13).

After the accession of Henry VII, many Welsh regarded him as a figure who would restore Wales to its former greatness.[19] Paradoxically, Elizabethans either feared the Welsh as potential rebels or dismissed them as harmless and simple rustics (Boling 47).[20] To the Elizabethans, a Welsh rebellion was still conceivable, even though the "rehabilitation of [Glendower] was in full flood now that his countrymen were well behaved" (47).[21]

In *1 Henry IV,* Shakespeare uses the language of the wild Welshwomen, Glendower, and Lady Mortimer to demonstrate "the fragile coexistence of English and Welsh, English and Other, in late Elizabethan England" (Lloyd 7). After Henry VIII's Acts of Union,[22] the English pressured the Welsh into learning and using the English language, which was the authorized language of government, law, and society. The Acts gave the Welsh all rights of English citizenship as well as economic opportunities, provided they embraced the changes specified in the Acts. Megan Lloyd explains that the eventual success of the Welsh people depended on their use of English. Pressures of commerce, education, and law all combined "to convince Welshmen that they had no civilised or ambitious future ahead of them except by claiming equality on those terms" (8). Success meant speaking English; retaining Welsh signaled failure. A linguistic chasm separated the people, and an irreverent attitude developed toward the Welsh language.

As a result of this suppression of their language, anti-English sentiment grew during the Tudor age even though it was common knowledge that Queen Elizabeth spoke some Welsh.[23] And this resentment was justified. The Welsh language was one of the few vestiges remaining of the old Briton ways and one that the Welsh were determined to preserve. Conversely, the English saw fit to exclude, avoid, or ignore the Welsh altogether. The language remained a potential threat, and ridiculing it was a way for the English to manage the "foreigners" (Lloyd 9).

Was Shakespeare aware of the long history of Welsh subjugation? Probably not, since his main historical sources were Holinshed and Hall, and Holinshed was not particularly fond of the Welsh.[24] But Shakespeare was most certainly aware of what was occurring during Elizabethan times; he would have witnessed the treatment of the Welsh by the English and would have found a way to write about it. In fact, two of his plays dealt specifically with the Welsh: *1 Henry IV* and *Cymbeline*.

Was Shakespeare's treatment of the Welsh in his plays fair or biased? This is somewhat more difficult to ascertain. Certainly, knowing that the use of the Welsh language was strongly discouraged, he would not have written something overtly pro-Welsh. However, he did create a character, embodied in the Englishman Edmund Mortimer, who shares a mutually loving and reciprocal relationship with Lady Mortimer. The Mortimers exemplify English-Welsh coexistence: They respect and love each other in spite of their differences. Even though Lady Mortimer remains incomprehensible and requires her father, Glendower, to translate for her and her husband, this shows the accord possible between the English and Welsh. Glendower, who is Welsh, studied in England. He learned the language of the people who, for centuries, sought to conquer his people, and he served

an English king, Richard II. Both Lady Mortimer and Glendower are strong, positive representations of the Welsh; although their language was banned and scorned, all attention is focused on them while they speak in their native tongue or while Lady Mortimer sings, and in this, Wales is a very strong presence (Lloyd 13).

While Welsh was still an unwelcome language in Shakespeare's time, it was spoken and performed onstage even though using Welsh in public was a reminder that the Welsh continued to defy English suppression. In fact, many Elizabethans who saw *1 Henry IV* would have understood Lady Mortimer's song, as well as the conversation between Glendower and his daughter, without difficulty (Lloyd 9). Knowing the tensions between the English and Welsh, especially with regard to the use of the Welsh language, Shakespeare demonstrated English and Welsh coexistence as a function of language. He showed a nation that was silent and repressed, yet remained resilient and strong.

The term *Wales* marks a perceived alienness of that country's inhabitants, an alienness that exists side by side with its opposite: the sense of Wales as something familiar. In the early modern period, people almost certainly questioned the place of Wales within English culture and society: Is it a foreign country or a distant province? Are the Welsh familiar outsiders or strange cohabitants of England? In creating such complex Welsh characters in *1 Henry IV*, Shakespeare conveys a sense of Welsh "resentment and threat" while still remaining optimistic that positive Anglo-Welsh relations can exist (Boling 47). And while the Welsh language is suppressed, it is not eliminated, as represented by Lady Mortimer. Shakespeare's Welsh characters offer "a model of assimilation and coexistence that spoke to an England aware of the need to control another unruly Celtic neighbor. Thus, seen by an audience aware of the problems with Ireland, Shakespeare's *1 Henry IV* presents a way for England and the Other, in this case Wales, to coexist" (Lloyd 7).

The Welsh suffered for close to 1,400 years at the hands of the English; their society was subjugated; they lost their lands and were forced to flee their homes; they lost their cultural identity including the name they gave themselves; and they were constantly forced to bow down to English kings and Marcher lords. Yet through it all, they managed to keep their values intact, including their history, their language, and their Celtic heritage. While it is true that they rebelled repeatedly throughout history, their rebellions were not uncalled for; where the English saw a stubborn, obstinate people that *had* to be conquered, the Welsh considered rebellion a means of survival. Even though Shakespeare may not have been familiar with Welsh history, he certainly portrayed them as a strong, worthy nation, deserving of people's respect.

END NOTES

1 For this paper, the Bedford edition of *The First Part of Henry the Fourth*, edited by Barbara Hodgdon, was used.

2 Many books and Web sites discuss Celtic migration, as well as Celtic culture. The following texts detail Celtic cultures, not just of the Goidelic or Britonic (Irish, Scottish, Welsh, and British) Celts, but the eastern European Celts as well. Peter Berresford Ellis's book *The Celtic Empire: The First Millennium of Celtic History 1000 BC–AD 51* is a very good starter for someone who is interested in getting to know the Celtic people and their history. Dáithí Ó Hógáin's *Celtic Warriors: The Armies of the First Great Peoples in Europe* is another good source, especially because it deals with the Roman and Celtic wars and the reasons for (or behind) the Celtic and Gaelic migrations. Another useful introduction is Iain Zaczek's *Chronicles of the Celts.*

3 The Romans first came to Britain in 58 B.C. under Julius Caesar, who invaded Britain twice. Caesar had just emerged victorious over the Gaelic Celt Vercingetorix when he set about conquering Gaul and all other Celtic nations, including Britain. He recorded his campaigns against the various Celtic tribes in *The Conquest of Gaul.* Although it was considered propaganda, the conquest of Gaul provides a unique first-person narrative written by one of the greatest generals in history, as well as an eyewitness account of early Celts. The Roman occupation started in earnest during the reign of the Emperor Claudius.

4 Much has been written about the Roman and Sub-Roman period in Britain. Charles Thomas's *Celtic Britain* and Christopher Snyder's *An Age of Tyrants: Britain and the Britons A.D. 400–600* are particularly good. Other sources for Roman and Sub-Roman times in Britain, including works detailing Romano-Briton culture, include Philip Wilkinson's *What the Romans Did for Us*, Peter Salway's *A History of Roman Britain*, Peter Hunter Blair's *Roman Britain and Early England: 55 B.C.–A.D.871*, and Peter Wilcox's *Rome's Enemies 2: Gallic and British Celts*. Many people believe that the period now known as Sub-Roman was the beginning of the Dark Ages—a terrifying period in which the barbarians laid waste to British culture. However, both literary and archaeological evidence suggests that this period was not as dark as people imagined. The Britons managed to keep their culture intact while incorporating many of the things they learned from the Romans into their lives. In the case of fifth- and sixth-century Britain, one finds a Celtic-speaking population with an elite Latin-writing element, both of which seemed, for the first time, to try to express their ethnic or political identity (Snyder 66).

5 The Irish were the people who settled in the island of Eire or Hibernia, or what is now known as Ireland. Many historians, Bede among them, describe how an Irish tribe—the Scotti—settled in northwestern Britain, above Hadrian's Wall, which the Romans had built to fend off Pictish raids into Britain. These Irish Scotti came to be known as the Dalriada, the name of the part of Britain where they settled. Before the ninth century, two races co-existed in Alba—the Picts and the Irish. It wasn't until the ninth century that the term Scottish came into use to describe the people who populated what would become known as Scotland. Since the Irish Scotti were the dominant tribe, the land was named after their people, and the Picts were subsumed within Irish-Scottish culture (Blair 57).

6 The Picts (the tattooed people of the North) were thought to be of Scythian (Scandinavian) descent. Bede writes:

> At first the only inhabitants of the island were the Britons ... it is said that some Picts from Scythia put to sea in a few longships, and were driven by storms around the coast of Britain, arriving at length on the north coast of Ireland. Here they found the nation of the Irish, from whom they asked permission to settle; but their request was refused ... so the Picts crossed into Britain, and began to settle in the north of the island, since the Britons were in possession of the south (45–46).

7 The group of Germanic invaders, including the Frisians, Angles, Jutes, and Saxons, were collectively known as the Saxons. The Jutes settled in Kent, while the Angles settled in southeastern areas. Bede and other sources claim that the Saxons were led by brothers named Hengist and Horsa, landing in Britain around A.D. 450. While archaeologists and historians insist this date is too late, the story *is* consistent with archaeological evidence. Blair writes that "Bede's racial division of the kingdoms of his own day is probably over-neat. ... [W]hen the kingdoms emerge in the light of day, there is much blurring at the edges" (55). What mattered, Blair writes, is not that the settlers were Angles, Saxons, or Jutes but that they belonged to the same broad culture as Scandinavia, Germany, and northern France. Early poems, jewelry, and customs were similar for these nations.

8 While both Britain and Ireland share Celtic roots, their cultures and languages are different and can be distinguished as either Goidelic or Brythonic. The Brythonic branch encompasses the languages spoken in Wales (Cymraeg or Welsh), Brittany (Breton), and Cornwall (Cornish). Irish, Scots, and Manx Gaelic derived from the Goidelic. While some Latin words were incorporated into Welsh during Roman rule, the languages used by the Cymry, Armoricans, and Dumnonians remained vernacular through the Middle Ages. Old Irish is the language spoken by the Dalriada (see note 5).

9 Geoffrey's *Historia* was considered an important text and a historically accurate document for some time. Over 200 manuscripts remain extant (a staggering number considering the survival rate of medieval manuscripts), but more importantly, the *Historia* was circulated not only in Latin but in French and English, both in England and on the continent. The large number of surviving manuscripts and their translations is proof of its popularity. However, while Geoffrey used Gildas, Bede, and Nennius as his historical sources as well as the *Annales Cambriae*, Welsh king-lists, and genealogies, many of his claims have been largely dismissed by modern scholars as spurious at best and completely fabricated at worst.

10 Four of the eight parts of Geoffrey's *Historia* (Parts 4-7, or §vi.4 through xi.2 in the 12-book version) detail Arthur's lineage and life. Because the *Historia* was considered a historical text for quite some time, the first cohesive narrative of the Arthurian cycle took shape and placed Arthur simultaneously within a historical and a fictional construct.

11 These Welsh or early Briton Arthurian tales vary greatly from the French versions (see next note). The Welsh legends dealt more with a Romano-Briton Arthur who ruled as *Dux Bellorum* or War Duke of Britain before ascending as High King. The Welsh/Briton Arthur was said to have been one of two leaders successful in waging war against the Saxons in the fifth century; the other was

Ambrosius, whose story has since been conflated with Arthur's (Ambrosius is often designated as Arthur's uncle, brother to Uther Pendragon, Arthur's father). French Arthurian legends include stories about Lancelot and the Round Table; neither existed in the Welsh tales, whereas the French have made these two, along with the Arthur-Guinevere-Lancelot love triangle, central to the Arthurian mythos.

12 Wace's *Brut* included Arthurian tales of the Round Table and Sir Gawain. It was Wace's *Brut* which eventually opened the door to the French Arthurian vulgate cycles, which included *Lancelot of the Lake, Sir Gawain and the Green Knight, The Knight of the Cart, The Quest of the Holy Grail,* and *The Death of King Arthur.* Chrétien de Troyes, who was associated with the court of Marie de Champagne (daughter of Eleanor of Aquitaine and Louis VII) in France towards the end of the 12th century, wrote many Arthurian romances under her patronage. His tales focus on courtly love (e.g., the Arthur-Guinevere-Lancelot saga). Sir Thomas Malory, who wrote *Le Morte d'Arthur* three centuries later, was highly influenced by Chrétien's *Arthurian Romances.* In turn, Malory's *Morte d'Arthur* was the model for Alfred, Lord Tennyson's *Idylls of the King.* The best known of these stories is Malory's. However, none of these French vulgate-derived stories is even remotely similar to the Welsh tales. For example, in the vulgate cycle, Lancelot, Galahad, and Gawain are Arthur's prime knights; in the Welsh tales, Cay (Kay in the French versions) and Bedwyr (Bedivere in the vulgate) are his left and right arms. Merlin assumes a supernatural, mystical persona in the French vulgate cycles, whereas Myrddin—the sometime-druid, warrior character in the Welsh tales—was instead teacher and bard in the Welsh legends. Many more of these differences exist, especially regarding Gwynhyfar (Guinevere) and Arthur's companions (knights).

13 The relation between the Isle of Avalon and Glastonbury has been studied etymologically. The Celtic name associated with Avalon (apples) was Ynis Witrin. Ynis Witrin, translated into English, meant Isle of Glass. The Saxon name Glastonbury also meant Isle of Glass. Glastonbury is located near the Severn (Bristol channel) and is close enough in proximity to Wales to give it meaning in an Arthurian context.

14 Proven in the 14th century to be forgeries.

15 Llewellyn is referring to himself as Snowdonia, a locale of Gwynedd, his home base in northern Wales.

16 The royal "he"; again, Llewellyn is referring to himself.

17 Glendower had studied at Westminster and had become a barrister. Afterward, Glendower served Richard II (Holinshed 139).

18 Holinshed reports that after this last attack, in which only Glendower escaped, he ended up in the Welsh mountains alone and starving. Holinshed suggests he may have committed suicide, although there is nothing that suggests this was the case.

19 Henry VII was part Welsh, part English. After the death of Henry V (Prince Hal), his wife, Katharine, married Owen Tudor, a Welshman. Owen Tudor is the grandfather of Henry VII. Henry himself promised that he would release the Welsh from "such miserable servitude as they have piteously long stood in" (Sullivan 15), a reference to the early Britons who had once ruled the island prior to the Anglo-Saxon invasion; the Britons the Normans knew were the descendants of the Britons who had been driven into Cornwall and Wales.

20 The English had a tendency to forget that the Tudors had Welsh blood. As Charles Thomas writes in *Celtic Britain*, "Yet the English overlook the brilliance, and the genesis, of the House of Tudor. Queen Elizabeth I ... was with her red hair, exuberance, linguistic ability, stamina and devious skills, a quintessential Welsh heroine" (Thomas 106).

21 This description of the Welsh as "well-behaved" probably stemmed from the fact that by this time, Wales had been annexed into England by Henry VIII. The union of England and Wales was accomplished by Parliament in 1536 and 1543, wherein:

> [T]he Marcher lordships were shired, English laws and county administration were extended to Wales. ... Wales was made subject to the full operation of royal writs, and to English principles of land tenure. The Act of 1543 dictated that Welsh customs of tenure and inheritance were to be phased out and that English rules were to succeed them. Welsh customs persisted in remoter areas until the seventeenth century and beyond, but English customs soon predominated (Guy 252).

While the English regarded the union as a civilizing vehicle for their Welsh neighbors, the Welsh viewed this as a crude annexation and *not* a union of two countries, since there was no treaty between negotiating parties, as was the case with Scotland. This led to resentment on the part of the Welsh.

22 See above. The part of Henry VIII's edict that offended the Welsh most asserts, "The people of the same dominion have and do daily use a speche nothing like ne consonaunt to the naturall mother tonge used within this Realme" (Lloyd 8). The Acts were designed to create a Welsh ruling class of fluent English speakers. By eradicating the Welsh language, the English hoped to destroy the Welsh identity and to achieve complete assimilation.

23 Elizabeth's nurse, Blanche Parry, taught her how to speak Welsh.

24 Holinshed created an image of the rebellious Welsh leader Owen Glendower as a mystical warrior in *The Chronicles of England, Scotland, and Ireland:*

> [F]or Owen conveyed himself out of the way, into his known lurking places, and (as was thought) through art magic, he caused such foul weather of winds, tempest, rain, snow, and hail to be raised, for the annoyance of the King's army, that the like had not been heard of; in such sort, that the King was constrained to return home (145).

Holinshed portrays Welshwomen as barbarous and uncontrollable by their men:

> This was a shrewd discomfiture to the Welsh by the English, on whome sinister lot lowred, at such a time as more than a thousand of them were slaine in a hot skirmish: and such shameful villanie executed upon the carcasses of the dead men by the Welshwomen … which is worthie to be recorded to the shame of a sex pretending the title of weake vessels, and yet raging with such force of fiercenesse and barbarisme. For the dead bodies of the Englishmen, being above a thousand lieng upon the ground imbrued in their own bloud, was a sight (a man would thinke) greevous to looke upon, and so farre from exciting and stirring up affections of crueltie … yet did the women of Wales cut off their privities, and put one part thereof into the mouthes of everie dead man, in such sort that the cullions hoong downe to their chins; and not so contented, they did cut off their noses and thrust them into their tailes as they laie on the ground mangled and defaced. This was a verie ignominious déed, and a woorsse not committed among the barbarous … and thus much by waie of notifieng the inhumanitie and detestable demeanour of those Welshwomen, after the conflict betwéene the English and the Welsh (271).

This barbarism, performed by Welsh women, only underscored the prevailing notion of the era: that not only were the Welsh wild and unruly, but they allowed their women onto the battlefield and that women—the so-called "weaker" sex—went to their tasks willingly, committing acts worse than those of any other barbaric peoples of the time.

Works Cited

Bede. *Ecclesiastical History of the English People.* Trans. Leo Sherley-Price. London: Penguin Books, 1968.

Berthelot, Anne. *King Arthur and the Knights of the Round Table.* New York: Harry N. Abrams, 1997. 33–43.

Blair, John. "The Anglo-Saxon Period (c. 440-1066)" *Oxford Illustrated History of Britain.* Ed. Kenneth Morgan. Oxford: Oxford University Press, 1984.

Boling, Ronald J. "Anglo-Welsh Relations in Cymbeline" *Shakespeare Quarterly* 51 (2000): 33–66.

Geoffrey of Monmouth. *Historia regum Britanniae.* Trans. Lewis Thorpe. London: Penguin Books, 1966.

Gildas. *The Ruin of Britain and Other Documents.* Trans. Michael Winterbottom. Sussex: Phillimore & Co., 1978.

Gillingham, John. "The Early Middle Ages (1066–1290)" *Oxford Illustrated History of Britain.* Ed. Kenneth Morgan. Oxford: Oxford University Press, 1984.

Griffiths, Ralph A. "The Later Middle Ages (1290–1485)" *Oxford Illustrated History of Britain.* Ed. Kenneth Morgan. Oxford: Oxford University Press, 1984.

Guy, John. "The Tudor Age (1485-1603)" *Oxford Illustrated History of Britain.* Ed. Kenneth Morgan. Oxford: Oxford University Press, 1984.

Holinshed, Raphael. "The Chronicles of England, Scotland, and Ireland." *The First Part of King Henry the Fourth: Texts and Contexts.* Ed. Barbara Hodgdon. Boston: Bedford/St. Martin's, 1997. 139–157.

Lloyd, Megan. "To Speak Welsh: Nonsense and Subversion in Shakespeare's *Henry IV, Part I*" *North American Journal of Welsh Studies 2* (2002): 7–4.

Morris, John. "Historical Introduction." *Gildas: The Ruin of Britain and Other Documents.* By Gildas. Michael Winterbottom, Sussex: Phillimore & Co., 1978. 1–4.

Norwich, John Julius. *Shakespeare's Kings: The Great Plays and the History of England in the Middle Ages: 1337–1485.* New York: Simon & Schuster, 1999.

Schama, Simon. *A History of Britain: At the Edge of the World? 3500 B.C.–1603 A.D.* New York: Hyperion, 2000.

Shakespeare, William. *The First Part of King Henry the Fourth: Texts and Contexts.* Ed. Barbara Hodgdon. Boston: Bedford/St. Martin's, 1997.

Snyder, Christopher A. *An Age of Tyrants: Britain and the Britons, A.D. 400–600.* University Park: Pennsylvania State University Press, 1998.

Sullivan, Garrett A. "Civilizing Wales: Cymbeline, Roads and the Landscapes of Early Modern Britain" *Early Modern Literary Studies* 4 (1998): 31–34.

Thomas, Charles. *Celtic Britain.* London: Thames & Hudson, 1997.

Williams, Brenda. *Saxons and Vikings.* Hampshire: Pitkin Unichrome, 2001.

Spectacles of Justice: Failure and Success of Punishment in *The Merchant of Venice* and *Richard II*

Courtney E. Rydel,
The College of New Jersey

Faculty Sponsor:
Professor Jo Carney,
Department of English

Abstract

William Shakespeare's plays *The Merchant of Venice* and *Richard II* feature pivotal scenes of public trial and judgment. These spectacles replicate many of the ways that justice operated in early modern Europe. Linking information about the contemporary judicial system and theories of punishment with the plays themselves, this essay illuminates how Shakespeare's characters manipulate or succumb to the power invested in spectacles of justice.

Introduction

Trial by ordeal once combined elements of corporal punishment, judgment of guilt or innocence, and execution. The accused underwent torture, such as plunging a hand into molten metal. If he or she was able to withstand injury or endure the pain, it was taken as proof of innocence, since God supposedly intervened to save the blameless. Although these aspects of corporeal torture, execution, and trial were gradually separated as European society moved from the medieval into the early modern period, they still retained connections. For the purposes of this paper, I am going to consider scenes of trial, torture, and execution before an audience as "spectacles of justice." Shakespeare incorporated both successful and aborted spectacles of justice into his plays *The Merchant of Venice* and *Richard II*. The fates of Shylock and Antonio in *The Merchant of Venice* and *Richard II* and Henry Bolingbroke in *Richard II* rest upon how they manipulate power within spectacles of justice.

In his seminal book, *Discipline and Punish: The Birth of the Prison,* Michel Foucault concentrates on the interplay of power among the participants in public executions and scenes of torture. Specifically, Foucault and scholars building off his work examine the roles of the audience, the body of the accused, and the importance of the monarch. All of these elements play into the failure or success of Shakespeare's spectacles of justice.

Playing to the Audience

For the spectacles of justice in the plays, I consider the audiences to consist of two groups: the audience on stage that exists within the fiction of the play and the audience of viewers or listeners outside the play. Both are important. As Frances E. Dolan notes, "Recent scholarship has drawn attention to the theatricality of public executions and to the relationships between the scaffolds and the stage in early modern Europe, emphasizing displays of power on both" (157). The audience was crucial in public executions and trials. Esther Cohen observes: "Implicit in the actual execution of justice was the perception of law as a vehicle for human relations. ... It existed and was used for the purpose of preserving the social order" (Cohen 287). Obviously, no message could be sent about human relations without an audience to witness the spectacles of justice.

As Foucault argues in *Discipline and Punish,* it is only in modern times that justice becomes focused on preventing crimes before they occur and redeeming the soul of the criminal by means of systems of surveillance

and control. Originally, he says, "The public execution is to be understood not only as a judicial, but also as a public ritual. It belongs, even in minor cases, to the ceremonies by which power is manifested" (47). "In short this power is exercised rather than possessed: it is not the privilege, acquired or preserved, of the dominant class, but the overall effect of its strategic positions—an effect that is manifested and sometimes extended by the position of those who are dominated" (26-27). Therefore, spectacles of justice are, at heart, ceremonies manifesting and creating power, through the interaction between those who are currently dominant and those who are dominated.

Thomas Laqueur, however, emphasizes the audience's agency. He argues that "The crowd, and particularly the carnivalesque crowd, was the central actor in English executions" (309). Laqueur records numerous telling anecdotes. At an execution for treason in 1785, seats went for two shillings; when the convicted was at the last minute pardoned, riots often broke out among disappointed spectators; contemporary diarists record lively interest, not revulsion, at public hangings (309-322). John Laurence in his *A History of Capital Punishment,* describes the audience at a typical early modern English execution:

> While the public execution was a source of terror, it was also a call to carnival, less an object lesson than a festival. 'The high and titled frequently booked windows overlooking the scaffold' and feasted their friends with 'elaborate breakfasts,' while 'the low gathered in their hundreds to witness a hanging, and the public-houses in the neighborhood often kept open the whole of the preceding night for their refreshment and enjoyment'. ... Beggars, harlots, and pickpockets worked the crowd. (213)

Sometimes this carnivalesque audience disrupted and subverted the spectacle of justice. Tactics included "Preventing an execution that was regarded as unjust, snatching a condemned man from the hands of the executioner, obtaining his pardon by force, and assaulting the executioners" (Foucault 59). Similarly, the audiences within the plays, namely the Venetian Christians and English nobility, possess power to legitimize or disrupt spectacles of justice.

Shylock and Richard fail to connect with their audiences onstage, even before their spectacles of justice begin. They take such pride in their own identities of Jew and king that they open a chasm between themselves and their audiences. Shylock has self-identified as a Jew repeatedly. He is proud to be a Jew and not a Christian, such as when he scorns Christian masques (2.5). As Mary Jane Metzger notes, for an early modern English audience "living and working 'honestly and unobtrusively' meant becoming invisible as 'former' Jews and convincingly performing the prerequisites for integration into English society" (54). Shylock does not keep a low profile or attempt to assimilate. Shylock is a highly visible, malevolent other who anticipates taking Antonio's life.

Richard, too, has estranged himself from his nobles. According to York, Richard foolishly listens to Bushy and Green, his associates who insinuate "Lascivious meters to whose venom sound/The open ear of youth doth always listen" (2.1.19-20). Richard makes rash decisions, such as halting the duel between Mowbray and Bolingbroke at the last moment and exiling the men, in front of an audience ready for action. Two plays later, in *Henry IV Part Two,* Mowbray's son says that when Richard ended the duel, "Then he threw down himself" (4.1.125). Richard remains oblivious to his problem because he believes "The breath of worldly men cannot depose/The deputy elected by the Lord" (3.2.52-3). When he seizes Henry Bolingbroke's estate while Henry is in exile, Richard again displays disregard for his audience. Before seizing the estate, Richard has told Bushy and Green that "[Henry's] coffers shall make the coats/To deck our soldiers for these Irish wars" (1.4.60-1). Yet Richard pretends to his onstage audience that he takes the estate because Henry is a traitor. He dismisses his audience: "Think what you will" (2.1.210); and the dissatisfied nobles do.

By contrast, Antonio and Henry Bolingbroke play to their onstage audience, Antonio perhaps with less self-awareness than Henry. Their identities as Venetian Christian and English nobles give them wider appeal than Shylock and Richard, and their earlier roles as wrongfully accused men in spectacles of justice earn them sympathy. Antonio's power over Shylock is accorded to him by the joint actions of Portia and the Duke, with the support (presumably) of the Christian community onstage. Henry Bolingbroke, however, personally earns his power in the eyes of his audience. Richard comments on how Henry cultivates his popularity: "Observed his courtship to the common people/How he did seem to dive into their hearts/With humble and familiar courtesy" (1.4.23-5). When Henry returns, he solidifies his position by enacting his own successful spectacle of justice against Bushy and Green.

Body Images

According to Foucault, "The disappearance of public executions marks therefore the decline of the spectacle; but it also marks a slackening of the hold on the body" (10). Yet in these spectacles of justice under consideration, there is still a firm hold on the body and a careful examination of the political power invested in the body. Foucault observes that "the body is also directly involved in a political field; power relations have an immediate hold on it" (25). Torture was, as Foucault dryly observes, "cruel but not savage" (40). Rules governed the gradations of torture. In the classical period in France, "The first degree of torture was the sight of the instruments. In the case of children or of persons above the age of 70, one did not go beyond this stage" (Foucault 40). So when Shylock whets his knife in plain view, he is torturing Antonio before even laying a hand on him (4.1.120). Antonio's body is the crux of this scene, as Shylock knows. He puts Antonio's pound of flesh at stake in the bond (1.3.140-147).

Yet, by putting Antonio's body in the spotlight, Shylock makes a crucial misstep. The Venetian Christians do not want Shylock to have the power involved in punishing Antonio's body. Just as early modern society often let minor offenses off with mere fines,[1] the Venetian Christians seek to remove Antonio's body from the spectacle of justice. Repeatedly, Bassanio offers Shylock 6,000 ducats, twice the contested sum, and Shylock refuses (4.1.83-6, 4.1.203-9). Shylock concentrates on punishing Antonio's body, a preoccupation often associated with older conceptions of justice. As Foucault observes, "[W]ith feudalism, at a time when money and production were still at an early stage of development, we find a sudden increase in corporal punishment—the body being in most cases the only property accessible" (25). Yet unlike a peasant, Antonio does not have only his body as "the only property accessible." Shylock closes off other options in his focus on Antonio's body. Ultimately, Antonio's spectacle of justice succeeds because it stops concentrating on the body. Instead, Antonio, the Duke, and Portia focus on the modern conception of punishment as a means of remediating criminals. Shylock is forced to convert to Christianity, ostensibly saving his soul, and loses half his estate as a fine (4.1.375-385).

On the other hand, Richard's problem is that he does not understand the necessary link between ritual spectacle and violence. While written in the 1590s, *Richard II* is set in the late Middle Ages, so Foucault's observations on medieval justice are pertinent: "To judge was to establish the truth of crime, it was to determine its author, and to apply a legal punishment" (19). Richard violates these principles. Truth is concealed, not established. Richard, who ordered the murder of Gloucester, is not legally punished; assassin Mowbray and innocent Bolingbroke are both exiled. Conversely, Richard divorces from bloodshed and corporal punishment the real spectacles of justice: Mowbray and Bolingbroke's appeal to Richard in 1.1 and the attempted trial by duel between them in 1.3. Leonard Barkan argues that Richard's interruption of the duel "reduces [the trial by combat] entirely to its ritual aspects

and deprives it of its natural, and bloody, resolution" (8). Gloucester's murder, the only violence Richard presides over, is separate from even a sham display of royal justice.

Richard's ineffectuality can be contrasted with Henry Bolingbroke's impressive dispatch and authority. Henry's effective handling of royal justice must come as a positive relief to the implicit audience—the English nobility of 1399 and the London theatergoers of the 1590s. As Barkan notes, while Richard is "suddenly taking his deposition for granted" without any actual crisis forcing him to it, "Bolingbroke is executing villainous Plantagenet allies" (10). Finally we have Henry Bolingbroke, a man who can say, "This and much more, much more than twice all this/Condemns you to the death.—See them delivered over/To execution and the hand of death" (3.2.28-30). Henry does not have a formal, modern trial for Bushy and Green, but he announces: "Yet to wash your blood/From off my hands, here in the view of men/I will unfold some causes of your deaths" (3.1.5-7). Henry's reasons are at least in part true, as the audiences—both fictional and real—know that Bushy and Green have been complicit in Richard's unjust seizure of Henry's estate. Two men are judged and sent to their execution in a scene consisting of 44 lines. In contrast, Richard expended hundreds of lines on the murder of Gloucester, but accomplished nothing. Foucault notes that "the right to punish, therefore, is an aspect of the sovereign's right to make war on his enemies" (48). The play illustrates this truth as Richard II fails to put down the Irish rebellion and Henry Bolingbroke stages his own rebellion and later puts down uprisings against himself.[2] Richard separates speech from violent action, whereas Henry connects them.

Heads of State

Spectacles of justice emphasize the effect of crime on the community. The sovereign both embodied the community and exerted the most power within the community, a power dynamic early modern monarchs exploited. Cunningham observes:

> To solidify their position, stabilize their power, and universalize an image of themselves, Tudor monarchs tried to organize unambiguous, 'divinely sanctioned' public punishments by controlling the interpretive play of trials and executions; in these events ... God (not the monarch) saw the truth, brought it to light, preserved the innocent, and punished the guilty. (209)

Foucault goes so far as to argue that "every crime constituted as it were a rebellion against the law and that the criminal was an enemy of the prince" (50). The sovereign *was* the law and the community and in large part dictated the script of spectacles of justice.[3]

Indeed there was a script, even if the audience had the carnivalesque power to disrupt it. Cunningham notes, "The textual self-consciousness appeared in trial records ... at every turn, words were recorded, commented on, and policed" (210-211). According to Dolan, "By the late 17th century, a voluble and assured performance by the condemned had become so standard, so expected," that one contemporary record of a woman's hanging "apologizes for her failure" to deliver a speech (169). To the watching crowd's disappointment, she did not "do anything on the scaffold but die" (Dolan 169). There were several roles available to the convicted in their final performances, such as confessing their crimes and repenting, proclaiming their innocence, or acting carefree and jocular on their way to their death.[4]

In *The Merchant of Venice,* Antonio initially is in the position of the convicted, and he welcomes death. Antonio proclaims, "I am a tainted wether of the flock/Meetest for death" (4.1.113-4). Yet the community and head of state, the Duke of Venice, fight for his life. Antonio says to the Duke, "I have heard/Your grace hath ta'en great pains to qualify/His rigorous course" (4.1.5-7). The Duke presses Shylock to be merciful and concludes his speech with: "We all expect a gentle answer, Jew" (4.1.33). Unable to disprove Shylock himself, the Duke awaits the help of "a learned doctor/Whom I have sent for to determine this" (4.1.104-6). Portia, the

legal consultant, first finds the loophole that saves Antonio and then discovers the law that allows Christians to reverse the situation and put Shylock on trial. Portia warns: "If it be proved against an alien/That by direct or indirect attempts/He seek the life of any citizen" half his goods and his life are forfeit (4.1.344-6). Antonio's person becomes sacred, symbolizing the entire community as the body of the sovereign does. Shylock is "other," and therefore an assassin. Shylock is the one requesting death now, but his plea is no more successful than Antonio's. The same power stands behind both decisions: the Duke of Venice.

Only in the final scene, in which he renounces his crown, does Richard's speech influence the process of spectacular justice. Earlier, Richard was unable to control through speech spectacles of justice over which he presided as king.[5] Now, the House of Commons has requested that Richard have judgment passed on him.[6] Northumberland orders him to read a paper listing "These accusations and these grievous crimes/Committed by your person and your followers" (4.1.213-4) so that "the souls of men/May deem that you are worthily deposed" (4.1.216-7). Richard refuses. Faced with Richard's obduracy, Henry stops Northumberland, who worries that "The Commons will not then be satisfied" (4.1.262). Richard dismisses this worry, saying, "They shall be satisfied" (4.1.263). Characteristically, Richard ignores his audience and speaks one of his most powerful soliloquies to his own reflection in a mirror.

Furthermore, Richard powerfully calls attention to his role as king,as representative of God's justice and the complete opposite of a condemned criminal. After being deposed, Richard says: "I find myself a traitor with the rest/For I have given here my soul's consent/T'undeck the pompous body of a king" (4.1.238-40). For him, regicide and suicide are the same act. He repeatedly refers to his divinely ordained kingship even as he is abdicating: "With mine own tongue [I] deny my sacred state" (4.1.199). Richard repeatedly upbraids the nobles as "traitors." By refusing to play his scripted role in the spectacle of justice Henry Bolingbroke presides over, Richard deprives the spectacle of some of its legitimacy.

Overall, Henry appears to control this spectacle of justice effectively; he allows Richard to make speeches, but Henry takes the crown for himself. Foucault links public execution with "a whole series of great rituals in which power is eclipsed and restored (coronation, entry of the king into a conquered city, the submission of rebellious subjects)" (48). Immediately after Richard is escorted to the tower, Henry announces: "On Wednesday next we solemnly set down/Our coronation" (4.1.309-10). The judgment of Richard is interwoven with the coronation of Henry. When Richard asks for a mirror, Henry repeats the request as a command (4.1.258). Later Richard asks to leave and Henry orders, "Go some of you, convey him to the Tower" (4.1.306). Henry takes every opportunity to display his authority.

The law for traitors provides insight into how high the stakes were for Bolingbroke and nearly every other male in *Richard II* since they were all considered traitors to one party or another. In 16th-century England the standard sentence for treason was:

> That the convict be drawn from the Tower on a hurdle through the city of London to the place of execution, 'there to be hanged till he should be half dead; that then he should be cut down alive, his privy parts cut off, his belly ripped, his bowels burnt, his four quarters set up over the four gates of the city, and his head upon London bridge' (Cunningham 220)

The purpose of the hurdle was to ensure that the convict, while being dragged behind a horse, did not smash his head on the street and thereby die relatively quickly and painlessly.[7] Hanging until "half-dead" also seems odd to a modern reader, until one realizes the difficulty of hanging someone expeditiously. In many village hangings, where the condemned swung from a tree after the cart on which he or she stood pulled away, the

friends of the convicted person would rush forward and tug on his or her feet to try to hasten death. According to Dolan, Henry VIII considered burning Anne Boleyn at the stake before relenting and having her merely beheaded for treason (164).

Since treason was the ultimate crime, on some level Henry Bolingbroke and his audiences must be disturbed by his usurpation. Therefore, prior to deposing Richard, Henry conducts a spectacle of justice to ameliorate his own guilt. Punishing Bushy and Green, two of Richard's unpopular advisors, allows Henry to displace his own guilt and his own treason onto others. Henry enacts a spectacle of justice which implicitly clears his own name, blaming Bushy and Green for making Richard "misinterpret me" (3.1.18) and thereby causing his exile. In listing their crimes, Henry accuses them of having "misled a prince, a royal king" (3.1.8). Richard's mistakes are not his own (for the moment, until it becomes convenient for Henry to accuse him of crimes worthy of deposition). Bushy and Green become the *real* traitors. Whereas Richard displaces responsibility onto Mowbray and solves nothing, Henry's shifting of responsibility allows him to gain power and resolve the issue of his disinheritance.

All's Well That Ends Well ... For Whom?

Regardless of Shylock's bloodlust and Antonio's anti-Semitism, Shylock had a seemingly legal case to take Antonio's life. Henry Bolingbroke, on the other hand, had no legal right to execute Bushy and Green or seize the throne. However, Shylock and Richard produce failed spectacles of justice, paving the way for Antonio and Henry to come out on top. As Esther Cohen suggests, the audience of a spectacle of justice is "like a magician's audience" and sees "exactly what they were meant to see and only that: the majesty of the law in its full efficacy" (300). When reading or watching Shakespeare's plays, we too join the audience.

WORKS CITED

Barkan, Leonard. "The Theatrical Consistency of Richard II." *Shakespeare Quarterly* 29.1 (1978): 5–19.

Cohen, Esther. "'To Die a Criminal for the Public Good': The Execution Ritual in Medieval Paris." *Law, Custom, and the Social Fabric in Medieval Europe: Essays in Honor of Bruce Lyon.* Ed. Bernard S. Bachrach and David Nicholas. Kalamazoo: 1990. 285–304.

Cunningham, Karen. "Renaissance Execution and Marlovian Elocution: The Drama of Death." *PMLA*, Vol. 105, No. 2. March 1990. pp. 209–22.

Dolan, Frances E. "'Gentlemen, I Have One More Thing to Say': Women on Scaffolds in England, 1563-1680." *Modern Philology* 92.2 (1994): 157–78.

Foucault, Michel. *Discipline and Punish: The Birth of the Prison.* Trans. Alan Sheridan. New York: Vintage Books, 1995. 3–69.

French, A. L. "Richard II and the Woodstock Murder." *Shakespeare Quarterly* 22.4 (1971): 337–344.

Laqueur, Thomas W. "Crowds, Carnivals and the State in English Executions, 1604–1868." *The First Modern Society: Essays in English History in Honor of Lawrence Stone.* Ed. A. L. Beier, David Cannadine and James M. Rosenheim. Cambridge: 1989. 305–55.

Metzger, Mary Janell. "Now, by My Hood, a Gentile and No Jew: Jessica, *The Merchant of Venice*, and the Discourse of Early Modern English Identity." PMLA 113.1 (1998): 52–63.

Shakespeare, William. *Richard II* and *The Merchant of Venice. The Norton Shakespeare.*

NOTES

1 "It is only when one digs below the level of the spectacular, reaching the mundane daily routine of justice, that one discovers how small a proportion of the offenders actually ascended the gallows." Cohen 299.

2 This consideration illuminates a scene in *Henry V*, 3.4. When *Henry V* hangs his former associate Bardolph for stealing, it seems harsh. Yet Henry V has to be able to enact spectacles of justice if he is fully to inhabit the role of the warrior-prince.

3 "The protagonists' lines were usually written for them." Cohen 289.

4 Dolan, Foucault, Laqueur. This observation is common to all three scholars' arguments.

5 A. L. French suggests the play hinges upon a speech Richard does not control: Gaunt's deathbed accusation that Richard killed Gloucester. In this reading, Richard's seizure of the estate becomes postmortem revenge.

6 Footnote, *Richard II* 4.1.145.

7 Cunningham 220.

The Effects of Hand-Raising on Student Speech

Christina Maffa,
The College of New Jersey

Faculty Sponsor:
Professor Felicia Jean Steele,
Department of English

Abstract

This study examines the effect of hand-raising on a student's speech register in classroom discussion. Based on observations at a middle school site, I have noticed that when students raise their hands preceding their comments in class, their register becomes more formal than the register they use when talking to their peers outside the classroom. If students are responding to comments or questions without raising their hands, their register remains at the level they would use with their peers. This change in register when students raise their hands signals an effort (whether conscious or unconscious) to use a more standardized English; students are therefore perceiving that their audience has changed from the familiarity of their peers to someone who is less familiar (the teacher or the class as a whole, to whom they are not as close). The use of Standard American English when a student's hand has been raised also corresponds to the nature of the response as being more objective, whereas when they speak without their hands raised, the response is often more subjective and emotionally based. I would argue, based on these observations, that hand-raising limits the extent of student response in a classroom because students feel as if they have to impress a certain audience as shown through the use of a more formal English. There is also an increased sense of community and interactional level achieved when students are contributing to a conversation but not raising their hands as suggested by the informality of their speech. As a result, teachers should explore alternative methods of facilitating classroom discussion as opposed to hand-raising.

Introduction

For this study, I conducted my observations at Community Middle School in the West Windsor-Plainsboro school district in an English class of seventh-graders. The classes that I concentrated my studies on were the second- and third-period classes of a first-year teacher. The average class size was 22; the period-two class consisted of nine boys and five girls; the period-three class of 19 boys and six girls. According to the teacher, the school has a mix of lower-middle-class and upper-middle-class students. Compared to the other middle school in the West Windsor school district, Community Middle School has a lower economic status. The students in both classes had been allowed to choose their own seats for this unit so the students were arranged mostly in gender-based clusters. For example, in the period-two class, a row of seven of the nine boys sat in the back of the classroom. Both classes were non-tracked, meaning that there were students of different ability levels in the class, including inclusion students.

The size of the classroom was fairly small, with the desks closer together than most classrooms and also more closely grouped near the teacher. The class, for the period, was engaged in a discussion on the book *The Giver* by Lois Lowry. In order to recreate the

atmosphere of the book, the teacher had taken down all the signs and posters around the room so that the walls were bare. The desks were also arranged in rows unlike their normal configuration, a double-horseshoe shape. The teacher remained in the front of the classroom, either walking back and forth in the front, or sitting on a desk in the front while talking to the students. The class period lasted an hour.

When students first walked into the classroom, they had a "do now" task written on the board. The students, for the most part, sat in their seats and began working on the question; however, they also brought in their conversations from outside the classroom that they continued while writing. These conversations, from what I could hear, were all conducted in an informal register because the students were speaking with their friends as they walked in. Most of them, since they were able to choose their seats, were also presumably sitting next to their friends so the conversations with those sitting next to them also remained in an informal register. This informal register was characterized by referring to each other as "dude," the common occurrence of the word "like," shortening "because" to "cuz," using "gonna," and other vernacular forms. While they were writing their answers to the "do now," the teacher told the students to quiet down once in the first class and twice in the second class, and they eventually complied.

After the "do now," the teacher and students engaged in a discussion about the "do now" and then extended that discussion to other aspects of the book. This discussion continued for about 30 minutes, taking up a good portion of the class period. Since the observation occurred in April, the students and teacher had already established rules for classroom discourse, and the rules focused on hand-raising. If a student had something to contribute to classroom discussion, he/she would raise his/her hand. The teacher then signaled to the students that she wanted them to speak by either pointing to them or calling their names.

Once students were called on after they had raised their hands, they often formalized their speech, changing from the register they were using when talking with their friends at the beginning of class. For example, a boy who had previously whispered to his friend, "Whatcha gonna write?" stated in his response to the teacher (referring to the characters in the book), "They were going to die." Also the word "because" remained as "because" instead of becoming "cuz" four out of the six times it was used after students raised their hands to speak. The word "like" still prevailed in the students' speech, but the filler word "um" was also now common. After students responded to a question posed by the teacher, she often replied with her own comments on what they said. The teacher's speech, unlike the students', remained in an informal register. She referred to the students as "you guys," described something one of the students said as "kinda cool" and also used "like" with great frequency in her speech.

Not all the classroom discussion, however, was conducted through hand-raising. At several points during the discussion, the class abandoned raising their hands and began calling out responses to each other's statements. The teacher allowed for these moments, listened to what the students had to say, and gave her own comments as well. After she took control of the topic and moved along to the next question or point, the class resumed their hand-raising. The students' speech at these moments when they were not raising their hands reflected the informal register they used when speaking to their friends at the beginning of class, even though they were engaged in a classroom discussion. For example, at the beginning of the discussion, the teacher asked what the students thought about the end of the book. The discussion started with students raising their hands. One student explained, "I was confused because it said there was music coming from somewhere so that might have been the community but it could've been Jonas imagining things." She used the word "because" and did not contract

"might have." As the discussion continued, some students started to respond to what their peers and the teacher were saying without raising their hands to have the teacher call on them. For example one student said in response to another without raising his hand, "No, it stinks cuz they might've died." The two words that were not shortened when one student raised her hand—"because" and "might have"—were shortened when another student called out his response. Another student when not raising his/her hand said, "Dude, that would be cool if..." This comment responded to something the teacher said. The informal register that the students used with each other at the beginning of class was now being used with the teacher to discuss the novel. The students who called out their responses were either those in close physical proximity to whomever spoke last, those sitting close to the teacher, or boys (who in both classes constituted the majority). In the third-period class, where the boys held an overwhelming majority, only two girls spoke at all.

After the class discussion of the novel, the teacher asked the students to write an ending to *The Giver* extending beyond Lois Lowry's. Students were given about 10 minutes to write a response. While they were writing, they remained, for the most part, silent because the teacher stressed that they should be working individually while they were writing. Some students, mainly the clusters of boys or the clusters of girls, would occasionally start talking to each other about what they were writing. These conversations were conducted in the more informal register again. For example, one boy said to another boy sitting next to him, "Get this, instead of ridin' down the sled, I'm gonna make them dissolve into acid." "Get this" was an informal way to attract his friend's attention. The student then replaced the velar nasal present in "riding" with the alveolar nasal present in "ridin'" which also is a characteristic of more informal dialect (Keisling 1). Once the discussion in the class became louder, signaling that most of the students were finished writing, the teacher got their attention by saying, "All right, who wants to read?" Five students' hands went up, but one boy added speech to his gesture of hand-raising—"Can I go first? Can I go first?" The teacher nodded at him, but the other students were still speaking to each other, so the student signaled that he was going to start by saying loudly, "All right, I'm gonna read." Even though he raised his hand, he still used "gonna" because he was still addressing everyone in the class informally as if they were his friends and he wanted them to be quiet.

When reading the endings of their stories, certain gender patterns emerged. Most of the boys started by reading what they had written. Then, they began to summarize rather than read directly from their sheets of paper anymore. All the girls read directly from the paper at all times. The boys' speech, therefore, was a little more informal than the girls'.

Based on the observation of the classes for these two periods, I believe that when students address only the teacher, they raise their hands and their register becomes more formal. When they address each other, either in a conversation that does not relate to class or one that does but they are only calling out a response, they use a more informal register. If the discussion continues in class, without the teacher enforcing hand-raising, the teacher soon becomes part of the informal conversation developing, and the students do not address her in the formal register anymore either. The informal register suggests that the students have achieved a greater level of comfort when expressing their opinions.

While the responses with the students' hands raised were intelligent and well considered, the discussions without hand-raising involved more students participating. In the discussion with the more informal register, students also seemed to be more emotionally involved, and this emotion implies a certain sense of subjectivity; they brought their own ideas and experiences to

the discussion. The discussion with no hand-raising was also much more animated because of the increased involvement of members of the class and the interest of students in what their classmates were saying.

Courtney Cadzen describes the mode of classroom discussion facilitated through hand-raising as "teacher nomination" and the discussion without teacher direction as "student self-selection" (82). Cadzen suggests that having students raise their hands to participate is not the only way discussion can be run in the classroom, but that student self-selection can also work well. Student self-selection provides a more student-centered approach to the teaching taking place in the classroom because it allows students to make the classroom discussion more relevant to them. Cadzen explains:

> Teachers may decide during some activities not to exercise their power to select student speakers. Instead of preallocation of turns by the teacher, there is then more local management of turn-taking by individual students at the moment of speaking. With this shift, classroom talk becomes more like informal conversation—not the same as conversation, because there is still the large group of potential speakers and the educational necessity to stick to an agenda, but closer to it. (82-3)

This idea of an informal conversation that occurs when the discussion is not controlled by the teacher is reflected in the observations at Community Middle School. The students sensed this informality because their discussion was not being directed by the teacher, and it was reflected in their informal register when they were not raising their hands. Giving students the ability to decide when they will speak allows them to feel more a part of the classroom environment. There is also less pressure to change their speech to a more formal Standard American English as shown through their register.

The more formal register used when hand-raising indicates a lack of familiarity with the other members of the classroom. Ithel Jones argues that this familiarity/friendship with classmates and the community it creates in the classroom is more conducive to a successful learning environment (70). The classroom becomes more of a place where students are learning from each other, instead of a place where students' voices only serve to respond to the teacher's questions. In the study she conducted, she compared students' language usage when discussing literature in the class with friends to their language when discussing literature with a non-friend. While the study was done with elementary school children, its results can also be applied to a middle school setting. She remarks "In general, the findings suggest that different aspects of social context, in this case peer relationships, differentially affect children's oral language use. This seems especially true for children's use of literate language, thus supporting the proposition that not all peer contexts are equally facilitative of literacy learning" (Jones 70). The students in the elementary school engaged in discussions with more "literate language" when they were speaking with their friends as opposed to non-friends. Jones defines literate language as "a specific oral register that shares many design features with the language used by teachers during reading and writing lessons, as well as the language used in early reading" (Jones 64). Therefore, students were engaged in discussions that were more beneficial to their understanding of a text if they were talking to friends. Jones observes, "A relationship [between students] implies the existence of an interactional history between individuals that influences both present and future interactions ... moreover, recent research suggests that this type of interaction is particularly important when children learn to use literate language" (64). Obviously, while all members of the classroom cannot be close friends, a more familiar relationship can be developed if students are allowed to speak more informally. Jones also explains that "close relationships originate as a result of repeated exposure or propinquity. Consequently, over time, repeated interactions of non-friends and

acquaintances could become more like the interactions of close friends" (Jones 65). If the teacher works to create a classroom community where students are not afraid to speak and feel comfortable in front of their peers, the language usage might also reflect that. The important aspect of the self-selected approach is that students are not only responding to the teacher but to every member of the classroom. Cadzen recognizes this:

> In a community of learners, students have to listen to and learn from each other as well as the teacher. That's the only way for them to learn during the time spent solving problems in a group rather than just working alone at more traditional seat work. Beyond careful listening herself, the teacher's responsibility is to help peer listening happen. (89)

While the self-selected approach allows students to learn from each other, it also gives more responsibility to the teacher to make sure everyone is respected and treated fairly.

This self-selected approach can also lead to inequalities that do not normally occur when the teacher controls the discussion. Some students may dominate the discussion and some may never speak at all. Some students may feel uncomfortable with the perceived lack of structure in this kind of discussion. The class, cultural, and gender factors that affect these styles all have to be taken into account. For example, at Community Middle School, when student self-selection took place, the discussion was much more dominated by boys than girls. The control of the discussion by males could have resulted from their majority in the classroom, but it might also have to do with females being hesitant to speak out of turn. Research also affirms the different conversational styles of boys and girls in the classroom. In Jones's study the data suggest that there could be different antecedents of metacognitive and literate language for male and female friendship groups. This proposition is consistent with previous studies suggesting that differing interactional styles of boys and girls, such as boys being more conflictual than girls, is a significant factor (71).

Cadzen presents several possibilities teachers have used in their classroom to overcome the inequalities that might present themselves in a student self-selected discussion. One teacher had an observer in the class create a bar graph recording how many times each student in the class participated in a student self-selected classroom discussion. The teacher then presented the graph to the class without the students' names on the graph. The class then discussed why certain bars were lower than others and what could be done to make it easier for these students to participate. Another teacher videotaped a classroom discussion and then showed the class the videotape. After watching, the class reflected on the videotape with regard to how they could make the conversation in the class more equitable.

In a study reported in the *Journal of Applied Behavioral Analysis,* three researchers compared the use of hand-raising to facilitate classroom discussion to the use of response cards on which students could write their answers to the teacher's questions. While response cards might not work in an English classroom where responses tend to be longer and more complex, the study did reveal some important information about a discussion that uses hand-raising:

> Most of the students preferred response cards over hand-raising, stating that they were more fun to use. Students also felt that they learned more when response cards were used. In fact, during hand-raising, some students appeared frustrated at times when they were not called on; some students stopped raising their hands at all, others put their heads down on their desks, and still others complained about not being called on. (70)

This frustration also occurred frequently in the class that I observed. If students were raising their hands and not being called on, they would either wiggle around in their seat and keep shifting position or put their hands down. Janani Narayan, who first conducted this study, explains that one benefit

of hand-raising over response cards is that "it could be argued that the hand-raising procedure entails a built-in bias for higher response accuracy. Students are more likely to raise their hands for questions that they believe they can answer, whereas with response cards, students are expected to give an answer to every question" (485). In an English classroom, however, teachers may not be as concerned with accuracy as with students sharing their ideas and experiences and connecting them to the text.

Not only did students in the study like using response cards more, but using the response cards also resulted in increased student scores on tests and quizzes covering the material in the discussion. Because more students can respond, "response cards allow the teacher to teach facts and definitions more effectively in conjunction with hands on experiences in which students apply those concepts" (Howard 69). In student self-selection, more students can also respond as shown by my observation at Community Middle School, and "learning is enhanced when the frequency with which students actively respond is increased" (Howard 63). Whether it is with response cards or through student-self selection, having more students participate is beneficial to their academic success.

The shift in register between the two discussion styles in the classroom reveals how each discussion style reflects the student's perception of the classroom environment. With student self-selection, students are able to see themselves as community members learning from each other, and their informal register reflects this. Changing the discussion style of a classroom can be very difficult. As Cadzen observes, "Teachers who are trying to make self-selection work with older students often find that well-learned habits, especially those literally embodied, as in hand-raising, are hard to set aside" (Cadzen 84). However, a student self-selected approach is worth using if it will create an environment where more students are involved and responding to each other, not just to the teacher. Allowing students to be more comfortable with their relationship to other members of the classroom through their language register can help them become more comfortable with making their voices heard and creating a classroom where all students' voices are central to the discussion.

BIBLIOGRAPHY

Cazden, Courtney B. *Classroom Discourse: The Language of Teaching and Learning.* Heinemann: Portsmouth, 2001.

Jones, Ithel. "Social Relationships, Peer Collaboration and Children's Oral Language." *Educational Psychology.* Vol. 22, No.1, 2002.

Howard, William L., Grossi, Teresa A., Gardener, Ralph III. "Effects of Response Cards on Student Participation and Academic Achievement: A Systematic Replication with Inner City Students During Whole Class Science Instruction." *Journal of Applied Behavioral Analysis.* Vol. 27, Spring 1994, 63–71.

Kiesling, Scott. "Men's Identities and Sociolinguistic Variation: The Case of Fraternity Men." *Journal of Sociolinguistics.* Vol. 2, Issue 1, February 1998.

Narayan, Janani S., Gardener, Ralph III, Howard, William L., "Using Response Cards to Increase Student Participation in an Elementary Classroom." *Journal of Applied Behavioral Analysis.* Vol. 23, Winter 1990, 483–490.

"A Woman's Party"[1]: Race, Class, and Gender Intersections in the Black Panther Party

Matthew Richman,
The College of New Jersey

Faculty Sponsor:
Professor Ann Marie Nicolosi,
Department of Women's and Gender Studies

Abstract

In a 1993 opinion piece in *The New York Times*, the writer Alice Walker responded to a new retrospective book by Black Panther David Hilliard.[2] Walker excoriated Hilliard, and the party by extension, for rampant misogyny and sexism. She claimed that men such as Huey Newton and David Hilliard embraced the idea of the Black Panther party after being ridiculed as "punks," both as individuals and collectively as black men. In the process of defaming Panther men as malevolent patriarchs, Walker dismissed the numerous women in the party as inconsequential and invisible. The paradigm of analysis for the Black Panthers that Walker establishes reflects the prevailing historical discourse surrounding the party. Despite the sometimes willful ignorance of women's experiences and voices in the Black Panthers, their participation—born out of particular intersections and interconnections of race, class, and gender—constituted perhaps the most vital force in the formation of the party and the continued propagation of its values.

Introduction

The Black Panther party, both in its ideological positions and political practice, was considerably more complex than Walker and its other detractors admit—largely as a result of layered experiences of race, class, and gender. Former Black Panther Safiya Bukhari asserted that "the primary struggle that the Black community faces is one against racist oppression and economic exploitation."[3] Consequently, the Black Panther party articulated a programmatic ideology for confronting the intersection of race and class oppression, drawing heavily from Marxism-Leninism, Fanonian thought, and other forms of revolutionary socialism. Contrary to media images and cursory historical explanations, the Panthers additionally recognized the importance of combating male chauvinism and sexism, both within their own ranks and in the society at large. Black women experienced race and class through their gender identities, a process which constituted a "multiple jeopardy" of simultaneous, interconnected race, class, and gender oppressions.[4] As a consequence, black women operated on two distinct levels in the Black Panther party, filling both socially defined "masculine" leadership and cadre roles, and, perhaps more importantly, the "feminine" positions in community programs which composed the backbone of the organization.

In this context, Panther women confronted the multiplicity of their oppressions in totality, refusing to subsume sexism to racism when it was not politically and socially expedient—or desirable. Like their historical foremothers in past movements, black women's agency in the party allowed them to define the terms of their activism in relation to the reality of their race-class-gender "multiple jeopardy." Thus, studying the Black Panther party necessitates a shift in the ways in which scholars examine Panther women and their actions. Furthermore, the uniqueness and importance of the party's

Community Survival programs require another alteration in thinking. The violent cadre actions and "traditional" leadership structure of the Panthers receive undue attention in comparison to their overall significance. In confronting and emphasizing the interconnectedness of their race, class, and gender oppressions, black women, through their actions, refuted many of the charges of sexism leveled against the Black Panther party.

Racism, buttressed by capitalist economics, systematically oppressed black men, women, and children throughout the entirety of their history in the United States. Beginning with slavery, the continuous exploitation of the surplus value created by black labor by the white capitalist oligarchy served to create a perpetual state of black underdevelopment.[5] Slaveholding America never compensated blacks for their over 250 years of forced labor while in bondage. After emancipation, black men and women sold their labor power for starvation wages. When later trained in skilled and professional occupations, blacks, even when they possessed superior abilities, received lower wages than their white counterparts.[6]

Black women experienced race and class differently, and arguably more harshly, than black men. Though patriarchy indeed thrived in pre-capitalist societies, under capitalism it converged with racism to constitute a particularly oppressive system for black women.[7] As compared to their brothers, the labor of working-class and poor black women faced increased devaluation because of "double jeopardy"—the intersection of race and gender.[8] Race and sex discrimination have historically limited the employment options of black women, generally relegating them to low-status, low-wage work.[9] Employers exploited the surplus value produced by black women's labor, as the "last hired and first fired," even more than that of black men.[10] Adding to this economic oppression, under a racist, capitalist patriarchy white men exerted power over black women through a long history of system-sustaining, terroristic forms of violence, particularly of a sexual nature.[11] Additionally, black women faced internal oppression within the black community. This patriarchy differed from that of the white community in its basis in "racism and self-hate."[12] Black women largely performed the unremunerated labor within the home, while simultaneously working for wages in the public sphere.[13] This "double day" has its roots in slavery, when after an exhausting day in the fields or master's home, the burden of domestic work and childcare fell on black women,[14] rendering them "slaves of a slave."[15]

The intersection of gender with interconnected race and class oppressions, beginning with slavery, imprinted successive generations of African Americans with "socially and psychologically ... deforming practices, which ... continue to live in the Black psyche long after" legal Emancipation.[16] The prevailing discourse indicated that the institution of slavery and its white masters systematically emasculated black men. During slavery and after Emancipation, black men failed to provide economically and physically for their families, according to white standards. Further, 1950s scholars such as E. Franklin Frazier, Kenneth Stampp, and Stanley Elkins posited that this process endowed black women with the masculinity that their men once possessed.[17] Black men after the end of chattel slavery competed for jobs with white men, while black women, though providing an easily exploitable pool of labor, did not face competition for jobs to the same degree of intensity as men. Therefore, as Safiya Bukhari averred, "The Black woman had to be the breadwinner as well as homemaker."[18]

Rather than directing their anger over unemployment and underemployment at the racist, capitalist system, black men treated black women as the scapegoat for their symbolic (and sometimes physical) castration by white America. At the time Huey P. Newton and Bobby Seale began to preach

the need for Northern ghetto self-defense, Daniel Patrick Moynihan issued his now discredited sociological study of the black family and community, *The Negro Family: The Case for National Action*. Relying heavily on E. Franklin Frazier's analyses, Monyihan's work effectively pathologized the black family, attributing economic, social, and political flaws to the mythical black "matriarch," who metaphorically castrated the men in her family and community.[19] According to Frances Beale, a prominent member of the Student Non-Violent Coordinating Committee (SNCC), the image of the black woman "has been maliciously maligned."[20] While Beale admits that black men indeed faced emasculation in racist, capitalist America, the confluence of race and class increased the harshness of black women's oppression when their gender was factored into this complex web of identity. Consequently, this discussion of black emasculation ignores black women's history of oppression: "She has been sexually molested and abused by the white colonizer; she has suffered the worst kind of economic exploitation, having been forced to serve as the white woman's maid and as wet nurse for white offspring while her own children were ... starving and neglected."[21] In light of this exploitation, that these experiences are misapplied in order to pathologize the black family and black womanhood is "the depth of degradation."[22] In the context of the black crisis of masculinity and subsequent excoriation of black women, the Black Panther party operated, constructing an ideology and image for itself reflecting the intertwined complexities of gender, race, and class oppressions in the black community.

The Black Power movement within SNCC, as elucidated by Stokely Carmichael and H. Rap Brown in the mid-1960s, marked a Civil Rights watershed and important precursor to the Black Panther party. The young proponents of Black Power separated themselves ideologically and tactically from the liberal, church-based structure of the Civil Rights movement, facilitating the growth of black militant organizations.[23] The reformist goals of the Civil Rights movement allowed for the beginnings of a black middle class, accompanied by a drift of a segment of the community to the right of the liberal black establishment.[24] The prospect of black capitalist development as a panacea to race and class oppression entered the mainstream when in 1968 Richard Nixon embraced

> black power, [not] as some of the extremists would interpret it—not the power of hate and division, not the power of cynical racism, but the power the people should have over their own destinies, the power to affect their own communities, the power that comes from participation in the political and economic processes of society.[25]

For Nixon and other white conservatives, the most efficient solution to America's race problem was capitalist economic development, supposedly breaking the cycle of black dependency on public assistance and thereby eliminating racism.[26]

The Black Panther party concurred with Nixon's assessment of the importance of economics in black liberation; however, rather than accomodationist black capitalism, the Panthers injected the rhetoric of class struggle into the fight against racism inherited from the Civil Rights movement. Consequently, socialism, particularly Marxism-Leninism, served as a primary strand in Panther ideology. "It is no accident," the Party's official newspaper, *The Black Panther,* intoned in 1970, "that the vanguard party—without chauvinism or a sense of nationhood—should be the Black Panther Party."[27] In a 1969 issue of *The Black Panther,* Bobby Seale stressed that "a democratic dictatorship by the proletarian class, the people who the Black Panther Party members come from, is the real reason for the Party."[28] They departed from Marxist rhetoric in its conception of the lumpenproletariat. Marx considered the *lumpen*—"that passively rotting mass thrown off by the lowest layers of old society"—as an easily bribed, potential fifth

column against the revolution.[29] Because of the intersection of race and class oppression, fluidity existed between the dichotomous categories of "proletariat" and "lumpenproletariat," and some segments of the black community passed freely between these lower economic strata.[30] Thus the Panthers supplemented their Marxism-Leninism with the political theory of Frantz Fanon, who endowed the indigenous lumpenproletariat with the role of revolutionary vanguard. The relevancy of Fanon's philosophies to the milieu of U.S. blacks electrified the Panthers, who had never found "a major Marxist-Leninist theoretician who was primarily concerned about the problems of Black people."[31] Consequently, the Black Panther party targeted for recruitment members of the black community who composed the lumpenproletariat in Frantz Fanon's sense—those who "have nothing to lose and everything to gain."[32]

The Black Panther Party (for Self-Defense) from its earliest days in 1966 consciously advocated masculinist, male-centered revolutionary theory and praxis as a response to this historical crisis of black masculinity, which rose to the forefront of societal discourse through the Moynihan Report. In their October 1966 "10 Point Program: What We Want, What We Believe," founders Huey P. Newton and Bobby Seale submitted eight non-gender-specific demands for the betterment of the black community, and two explicitly for men: "We want all black men to be exempt from military service" and "We want freedom for all black men held in federal, state, county and city prisons and jails."[33] These demands, while necessary to stem the slavery-like incarceration of black men and their use in Vietnam as cannon fodder, ignored the long-standing pervasiveness of violence against black women in the United States.[34]

As indicated by their 10 Point Program, the Panthers directed their initial recruitment at men, specifically the "brothers off the block" who comprised the lumpenproletariat. The preliminary issue of *The Black Panther*, the official organ of the party, contained this exhortation to male members of the community:

> So, Brothers **and** Sisters everywhere: righteous BLACK POWER organized is where its [sic] at. The BLACK PANTHER PARTY FOR SELF-DEFENSE really has something going. These brothers are the cream of Black Manhood. They **are there for the protection and defense of our** Black Community ... BLACK MEN!!! It is your duty to your women and children, to your mothers and sisters, to investigate the program **of the PARTY.**[35]

The rhetoric of the reclamation of masculinity, set in a discourse that scapegoated black women for the castration of their men, ignored the economic and racial oppression of women, consequently dismissing their agency in struggles for social change.

From these raw ideological principles, it is not readily apparent that women composed approximately two-thirds of the party's membership at its height.[36] Because a wide array of forces aligned themselves against the Black Panther party, including the corporate media, politicians, police officials, and the Federal Bureau of Investigation, public perception of the group must be tempered with a dose of gendered reality.[37] Women's perceived public invisibility in the Black Panther party resulted from a complex mixture of these external factors, as well as Panther men's androcentric self-representation and some top-down organizational sexism. Compounding these immediate factors, black women's historic invisibility—the paradigm that establishes the category "black" as "black men," and the category "women" as "white women"—further defined the image and interests of the Black Panther party as male-centered. However, former Panther Malika Adams stressed that, "Women ran the BPP [Black Panther Party] pretty much. ... We actually ran the BPP's programs."[38]

Reflecting slavery's legacy in terms of gender roles, where the social construction of gender retained fluidity (if only for the convenience of the masters), women in the Black

Panther Party performed both the "male" and "female" roles.[39] Operating under the maxim that "a revolutionary has no gender," though often defined utilizing a male paradigm, Panther women participated in cadre actions and party leadership, sometimes committing violence, getting arrested, and facing legal charges. More importantly, however, women in the party occupied the roles defined as "female" by the dominant society, organizing and staffing the crucial Panther Community Survival programs. Black women, with their inherent agency, defined the terms of their participation in the Black Panther party through the prism of their historic race, class, and gender "multiple jeopardy."

In numerous accounts and recollections of the New Left, which existed concurrently with the Black Panther party, women (primarily white) suggest the large degree to which the men marginalized them to positions deemed "feminine." Barbara Haber, wife of Students for a Democratic Society "heavy" Al Haber, remembers: "I was kept from any real decision-making. For five years in SDS I never so much as chaired a workshop, much less made a speech. I was always being asked to stuff envelopes, and my husband was always being asked to make speeches."[40] Not only did this distinct gendered division of labor exist in SDS, but both the organization and the society at large degraded the work that women performed. Their experiences with sexism in the New Left provided much of the initial impetus for the women's caucuses and consciousness-raising groups that gave birth to the women's movement.

However, the women in the Black Panther party experienced sexism differently from those in the predominantly white New Left. In addition to interlocking race and class oppressions, black women carried baggage from slavery that white middle-class women lacked. Rather than enforced compliance with a lily-white standard of womanhood, the twin specters of the matriarchy thesis and Sapphire stereotype hung over black women in the Panthers.[41] Under these myths, the masculinity that whites ostensibly stole from black men and endowed to black women afforded women in the party a modicum of freedom in terms of organizational gender roles. Though the double-edged nature of these stereotypes pathologized their activities by contributing to the Sapphire myth—that of the emasculating, powerful black woman—they allowed Panther women to navigate between socially constructed masculine and feminine aspects of the party.

With this gender freedom, Panther women climbed the ladder of conventional leadership and participated in activities normally reserved for men. When Tarika Lewis and Regina Jennings joined the Panthers, the male leadership expected them to conform to the predominant notions about masculinity including physical training, the use of firearms,[42] and push-ups for discipline.[43] Elaine Brown, Kathleen Neal Cleaver, Ericka Huggins, and Afeni Shakur all rose to positions of prominence within the party leadership structure. These women were not exempt from the hard-nosed ruthlessness associated with "masculine" leadership—a grim indication of equality. Elaine Brown, appointed national Party Chairperson in Huey Newton's exile, demonstrated this egalitarianism of a violent nature in ordering the beating of a recalcitrant comrade: "Their punishment became unmerciful. When he tried to protect his body by taking the fetal position, his head became the object of their feet. The floor was rumbling, as though a platoon of pneumatic drills were breaking through the foundation. Blood was everywhere. Steve's face disappeared."[44]

As a consequence of this deviation from gender norms, Panther women faced similar persecution at the hands of the authorities. Or, as Bobby Seale once quipped when discussing the gender dynamics of the party, "The sisters can get arrested, too, just the same as the brothers."[45] Eldridge Cleaver's statement on Ericka Huggins, an East Coast Panther charged with murder in the New

Haven trial of 1969, indicates an important and often overlooked aspect of gender relations within the Black Panther party: ... "[T]hey didn't put her in a powder puffed cell. They did not make life easy for her. *But the pigs recognized a revolutionary woman to be just as much a threat as a revolutionary man*"[46] (emphasis added). Both the radical outlook of the Black Panther party and the immediacy of its fight against race- and class-based oppression shaped its gender ideology. Panther women, lacking many of the constraints that gripped their counterparts in the predominantly white New Left, advanced modestly in the male-defined power structure.[47] On the other hand, the purported equality of the "revolutionary has no gender" philosophy ignores particularities of black women's double jeopardy in the context of the Black Panther party. For example, Naima Major remembers that she was "devoted to the black revolution and the ten point program [and] commenced *with baby in sling* to doing the hard community work"[48] (emphasis added). Like black women since slavery, Major shouldered the burden for motherhood and childcare, while simultaneously working in the community. Similarly, Regina Jennings contended with the unwanted sexual advances of her section leader, which nearly forced her from the party.[49] Though Kathleen Neal Cleaver stresses that the party dealt harshly and resolutely with allegations of sexism, a step that the larger society had yet to take,[50] a gender analysis must join the multifarious interconnections of race and class identity traditionally applied to the party.

Therefore, scholars must consider women's major lines of participation and influence in the "feminine" side of the Black Panther party—its important and often overlooked Community Survival programs.[51] The Black Panther party's community work filled a void left by capitalist exploitation experienced through blacks' racial, and in many cases, gender identities. Because of the intersection of race and class oppression, a disproportionate number of black women and men lived with "deplorable housing, poor health care services, an unresponsive criminal justice system, inadequate diets, and substandard education."[52] Introduced in 1968, the programs included the most well-known Free Breakfast for Children and the liberation schools, as well as free clothing, medical treatment, and transportation.[53] The Panthers instituted these services to provide immediate relief to their communities, to inspire the community to militancy if the need arose, and to demonstrate the hypocrisy of a government that "did so little with so much," while the party members "did so much with so little."[54] As former Party Chairperson Elaine Brown emphasizes, the programs kept the work of the party politicized without resorting to masculinist gunplay, preventing the predominantly male cadre and central leadership from "being obliterated" by the police and FBI.[55] The inception of the Community Survival programs was significant because it confirmed a physical and very real commitment by the Panthers to their ideological claims of existing as the vanguard and voice for the [lumpen]proletariat. Ultimately, the Community Survival Programs served to legitimize the Panthers' ideology because "social practice is a criterion of truth."[56]

Women comprised the backbone of these Community Survival programs. While women served in national leadership positions and contributed to masculine-defined cadre work, the Black Panther party organized itself along gender lines, similar to other leftist groups and the society at large. The Community Survival programs, by nature, "often represented an extension of 'traditional' roles for women in the family: nurturers, caretakers of children, transmitters of morals."[57] Though men participated in Panther community activism as well, the burden of leadership and personnel for the programs fell on the women.[58]

Ironically, FBI Director J. Edgar Hoover perhaps accorded the most recognition to

Black Panther women and their work when he described the Free Breakfast for Children program as "the best and most influential activity going for the BPP and as such, is potentially the greatest threat to efforts by authorities ... to neutralize the BPP and destroy what it stands for."[59] These Panther women partook in meaningful, utilitarian work, whose effects far outlasted the violent aspects of the party. For example, the Oakland Community School, a liberated institution established by the Panthers and operated primarily by women, folded in 1982.[60] At this point, the majority of the male cults-of-personality were dead, incarcerated, or excommunicated from the party. Eldridge Cleaver, among the most well-known of the male figures, emerged a born-again Christian after returning from Algeria in 1975 and declared, "I'd rather be in jail in America than free anywhere else."[61] Conversely, the Black Panther women involved in the community programs constantly contributed to the propagation of Black Panther values and ideology in their communities; however, despite their dedication, they failed to receive the attention, be it positive or negative, accorded to the male Panthers as symbols of violent revolution.

Thus, from the example provided by the gender politics within the Black Panther party and its Community Survival programs, it becomes necessary for historians to shift the center of the discourse on activism. Both in and out of social movements, the duties male leaders and the society at large deemed as important and noteworthy were rarely ascribed to women. Former Panther Tracye Matthews recalls that "women were responsible in terms of both leadership and personnel for key party programs ... yet the media image of the Party was and is male-centered."[62] Studying these movements necessitates the redefinition of prevailing notions as to what constitutes important and revolutionary work. Furthermore, the particular political, social, and economic standpoints and identities of black women figure heavily into potential historical analysis. The multiple jeopardy black women confront in their lives—among others, the confluence of racism, classism, and sexism—necessitates the development of "a political ideology capable of interpreting and resisting that multiple jeopardy."[63]

Other historical examples provide evidence for the relevancy of such a project. In the male-dominated Universal Negro Improvement Association, Marcus Garvey's second wife, Amy Jacques-Garvey, unofficially ascended to the top leadership position while her husband languished in prison. Bound by a double jeopardy of gender and race oppression, Jacques-Garvey advocated a "community feminism," wherein female Pan-Africanists challenged sexism while remaining within the context of black nationalist struggle.[64] Community feminism allowed Jacques-Garvey and other members of UNIA to reconcile seemingly paradoxical helpmate and leadership roles, embracing "the multiple identities of black women."[65] Secondly, women who joined the Black Panther party undoubtedly watched closely the participation of the Civil Rights women who preceded them. Kathleen Neal Cleaver remembers hearing Diane Nash "speaking at Fisk University, [and] leading Black and White Freedom Riders onto Greyhound buses that got set on fire when they reached Alabama."[66] Nash and others wholeheartedly believed in the necessity of a political movement for the eradication of racism; thus, they "refrained from making critical judgments about the movement," or the role of women therein.[67] Within both the Universal Negro Improvement Association and the Civil Rights movement, black women actively decentered the epistemological model of feminist theory, in which white women universalized their particular experiences in constructing discourses on gender.[68] These women, embracing the confluence of their various identities, constructed a viable theory and praxis for their participation in social, political, and economic struggles.

Black women in these male-dominated movements possessed the agency required to complete their work as activists.[69] They simultaneously addressed their multiple race, class, and gender oppressions. Furthermore, black women's inherent agency allowed them to recognize the fluidity of their identities and the variability of oppression contained within each identity,[70] while acknowledging that when taken as an aggregate, these oppressions are inseparable.[71] Simply, since black women face at least one other facet of oppression as compared to white women, their histories cannot be viewed in a strictly parallel fashion.[72]The gender politics of the Black Panther party must be analyzed in this context as well.

The experiences of Panther women provide evidence for the necessity of a paradigm shift in analyzing the gendered realities of activist work and the inextricable interplay of race, class, and gender. That a number of Panther women, both high-profile and rank-and-filers, offer positive recollections and analyses (as well as constructive criticism) of the Black Panther party suggests their refusal to subsume the class- and race-based struggle to Women's Liberation defined on a white model.[73] The women involved in the Black Panther party exercised agency in participating in the Black Liberation struggle, when many of their white counterparts, stressing the primacy of gender, urged them to join the Women's Liberation Movement (WLM). However, in situating themselves within a socialist movement for racial liberation, Panther women often shared more in common with their historical mothers and contemporary brothers than with the white women concurrently agitating in the burgeoning WLM.[74] For instance, the WLM demanded the freedom to work outside of the home, a "freedom" black women always possessed as a result of race and class oppression.[75] The white feminist universalization of "womanhood" failed to apply to the historical and sociological reality of black women, providing impetus for them to reject the WLM in favor of Black Liberation movements.

Undoubtedly, the party contained chauvinistic and misogynistic leaders and members. Panther women, however, addressed this inevitable sexism on their own terms, rather than on those that white feminists established. Black Panther women declared the inseparability of race, class, and gender in their own identities and exercised agency in determining the realities of their participation in the party.[76]

Furthermore, these complex, intertwining identities necessitate a shift in the discourse on the gendered division of labor within the Black Panther party. While select Panther women rose through the ranks of the party and secured "male" leadership positions, women primarily participated in their local Community Survival programs.[77] In Panther community activism, women assumed gender roles reflective of those of the larger society, such as cooking and elementary education. However, the Community Survival programs outlasted any other sector of the Black Panther party. The Black Panther community work portended the greatest chance for the fulfillment of their ideological goals—imparting a revolutionary consciousness to the proletariat and particularly the lumpenproletariat—if not for the strategically targeted destruction by the FBI and police. Thus, thinking about the political and social expediency of various forms of activist work requires a paradigm shift. In terms of these activist groups, women rarely performed work that received showers of praise as "radical," both because of the capitalist patriarchy of U.S. society and the internal sexism of leftist groups. Women in the Black Panther party, while achieving a degree of "masculine" leadership, predominantly contributed to the Panther's highly successful "feminine" Community Survival programs.[78] Observers, for a number of plausible reasons—unfair mass media coverage, government infiltration, and short-sighted male Panther self-representation—assumed that the male-centered

gunplay and confrontations with police represented the quintessence of the party. Though women participated in these activities as well, they primarily staffed the Community Survival programs, which stand as the party's most positive and long-lasting contribution.[79]

As exemplified by Amy Jacques-Garvey and Diane Nash, holding "masculine" leadership positions while simultaneously performing work deemed "feminine" is not a paradox for many black women activists. Black Panther women embodied this duality admirably. They functioned on a multitude of levels within the organization, and in providing the backbone for the Community Survival programs, sustained the ideals of the Black Panther party long after its organizational demise. As their work in the party demonstrates, Panther women successfully navigated the confluence of their race, class, and gender oppressions, addressing their multiple identities on their own terms.

NOTES

1 Mumia Abu-Jamal, *We Want Freedom: A Life in the Black Panther Party.* Cambridge, MA: South End Press, 2004. 159.
2 Alice Walker, "They Ran on Empty." *The New York Times.* 5 May 1993.
3 Safiya Bukhari, "On the Question of Sexism in the Black Panther Party." *Reflections, Musings, and Political Opinions.* Unpublished manuscript, ca. 1997. 3.
4 The term,"multiple jeopardy," comes from Deborah K. King's essay, "Multiple Jeopardy, Multiple Consciousness: The Context of a Black Feminist Ideology." *Feminist Social Thought: A Reader.* Ed. Diana Tietjens Meyers. New York: Routledge, 1997. 227.
5 Manning Marable, *How Capitalism Underdeveloped Black America: Problems in Race, Political Economy and Society.* Boston: South End Press, 1983. 7.
6 Ibid., 7.
7 Ibid., 10.
8 The term "double jeopardy" comes from Frances Beale's essay, "Double Jeopardy: To Be Black and Female." Ed. Barbara A. Crow. *Radical Feminism: A Documentary Reader.* New York: New York University Press, 2000.
9 King, 224–225.
10 Bukhari, 1.
11 Deborah Gray White, *Ar'n't I A Woman?: Female Slaves in the Plantation South.* New York: W.W. Norton & Company, 1999. 164–165.
12 Bukhari, 1.
13 King, 224–225.
14 White, 112–113.
15 Beale, 155.
16 Abu-Jamal, 172.
17 White, 17-22.
18 Bukhari, 1.
19 White, 166.
20 Beale, 155-156.
21 Ibid.
22 Ibid.
23 Abu-Jamal, 159-160.
24 Rod Bush, *We Are Not What We Seem: Black Nationalism and Class Struggle in the American Century.* New York: New York University Press, 1999. 193.
25 Robert L. Allen, *Black Awakening in Capitalist America: An Analytic History.* New York: Doubleday and Company, 1969. 191.
26 Ibid., 192.
27 Huey Newton, "Minister of Defense, Black Panther Party, U.S.A." *The Black Panther,* 29 August 1970. Reprinted in G. Louis Heath, *The Black Panther Leaders Speak,* 77.
28 Bobby Seale, *The Black Panther.* 12 July 1969: 12-13. Reprinted in G. Louis Heath, *The Black Panther Leaders Speak,* 58.
29 Karl Marx and Frederick Engels, *The Communist Manifesto.* New York: International Publishers, 1948. 20.
30 Chris Booker, "Lumpenization: A Critical Error of the Black Panther Party." Jones, Charles E., ed. *The Black Panther Party Reconsidered.* Baltimore: Black Classic Press, 1998. 337-8.
31 Eldridge Cleaver, "On the Ideology of the Black Panther Party." *The Black Panther,* 6 June 1970. Reprinted in G. Louis Heath, *The Black Panther Leaders Speak,* 75.
32 Frantz Fanon, *The Wretched of the Earth.* New York: Grove Press, 1963. 61.
33 Philip S. Foner, *The Black Panthers Speak.* Philadelphia: J.B. Lippincott Company, 1970. 3.
34 Beale, 156.
35 Foner, 12.
36 Kathleen Neal Cleaver, "Women, Power, and Revolution." Kathleen Cleaver and George Katsiaficas, ed. *Liberation, Imagination, and the Black Panther Party: A New Look at the Panthers and Their Legacy.* New York: Routledge, 2001. 125.
37 For an exhaustive description of the U.S. sanctioned internal war against the Black Panther party, see Ward Churchill's "'To Disrupt, Discredit, and Destroy': The FBI's Secret War Against the Black Panther Party." Kathleen Cleaver and George Katsiaficas, ed. *Liberation, Imagination, and the Black Panther Party: A New*

Look at the Panthers and Their Legacy. New York: Routledge, 2001.

38 Trayce Matthews, "'No One Ever Asks, What a Man's Place in the Revolution Is': Gender and the Politics of The Black Panther Party 1966-1971." Jones, Charles E., ed. *The Black Panther Party Reconsidered.* Baltimore: Black Classic Press, 1998. 291.

39 White, 22.

40 Dierdre English, et al. "The Impasse of Socialist Feminism: A Conversation." *Women, Class, and the Feminist Imagination: A Socialist-Feminist Reader.* ed. Karen V. Hansen and Ilene J. Philipson. Philadelphia: Temple University Press, 1990. 301.

41 King, 227.

42 Matthews, 307.

43 Regina Jennings. "Why I Joined the Party: An Africana Womanist Reflection." Charles E. Jones, ed., *The Black Panther Party Reconsidered.* Baltimore: Black Classic Press. 1998. 261.

44 Elaine Brown. *A Taste of Power: A Black Woman's Story.* New York: Anchor Books, 1992. 369–371.

45 Angela D. LeBlanc-Ernest, "'The Most Qualified Person to Handle the Job': Black Panther Party Women, 1966-1982." Charles E. Jones, ed., *The Black Panther Party Reconsidered.* Baltimore: Black Classic Press, 1998. 315.

46 Cleaver quoted in Abu-Jamal, 175.

47 King, 227.

48 Abu-Jamal, 177.

49 Jennings, 262.

50 Cleaver, "Women, Power, and Revolution," 126.

51 Matthews, 291.

52 JoNina M. Abron, "'Serving the People': The Survival Programs of the Black Panther Party." 178.

53 Ula Y. Taylor and J. Tarika Lewis. *Panther: A Pictorial History of the Black Panthers and the Story Behind the Film.* New York: New Market Press, 1995. 99–110.

54 Ibid., 99.

55 Brown, 248.

56 Yvonne King, "Panel Discussion and Q & A," Black Panther Film Fest. Church of the Advocate, Philadelphia. 20 February 2005.

57 Matthews, 291.

58 Ibid.

59 Ibid., 292.

60 LeBlanc-Ernest, 325.

61 Taylor and Lewis, 126.

62 Matthews, 291.

63 King, 236.

64 Ula Y. Taylor, "'Negro Women Are Great Thinkers As Well As Doers': Amy Jacques-Garvey and Community Feminism in the United States, 1924–1927." *Journal of Women's History.* Summer 2000, 12, 2. 105.

65 Ibid., 106.

66 Cleaver, "Women, Power, and Revolution." 123.

67 Anne Standley, "The Role of Black Women in the Civil Rights Movement." Vicki L. Crawford, Jacqueline Anne Rouse, and Barbara Woods. ed. *Women in the Civil Rights Movement: Trailblazers & Torchbearers, 1941–1965.* Bloomington: Indiana University Press, 1990. 187.

68 Taylor, "'Negro Women Are Great Thinkers As Well As Doers': Amy Jacques-Garvey and Community Feminism in the United States, 1924-1927," 106.

69 King, 225.

70 Ibid., 223.

71 Ibid., 221.

72 Ibid.

73 See the first-person accounts of the following former Panther women: Brown, *A Taste of Power: A Black Woman's Story;* Cleaver, "Women, Power, and Revolution"; Regina Jennings, "Why I Joined the Party: An Africana Womanist Reflection"; Matthews, "'No One Ever Asks, What a Man's Place in the Revolution Is': Gender and the Politics of The Black Panther Party 1966-1971."; Assata Shakur, *Assata: An Autobiography.* Chicago: Lawrence Hill Books, 1987.

74 Beale, 160.

75 King, 236-237.

76 Ibid.

77 Matthews, 291.

78 Ibid.

79 Ibid.

Functionalism and the Aristotelian Mind

Joshua B. Wright,
The College of New Jersey
Faculty Sponsor:
Professor John Sisko,
Department of Philosophy and Religion

Abstract

This paper presents a new direction in the contemporary debate over Aristotle's status as a functionalist (i.e., a philosopher who believes that the mind can be best understood as an algorithm executing on arbitrary hardware). Efforts at deciding this conflict have thus far focused on Aristotle's theory of perception and have attempted to argue by analogy to the mind. The results of these attempts, however, have proven inconclusive. In recognition of this impasse, this paper reframes and addresses the question of Aristotelian functionalism by returning to fundamental concepts and source materials of *direct* relevance.

To this end, the paper offers an analysis of modern functionalism, explicating its essential features and aims and comparing this doctrine to the understanding of function in Aristotelian philosophy. It concludes that Aristotle's conception of "functionalism" is not in harmony with modern varieties. Finally, the paper explores Aristotle's direct source material on the subject of the mind and concludes that what Aristotle says in particular about the mind is not favorable to functionalism, making a special exception for mental functionalism unlikely.

Introduction

Since the publication of M. F. Burnyeat's essay "Is an Aristotelian Philosophy of Mind Still Credible?" the question of whether Aristotle's account of the relationship between form, matter, and the soul implies mental functionalism has become the subject of considerable controversy. Since the answer to this question would have far-reaching consequences for our understanding of Aristotle's theory of soul, Burnyeat is to be commended for initiating this debate. Unfortunately however, the structure of Burnyeat's seminal essay has led participants in the discussion to focus on the Aristotelian theory of perception, a matter that I take to be only indirectly related to the issue at hand.

This essay, then, will take a different approach from most others in the literature on functionalism and the Aristotelian mind. First, as "function" unquestionably plays a vital role in Aristotle's theory of the soul, we will lay the appropriate groundwork for comparative analysis by examining what modern functionalism is, and what it is not; we will examine the sorts of commitments that a proponent of functionalism is accountable to. Next, we will see if Aristotle's notion of "function" in general resonates with contemporary functionalism—are the aims and process the same or similar? Lastly, and most importantly, we will investigate what Aristotle says in particular about the mind and its relation to the body. Only after these considerations have been satisfied can we hope to acquire an accurate picture of Aristotle's own position, and it is my belief that this position is not one that can rightfully be described as functionalist in any contemporary sense.

Functionalism

Theories on the relationship between the mind and the body are typically concerned with answering three fundamental questions: (1) What are the necessary conditions for a being to have a mind? (2) What are the sufficient conditions for a being to have a mind? (3) What is the nature of a mind? Functionalism is part of a family of theories such as logical behaviorism that assert that these three questions ultimately collapse into one: What does the mind *do*? (Morton 298). In answer to this question, functionalism offers that "mind" is best understood as the execution of a particular special algorithm or a member of a certain class of special algorithms (which may be infinite in size—Hilary Putnam's most recent position) by a system. Perhaps the most important consequence of this theory is "compositional plasticity": Since the sufficient conditions for mindedness are satisfied by the execution of the mind algorithm, any physical system with the requisite formal properties to instantiate the algorithm is a candidate for having a mind; what a system is constructed out of has, in and of itself, no bearing on its potential to have a mind.

Several points of clarification are in order, however. Recall that functionalism is an answer both to the question of necessary and sufficient conditions as well as an answer to the question of nature—when a mind is, as well as what a mind is. It is not enough that some algorithm merely be associated with mental activity; to satisfy the functionalist thesis, mental activity *in toto* must be the execution of an algorithm. To understand this distinction and its importance for functionalism, I will introduce two examples. At least one of these cases will undoubtedly appear absurd. Nevertheless, I wish to argue that both present essentially the same situation, and that neither is compatible with functionalism.

Suppose that in the year 2010, humankind makes an amazing discovery: The foundation of the human mind, far from being any collection of neurons, is actually a very small lump of undifferentiated blue tissue in the center of the brain. In fact, this particular shade of blue (thereafter known as "hyper-intelligent blue") is the one necessary and sufficient condition for consciousness in general. Whatever is painted hyper-intelligent blue will thereby become a minded being, although it will not be able to perceive the environment without a suitable collection of perceptual faculties. Now, in this case, there is a very simple "algorithm" (in the broadest sense) for becoming minded and preserving a mind—stay blue. Nevertheless, this situation obviously does not satisfy functionalism, as the principle for creating a mind has no formal relation to the activities of intentional states within the mind; it is a totally coincidental relationship.

Imagine another hypothetical situation, one in which scientists discover the principle underlying human neurobiology that allows the brain to give rise to consciousness. They learn that when a set of working perceptual organs are brought together within a living body and wired together in a certain way, that a minded being comes into existence. Crucially however, the researchers also observe that the activities within the organs and this communication matrix, "the brain," *underdetermine* the doxastic content of the mind. Although there is a considerable degree of formal correspondence between brain activity and mental life to be found (in fact, the researchers become quite adept at determining the qualitative content of a subject's mind by examining patterns of activity within his or her brain), the scientists are nevertheless unable to find a complete correspondence between intentional states and brain activity, even on a subject-by-subject basis.

Let us also suppose that some years later it becomes possible to create artificial functional analogues of the human brain and sensory organs, and scientists discover that when these are composed into a cohesive system that a minded being is also created

(for verification: Like a human being, it too displays intelligent behavior that cannot not be fully accounted for by reference to the activities in its hardware). Pressing on, they ultimately provide compelling empirical evidence that any system that is capable of instantiating the functional relationships defined by their model will have a mind, regardless of composition.

Although our second example certainly presents a more plausible scenario than the "hyper-intelligent blue" case, from the standpoint of functionalism both are equally unacceptable. The second example makes the functional activities of the brain and its sensory organs more than coincidentally related to those of mental life, and even satisfies the criterion of compositional plasticity, but it still leaves the functional activities of the mind underdetermined by those of the brain. Having a functional system organized in a special way is a necessary, sufficient, and essential *condition* for having a mind in example two, but it is not all there is to having one—more is going on in the mind than is going on in the functional system of the brain and its sensory organs. In short, the mind is not an instantiated algorithm; the essential criterion of identity is not satisfied.

Thus, in concluding the foregoing discussion we must find that Aristotle has a set of rather specific and strict standards to meet if his hylomorphism is to count as functionalism. It is not sufficient that Aristotelian "functionalism" provides a set of necessary and sufficient functional conditions for the emergence of a mind, or the possession of a mind. Mind must *be*, for Aristotle, an enmatterable, multiply-realizable form, whose functional activities underlie and circumscribe all mental activities. If Aristotle does not seem to support this thesis, then we have no justification for portraying him as anything like a contemporary functionalist.

Regarding Perception

As this essay will give almost no consideration to what has proven to be the focal point of the general discussion on Aristotle and functionalism, the Burnyeat/Sorabji debate, I thought that it would behoove me to offer at least a brief justification for my decision (beyond the opinion expressed in the introduction) before moving on to Aristotle's doctrine of the soul.

Quite simply, my choice to avoid Burnyeat/Sorabji reflects my belief that this discussion is, essentially speaking, quite irrelevant to the issue of whether Aristotle is or is not providing a functionalist thesis in *De Anima.* Whether it is Professor Burnyeat or Professor Sorabji who describes "taking on form without matter" accurately, the issue of Aristotle and functionalism remains; either account of perception is potentially compatible with, or exclusive with respect to, mental functionalism. This is perhaps easiest to see in the case of Sorabji's theory of perception. Even if we were to assume that the Aristotelian doctrine of perception provides a model for mental acts, and that Sorabji's description of literal qualitative assimilation is correct, we could still not thereby conclude that Aristotle was advancing functionalism. After all, Aristotle might believe that by their very nature only certain sorts of matter were capable of taking on forms in the proper way to make mental life possible, thus violating the compositional plasticity requirement. Putnam and Nussbaum, in their reply to Burnyeat, actually seem to realize this theoretical disconnect (Nussbaum 36) but curiously go on to spend most of their time discussing perceptual tangents anyway.

Function and the Soul

It is certainly the case that *De Anima* and other texts such as *Parts of Animals* present us with an account of the soul that stresses the function of an organism's faculties as being essential to their nature. This is apparent very early on in *De Anima* in sections such as 412b10-413a10, where it is stated that if an

organ of perception such as the eye had a soul, that its soul would be its functional actualization as an organ.* In *Parts of Animals* (641a1, for instance), Aristotle notes that a hand can only be a hand when it functions appropriately within the context of a living organism, and that when severed, or dead, there is a sense in which it is not a hand at all, except in name only. We are even told at 408b20 in *De Anima* that sight could be restored in an old man by replacing a poor eye with the "right kind of eye." Furthermore, like modern functionalists, Aristotle is deeply interested in combating reductionism. In both *Parts of Animals* and *De Anima,* Aristotle criticizes all attempts to equate the soul of a creature with its matter (for example, in *Parts of Animals* contra Democritus at 640b30, and in *De Anima* at 405a30-405b10 against Hippo and Critias, and more substantively at 409b20-410a10 against Empedocles). Moreover, we are confronted again and again with statements asserting that in our attempt to understand living organisms, a knowledge of formal properties is superior to one confined to material properties. Consider *Parts of Animals* 641a30:

> [I]nasmuch as it is the presence of the soul that enables matter to constitute the animal nature, much more than it is the presence of matter which so enables the soul, the inquirer into nature is bound to treat of the soul rather than of the matter[.]

So, it is certainly clear that Aristotle's theory of function and the soul has resonances with tenants of contemporary functionalism; however, it remains to be determined whether these resonances are only superficial or reflect deeper harmonies.

First, it must be noted that Aristotle's functionalism is a rather different sort than contemporary functionalism. Aristotelian functionalism is essentially teleological in its concerns and orientation. Why do creatures have the various parts that they do, and how does a creature's constitution reflect its form of life? These are the sorts of questions that consume *De Anima.* A creature's soul is an *externally* qualified functional system; the function of a faculty is determined by the role that it plays in helping a creature to interact with its *environment.* By contrast, functionalism is principally concerned with the relationship between internal functional mental states, that is, it is an *internally* qualified functional system (Morton 314). Although sets of physical conditions such as "boiling blood" will never for Aristotle be capable of providing sufficient conditions for a state such as anger given the functional requirement of perceiving a slight, such relations are not purely internal; they depend upon an *external* functional relation to final causes (Irwin, 131). Recall that for Aristotle, I can feel angry, and even *believe* myself to be angry, and still fail to be angry (if I lack the appropriate final cause). Moreover, Aristotle seems to hold the notion of a general "definition" of soul in

* Although Aristotle also raises the issue of artifacts here and in *Parts of Animals,* I will not delve into that discussion in this essay, as artifacts seem to be introduced only as roughly analogous models to the souls of living creatures, and there are very good reasons for thinking that the way in which function determines the nature of an artifact is not a good model for the relationship between function and the soul anyway. For instance, inanimate objects can receive what might be called "intentional blessing"—I can make a suitable natural object into a tool by using it in a certain way, or even by merely *recognizing* that it could be used in a certain way, all without changing anything in the object itself; the same natural object could prove to be excellent for two or more quite different uses, neither of which has any logical priority. Furthermore, the functionality of inanimate objects (and of machines in general) seems to be entirely externally qualified with respect to *minded beings.* No matter how refined or apparently well suited to any task, a non-living object that results from natural processes can never have a "function" prior to an act of intentional blessing. Even if by some freakish convergence of coincidences a tornado created a perfect replica of a Frank Lloyd Wright prairie style home out of the debris of a town, it would still not be a house. Living creatures on an Aristotelian account do not share this limitation. Their very mode of existence assures them of a certain kind of soul—one might say that their functionality is externally qualified with respect to the *environment* (consider Aristotle's account of the spontaneous generation of various mollusks). Nor can living beings come about by intentional blessing—no matter how much I treat my little wooden puppet like a child, he will never become a real boy by virtue of such acts alone.

fairly low esteem; at 414b20-25 in *De Anima*, Aristotle asserts that, although a general definition of soul must exist, it does not capture anything essentially important for the reality of a living organism. Rather, each organism must be investigated in particular, according to its own definition. Obviously, this does not fit very well with functionalism and its notion of a multiply-realizable algorithm for the mind.* Of course, it may be offered that we would be in error to use a functional picture of biology to argue against the possibility of an account of mental functionalism. That logic, however, cuts both ways; what we need to do is examine Aristotle's discussion of the relation between mind and body directly.

Before turning to Aristotle's comments on mind and body, however, it is worth noting that it is by no means clear that Aristotle is offering a fully "functionalist" picture even in his biology. Nowhere is it stated explicitly in *Parts of Animals* or *De Anima* that a creature is defined *exclusively* by its form. Usually, Aristotle encourages a formal orientation in how we study and classify animals, but he readily admits that the best natural scientist is familiar with both formal and material causes—they are both intimate parts of what a creature is:

> Now which of these is really the natural philosopher? The man who ignores the formula and is only concerned with the matter, or the man who is only concerned with the formula? Probably the man who bases his concept on both. ... The natural philosopher's concern is with all the func tions and affections of a given body, i.e. of matter in a given state.
> —*De Anima*, 403b5-403b10

Indeed, Aristotle's doctrine of final causes oftentimes seems to violate the spirit of contemporary functionalism directly. Consider, for instance, his notion of hypothetical necessity, whereby the needs of a creature determine a fairly narrow set of possible material substrates that may be used to instantiate an organ:

> There is, however, the third mode, in such things at any rate as are generated. For instance, we say that food is necessary in neither of the two modes, but because an animal cannot possibly do without it. This third mode is what may be called hypothetical necessity. For if a piece of wood is to be split with an axe, the axe must of necessity be hard; and, if hard, must of necessity be made of bronze or iron. Now exactly in the same way the body, since it is an instrument—for both the body as a whole and for its several parts individually are for the sake of something—if it is to do its work, must of necessity be made of such and such a character, and made of such and such materials.
> —*Parts*, 642a1-642a10

Now, although Aristotle allows that there may be some variation in the materials chosen by hypothetical necessity for the creature's material constitution, this by no means supports the sort of radical multiple realizability required by functionalism. To the contrary, Aristotle's emphasis is on the particularity of the material choices made by hypothetical necessity and on the necessary strength behind those choices. We would not expect to see beings with the same modes of life with radically different material constitutions. As Aristotle goes on to say:

> It is plain then that there are two modes of causation, and that both of these must, so far as possible, be taken into account, or that at any rate an attempt must be made to include them both; and that those who fail in this tell us in reality nothing about nature.
> —*Parts*, 642a15

So, a crucial part of the formal definition of a creature will be the hypothetically necessary principles that determine *just* what specific

* It must be admitted that not all versions of functionalism stipulate a single algorithm, or even a small family of algorithms, as being mental algorithms. Putnam, for instance, has given up this very notion in favor of a version of functionalism that allows for both compositional *and* computational plasticity (this is stated explicitly in *Essays on*, 48, for instance). However, there must be a limit to such plasticity, lest we risk the possibility of *any* functional system having a mind. Any serious philosophy of functionalism must be somewhat exclusive, after all, and it should at least enlighten us as to the general nature of mental acts. Aristotle's comments tend to suggest, by contrast, that a general definition of the soul would be useless because it would be *vacuous*, which is a very different sort of weakness than would be exhibited by a mature functionalism in the face of computational plasticity.

material basis that it does have. This is quite contrary to functionalism; good functionalist theories are not supposed to be in the business of making choices with such specificity. Nevertheless, for Aristotle, this process is an essential part of the "functionalism" of a living creature.

We must also be careful of reading too much into Aristotle's exhortations in *Parts of Animals* to concern ourselves with matters of form. Although *Parts of Animals* makes it very clear that a proper scientific methodology will produce classifications that track metaphysical facts (and that improper methodologies will all fail because they cheat reality in some way, e.g., the "method of dichotomy" in animal classification), Aristotle was not in *Parts of Animals* attempting to provide a *universal* methodology whereby *all* entities can be suitably classified as to their metaphysical status. Aristotle's aim was to provide a metaphysically founded program for zoology, i.e., a way of categorizing and understanding *natural living things*. As such, there is little reason to believe that it would be a reliable guide to making metaphysical judgments about androids, cyborgs, neural network simulations running on computers, or synthetic beings in general. It must be remembered that natural life forms exist within a very special teleological context for Aristotle—their structure is a necessary consequence of the way that the world is. In fact, there is a sense in which the structure of the world and the structure of the creatures that inhabit it it is actually a *mutually* entailed relationship, given that the species are eternal (they have been around for as long as the world has). A dolphin could not be anything other that what it is; its entire being, down to its material makeup, is determined by necessity. This sort of view allows a theorist to make assumptions when discussing natural creatures that could not be made when examining beings in general.

Indeed, it may very well be the case that Aristotle's confidence in our ability to speak primarily about living beings on a formal level profitably is a consequence of his belief that the material constitution of a living thing is very specifically determined by hypothetical necessity. After all, if it really *were* possible for animal souls to be instantiated in all sorts of different materials, it would seem that we would be committed to a host of complicated considerations that are not even mentioned in *Parts of Animals* or *De Anima*. Nowhere does Aristotle worry, for instance, about what sort of functional descriptions are necessary so that we may differentiate a real instance of a life-form class from a mere automaton. It is taken for granted that we are dealing with natural things, and that natural things must be composed in certain ways.

In the final analysis, I think it must be said that Aristotle's general theory of hylomorphism cannot be regarded as essentially functionalist—there are simply too many conflicting elements. Whether the problems with attempting to marry a functionalist theory of mind to Aristotle's theory of soul are so great as to be insurmountable is not a problem that I intend to address, however. First, I think it is likely to be a futile endeavor; there is simply not enough information to exclude the possibility completely. Second, I think that more compelling text is available in what Aristotle says on the subject of the relationship between mind and body specifically, text to which we shall now turn.

The Aristotelian Mind-Body Relationship

Recall for a moment the second scenario in our earlier discussion of functionalism; I hope that this was familiar to the reader from another context. Indeed, up until my mention of the discovery of multiple realizability, what was really being described was a modernized version of Aristotle's own basic picture of mental life, with the common sensorium moved from some mysterious organ around the heart to the brain. What's more, this is a very conventional, well-founded, modest interpretation of Aristotle's relationship between the mind and body as related

in *De Anima*. It is also a picture that, as I have already argued, is incompatible with functionalism. I will now go about reinforcing that assertion and bringing in textual evidence to offer direct support that this is indeed Aristotle's view on the relationship between the mind and body.

Just what does Aristotle say about the mind-body problem? Although one often reads comments suggesting that Aristotle says very little about the mind in its relation to the body, in truth he says quite enough for us to form a significantly detailed picture of mental life—one that does not favor a functionalist interpretation. First, let us refine the issue by *excluding* along with Aristotle everything that is non-essential to the mind, the rational faculty of a human being. Today, we tend to allow a very wide range of phenomena membership in the class of "mental content"—emotional states, desires, acts of recollection, intentional states, etc. Aristotle, however, not unlike Descartes, takes a somewhat less liberal view; the Aristotelian "mind" is far more narrowly defined. Indeed, Aristotle excludes from the domain of the mind such faculties as those of memory and recollection, sense reception and integration, and even of imagination (a nice summary is given in 429a1-429b10). All of these faculties are physically instantiated—the mind, by contrast, is said by Aristotle to lack an associated organ (429b20), and that this lack is moreover actually crucial to its ability to receive the appropriate forms.

Aristotle's exclusion of the aforementioned faculties from the mind allows us to eliminate a significant number of potential red herrings in our investigation. First, it is clear that arguments by analogy with respect to other faculties of the soul such as the common sensorium will fail—the form of the mind is simply not like the form of these organs. We are told that it is somehow "more divine" (408b20) and presumably survives the death of the individual in some sense, that it lacks an organ, and that it does not have an ordinary mean state with respect to its natural objects like the organs of perception (429b1-429b10). Accordingly, the passage that Putnam interprets as having functionalist overtones (Nussbaum 39) cannot, even if his interpretation is accurate, have direct relevance for our discussion of mental functionalism. Aristotle has, after all, been quite explicit in stating that animals lack a rational faculty (a true mind) and thus may be led about quite haphazardly by their malfunctioning imaginative intuitions:

> Again, because imaginations persist in us and resemble sensations, living creatures frequently act in accordance with them, some, viz., the brutes, because they have no mind, and some, viz., men, because the mind is temporarily clouded over by emotion, or disease, or sleep.
> —*De Anima*, 429a5-429a10

Thus, even if we could describe animals *entirely* in functionalist terms, this fact would not allow us to extend the analysis to the nature of the mind. It is simply a different sort of "thing" from the faculties that control animals—they haven't minds at all.

What then of the mind on its own terms? One theme that is hammered over and over again is Aristotle's reluctance to associate mental activity with any bodily organ. This is not confined to an isolated passage, but characterizes his treatment throughout the entire text whenever the relationship between the mind and soul arises. From the very beginning of *De Anima* Aristotle offers that the mind is a special case, and that it perhaps does not have a physical organ unlike other faculties of the soul. His assertions become ever more bold from there. At 407b3 we are even told that "*the inescapable association of the mind with the body would be wearisome; such a conception must be rejected, if it is true that it is better for the mind to be without the body, as is usually said and widely accepted*"—a thesis that Aristotle seems to accept, given his affinity for "useless knowledge" and a disembodied Prime Mover who sits in endless contemplation of the most practically irrelevant ideas available.

As Putnam would likely point out, however, the fact that in human beings the mind is not

located in the body would not *conclusively* prove that functionalism is false—after all, that could happen to be a contingent characteristic of the way that we are made up. Some other creature might exist whose mind was instantiated in its body. The problem is that this reply is unavailable to the student of Aristotle, as *De Anima* makes it abundantly clear that one would, for a variety of reasons, reject *a priori* the suggestion that a mind might be physically instantiated—it is an inherently bad idea on a number of levels. Such a theory would violate hypothetical necessity, fail to account for the differences in the way that the mind works as compared to perceptual organs, and even be aesthetically repulsive from Aristotle's perspective. Most pressingly, Aristotle's belief that the non-material constitution of the mind is *essential* to its ability to do its job evidences an outlook totally alien to functionalism. Aristotle is explicitly worried that physical systems cannot even *undertake* the sort of processes that are crucial to the mind. This is a very different kind of anti-reductionism to that held by functionalists. While functionalists agree that mental states cannot be totally reduced to physical states, this is because they believe we must attend to the far more crucial functional processes that physical systems *undergo*. Aristotle, by contrast, does not even wish to "associate" (let alone instantiate) the processes of the mind with those of a physical organ. This is an entirely different stance, and one that violates an essential tenet of functionalism—compositional plasticity. We must accordingly conclude that Aristotle simply cannot be a functionalist in the contemporary sense.*

I think it is telling that about halfway through *Changing Aristotle's Mind*, Putnam and Nussbaum seem to forget the issue of compositional plasticity altogether. One reads a great deal about "intentional systems" that will not go away or be reduced, but very little about whether Aristotle really believes that these intentional systems could be realized with, say, a very large number of Chinese men. Perhaps this is because once we have left the somewhat ambiguous domain of perception, Aristotle's obvious distaste for exactly the sort of thesis that Putnam and Nussbaum are advancing becomes overwhelmingly apparent. Can anyone, having read *De Anima*, really claim with a straight face that Aristotle would support the hypothesis that a mind could be realized in a suitably contrived system of water pumps?

The pleasant surprise implicit in the conclusion we have reached is that every major figure involved in the controversy over Aristotle and functionalism, for whatever flaws may be present in his or her respective account, has nevertheless managed at worst to win for losing. Burnyeat has been right all along—Aristotle is not a functionalist. This is not, however, a direct consequence of his rejection of a Sorabjian theory of "taking on form without matter" (indeed, much evidence suggests that literal assimilation is the most elegant, plausible way to interpret Aristotle's theory of perception). For their part, Putnam and Nussbaum are ultimately right; although Aristotle is not a functionalist, the general Aristotelian picture of the mind-body relationship does not have to be junked. In truth I think that Aristotle is offering us a theory that more or less reflects our intuitions about how the mind is related to the body. Although we nearly all believe that the functioning of the brain is a *part* of what defines human consciousness, I would venture to say that even at the dawn of the

* One possible reply to this assessment is that Aristotle might still have been a functionalist *in principle*, only one mistaken about the capabilities of matter. Perhaps he was simply too skeptical about the sorts of processes that material stuff could undergo—had he known more, he might have conceded that material structures could be acceptable vehicles for mental acts. This objection, however, proves too much, for taken to its logical conclusion it implies that we are nearly all closet functionalists (particularly if the person advancing this objection is committed to computational plasticity!). To claim that anyone is a functionalist who would concede that material systems could carry out mental processes if only they could do what minds can do is paradoxically to subsume a great many anti-functionalist thinkers under the banner of functionalism.

21st century, only a minority of us believe that the mind is completely identifiable with some sort of program "running" on the brain. It may very well be the case that there simply are, in the world, mental properties that are associated with certain forms actualized in certain kinds of matter, although those forms are not identical with those properties. In the end, is this hypothesis any more fantastic than the apparent paradoxes, such the Chinese Room problem, missing qualia, the problems of identity, etc. that functionalism forces us to reckon with? Indeed, the final irony of this debate may be that in their rush to defend Aristotle's relevance, Putnam and Nussbaum have struggled to justify a controversial contemporary interpretation of the antique philosopher that is actually more problematic than the orthodox view suggested by a casual reading of his own writings.

BIBLIOGRAPHY

Aristotle. Hett, W. S., Trans. *Aristotle VIII: On the Soul; Parva Naturalia; On Breath.* Cambridge, MA. Harvard University Press, 2000.

Irwin, Terence. *Classical Thought.* New York. Oxford University Press, 1996.

Morton, Peter Alan. *A Historical Introduction to the Philosophy of Mind.* New York. Broadview Press, 1997.

Nussbaum, Martha, Ed. & Rorty, Amelie, Ed. *Essays on Aristotle's De Anima.* Oxford. Clarendon Press, 1992.

Physical Appearance and Self-Esteem in College Students

Adriana Pilafova,
The College of New Jersey
Faculty Sponsor:
D.J. Angelone and Katrina Bledsoe,
Department of Psychology

Abstract

The purpose of this study is to examine the relationship between self-esteem and appearance for male and female undergraduate students. It was hypothesized that men would have a higher self-esteem and body-esteem than women. It was also hypothesized that lower scores on the body mass index would be associated with greater self- and body-esteem. The sample consisted of 72 men and 81 women from a small northeastern college. Participants completed the Rosenberg Self-Esteem Inventory, a body-esteem inventory, and a general questionnaire. A series of independent-groups t-tests and a one-way ANOVA revealed statistically significant relationships lending support to both hypotheses. It appears that men have higher self-esteem and body-esteem than women. Also, lower body mass scores appear to be related to higher body-esteem and self-esteem for all participants. Implications and limitations of this study are discussed.

Introduction

In the complexity of today's society there are many things that can affect self-esteem and the way people view their own physical appearance. Men and women are bombarded with images of individuals exemplifying a perfect physique (Hewitt, Flett, & Ediger, 1995). This may increase self-imposed pressure to maintain youth and meet the unrealistic beauty standards set by society. These constant demands, along with the need to receive approval and acceptance from others, often may lead to decreased self-esteem (Crocker, Luhtanen, Cooper, & Bouvrette, 2003).

The evolutionary perspective presents one way of understanding the focus on physical appearance and attractiveness in society. This view suggests that reproductive potential is the most essential quality of a future female mate. Cues that signal this potential are seen as the most important determinants of attractiveness and may include good mothering ability, femininity, and resistance to diseases (Wade, Shanley, & Imm, 2003). Thus, men are attracted to women who possess these characteristics, and women attempt to fit this "model" to attract men, further perpetuating the societal focus on appearance. Additionally, as people age they may become more prone to societal pressures to marry and raise a family. According to the evolutionary view, women will attempt to maintain their attractiveness in order to fulfill those societal expectations (Wade, Shanley, & Imm, 2003). This may lead to a decrease in self-esteem over time. Thus, self-esteem for older women may be lower than the self-esteem of adolescent girls.

Furthermore, appearance has been linked to self-esteem. According to the self-determination theory, self-esteem is higher when it is based on more abstract variables, such as values and distinctive features of one's personality, than when it is based on tangible attributes such as appearance (Crocker, Luhtanen, Cooper, & Bouvrette, 2003). Thus, it is expected that people who base their self-esteem on variables such as appearance will display

lower self-esteem. Women conforming to the above-stated characteristics of attractiveness may have lower self-esteem than those who rely on more abstract variables.

Other theories attempt to provide insight into the relationship between physical attractiveness and self-esteem. One approach uses self-concept, which may also play a role in the relationship between self-esteem and appearance (Garcia, 1998). Self-concept can be defined as a variable that combines global self-esteem, likability, lovability, body functioning, and body appearance (Garcia, 1998). Poor self-concept has been linked to physical anxiety in men and women. The more anxiety people have about their appearance, the lower their self-esteem tends to be. Furthermore, the more physical anxiety people display, the less they consider themselves attractive (Garcia, 1998).

The evaluation of physical appearance was found to be the strongest predictor of self-esteem (Crocker, Luhtanen, Cooper, & Bouvette, 2003). Some have defined self-esteem as an individual evaluation of attributes (Coopersmith, 1967). Others have defined self-esteem as a combination of positive and negative self-evaluations across a variety of domains (Piers & Harris, 1969) as well as a global construct without any specific domains (Rosenberg, 1979). Because of its wide acceptance, the current study used this last definition.

Much research has highlighted a general connection between weight and self-esteem. Overweight adolescents and adults tend to have lower self-esteem than individuals of normal weight (Felker, 1968). However, for boys, the relationship between weight and self-esteem is strongest during middle adolescence, with overweight boys displaying lowest scores of self-esteem (Mendelson & White, 1985). For girls, the relationship between weight and self-esteem is strongest during late adolescence, with overweight girls displaying lowest scores of self-esteem (Mendelson & White, 1985). While self-esteem in adolescents is an important variable, self-esteem in undergraduate college students calls for exploration.

For women, other people often affect their self-esteem. They are more likely to view themselves through the eyes of others and be preoccupied with the tangible attributes of appearance (Crocker, Luhtanen, Cooper, & Bouvrette, 2003). This focus can then lead to decreased self-esteem. Moreover, women view being overweight as socially inappropriate behavior (Mendelson, White, Mendelson, 1996), counter to societal standards of attractiveness, and likely to decrease mate desirability. Women who view their appearance more positively and consistent with societal standards tend to have higher levels of self-esteem (Mendelson, Mendelson, & White, 2001).

There has been a lack of research exploring the self-esteem of men. Past research has instead often combined self-esteem measures in men with measures of body mass and body-esteem (Mendelson, Mendelson, & Andrews, 2000). Body-esteem is a global construct that refers to people's self-evaluation of their physical appearance (Mendelson, Mendelson, & Andrews, 2000). Obese and overweight adults tend to have low body-esteem (Hendry & Gillies, 1978). In addition, women tend to have lower body-esteem than men (Gray, 1977; Mendelson & White, 1985; Mendelson, White, & Mendelson, 1996). Regardless of weight, high body-esteem is related to high self-esteem (Mendelson, Mendelson, & Andrews, 2000).

Contrary to popular beliefs, both men and women experience great pressure to be physically appealing. For men, the ideal body consists of minimal body fat and great muscular development (Olivardia, Pope, Borowiecki, & Cohane, 2004). They tend to believe that their bodies are fatter and further from the ideal than actual measurements indicate, which leads to decreased self-esteem and body-esteem (Olivardia, et al., 2004). Interestingly, when women are asked to rate men's appearance they favored a significantly less muscular and lean body

type. The way men view themselves does not directly relate to the way women perceive them (Olivardia, et al., 2004). This might be caused by the overidealization of the lean figure and the emphasis on weight reduction that has affected men and women in general (Garner, Garfinkel, Schwarts, & Thompson, 1980). However, in spite of this thinness ideal, some evidence suggests that men are conforming to varying social norms. That is, the effects of a person's perceived physical characteristics may depend on the specific social norms that the person is exposed to. For example, an overweight man may regard being overweight more positively, which can lead him to have higher self-esteem (Mendelson, White, & Mendelson, 1996).

Most research, though, provides evidence that both genders, under significant pressure to maintain a perfect body image, experience negative effects on their self-esteem (Mendelson, Mendelson, & Andrews, 2000). The societal pressure to maintain a slender physical appearance has turned body mass index into a standard of achievement (Mendelson, et al., 2000). Body mass index (BMI) is defined as a person's mass in kilograms divided by the square of the person's height in meters. It serves as a measure of leanness; a high score represents a less lean body, and a low score represents a more lean body.

BMI is a strong predictor of self-esteem for adolescent boys and girls, and this relationship does not differ between genders. People with high BMI scores tend to have lower self-esteem scores regardless of their gender (Mendelson, Mendelson, & Andrews, 2000). The relationship between BMI and body-esteem has also been found to be statistically significant (Mendelson, et al., 2000). People with high BMI scores tend to have lower body-esteem, with boys and girls having different ideal BMIs. Boys and girls tend to have higher body-esteem with lower BMI scores, although boys were more satisfied with slightly higher BMI scores (Mendelson, et al., 2000). This relationship does not appear to change over time.

The purpose of the current study is to investigate the connection between self-esteem, body-esteem, and BMI. This study also seeks to provide insight into the relationship between self-esteem and body satisfaction in undergraduate college students. Previous research suggests a relationship between self-esteem and physical appearance satisfaction. The current study will further investigate this relationship, specifically addressing possible gender differences. The first hypothesis is that male college students will have higher self-esteem and body-esteem than female college students. The second hypothesis is that higher scores on self-esteem and body-esteem measures will relate to lower scores on the BMI.

METHOD

Participants

Participants were male (n = 72) and female (n = 81) undergraduate students from a small northeastern college. Participants were sampled through a convenience (n = 18) and a cluster (n = 135) methodology. For the convenience sample, students signed up for the study via an online program and received credit for psychology courses. Students participated in groups with a maximum size of 10 in a college classroom. For the cluster sample all dorm names were placed in a hat, and five dorms were randomly selected. Next, one floor was randomly selected from each dorm. All residents of the selected floor were asked to participate. Those who agreed to participate were given candy as an incentive. Students participated in groups with a maximum size of two in their dorm rooms.

The mean age of the participants was 19.4 years (SD = 0.87). Regarding grade level, 12.4 percent of the participants were freshmen, 64.1 percent were sophomores, 15 percent were juniors, 7.2 percent were seniors and 0.7 percent were graduate students. Regarding ethnicity, 82.4 percent of the participants were Caucasian, 3.3 percent were African American, 7.2 percent were Asian,

3.3 percent were Hispanic, and 3.3 percent self-identified as an "other" category.

Measures

The participants were asked to complete three questionnaires: (1) the Rosenberg Self-Esteem Scale (Rosenberg, 1965), (2) the Body-Esteem Scale for Adolescents and Adults (Mendelson, Mendelson, & White, 1997), and (3) the general questionnaire. Participants were asked to report demographic information as well as weight and height approximations.

The Rosenberg Self-Esteem Scale (Rosenberg, 1965) was initially developed as a self-report measure of self-esteem. Participants are asked to rate a level of agreement with 10 statements on a four-point Likert Scale (1 = *strongly agree,* 2 = *agree,* 3 = *disagree,* and 4 = *strongly disagree*). For this scale, lower scores signify higher self-esteem. Sample questions include, "*As a whole I am satisfied with myself*"; "*I feel that I'm a person of worth, at least on an equal plane with others*" and "*All in all, I am inclined to feel that I am a failure.*" Responses to the scale were averaged for all participants to provide a mean self-esteem score.

The Body-Esteem Scale for Adolescents and Adults (BESAA) (Mendelson, White, & Mendelson, 1997) is a measure of self-perceived appearance. Participants are asked to rate a level of agreement with 23 statements on a five-point Likert Scale (0 = *never,* 1 = *seldom,* 2 = *sometimes,* 3 = *often,* and 4 = *always*). For this scale, higher scores represent a higher body-esteem score. The scale consists of three subscales: body-esteem appearance, body-esteem weight, and body-esteem attribution. Participants are blind to the subscales associated with each statement. BE-appearance highlights participants' general feelings about their physical appearance. Some examples include: "*There are lots of things I'd change about my looks if I could*" and "*I'm pretty happy with the way I look.*" BE-weight targets weight satisfaction. Some examples include: "*Weighing myself depresses me*" and "*I really like what I weigh.*" BE-Attribution targets a participant's opinion on how others evaluate his or her body and appearance. Some examples include, "*Other people consider me good looking*" and "*People my own age like my looks.*" Responses to the scale were averaged for all participants to provide a mean body-esteem score.

The final measure is the general questionnaire. It consists of seven questions that were developed by the current researcher. Participants are asked to disclose their age, gender, ethnicity, major, grade level, weight, and height. Body mass index (BMI) was calculated using the formula (Weight / Height^2) x 703. The BMI scores were then divided into three categories: *underweight/normal, normal/overweight,* and *overweight/obese.* There were 61 participants in the *underweight/normal* category (22 men and 39 women), 47 participants in the *normal/overweight* category (27 men and 20 women), and 45 participants in the *overweight/obese* category (23 men and 22 women).

Procedure

Participants responded to the survey in 20 to 45 minutes. For the convenience sample, participants were invited to complete the survey in a classroom. After all the individuals were seated, the informed consent form was distributed. Once it was completed and collected, the participants were given the Rosenberg Self-Esteem Scale (Rosenberg, 1965). After participants completed the Self-Esteem Scale, they received the BESAA. Finally, participants were asked to complete the general questionnaire. Upon completion of all three questionnaires, all participants were fully debriefed.

For the cluster sample, all participants completed the survey in their dorm rooms. In order to ensure anonymity, since their roommates may have also been completing the survey, participants were instructed to sit at their desks. After they were settled at their desks, the informed consent form was distributed and signed. Next, the survey was distributed, and the researcher waited outside the room for the students to finish. Once the participants had completed all three questionnaires they were asked to place them in

a box to foster confidentiality. Next, participants were fully debriefed. Before leaving, the researcher provided candy to the participants as incentive for their participation.

Results

For the convenience sample, most participants' BMI scores were in the *underweight/normal* category (M = 22.97, *SD* = 4.17). The mean self-esteem score was 1.83 (*SD* = 0.43) and the mean body-esteem score was 2.21 (*SD* = 0.84). In the cluster sample, most participants' BMI scores were in the underweight/normal category as well (M = 23.36, *SD* = 3.64). The mean self-esteem score was 1.87 (*SD* = 0.41) and the mean body-esteem score was 2.34 (*SD* = 0.58).

A series of independent-groups t-tests and a one-way analysis of variance were performed to determine any potential differences between the two samples. Analyses revealed a lack of statistically significant differences between the two samples; therefore, all subsequent analyses include combined data from both samples. Thus, the majority of participants' BMI scores were in the *underweight/normal* category (M = 23.31, *SD* = 3.69). Also, the mean self-esteem score for all participants was 1.86 (*SD* = 0.41) and the mean body-esteem score was 2.33 (*SD* = 0.61).

The first hypothesis that self-esteem and body-esteem would vary by gender was examined through a series of independent-groups t-tests. The first analysis assessed the relationship between gender and self-esteem. Overall, men (M =1.79, SD = 0.38) reported higher self-esteem than women (M = 1.93, SD = 0.43), $t(151) = -2.10$, $p = .038$. The second analysis assessed the relationship between gender and body esteem. Overall, men (M = 2.50, SD = 0.52) reported higher body-esteem than women (M = 2.17, SD = 0.64), $t(151) = 3.40$, $p = .001$.

The second hypothesis that higher self-esteem and body-esteem scores would be associated with BMI scores was examined through a series of one-way analyses of variance with follow-up Tukey tests. The relationship between BMI and self-esteem was statistically significant, $F(2, 150) = 3.96$, $p = .021$, $\eta^2 = .05$. People who had lower BMI scores had higher self-esteem scores. Significant differences were found only between the *underweight/normal* (M = 1.76, SD = 0.40) and the *overweight/obese* categories (M = 1.99, SD = 0.32). People in the *underweight/normal* category had significantly higher self-esteem than those in the *overweight/obese* category.

The relationship between BMI and body-esteem also was significant, $F(2, 150) = 15.64$, $p < .001$, $\eta^2 = .17$. People who had lower BMI scores had higher body-esteem scores. For body-esteem, significant differences were found between all the groups. People in the *underweight/normal* category (M = 2.60, SD = 0.52) had significantly higher body-esteem than those in the *normal/overweight* category (M = 2.29, SD = 0.65) and those in the overweight/obese category (M = 1.99, SD = 0.51). In addition, people in the *normal/overweight* category had higher body-esteem than those in the *overweight/obese* category.

Discussion

In the end, both hypotheses were supported by the findings of the current study. For the first hypothesis, men had significantly higher self-esteem and body-esteem than women. Moreover, lower scores on body mass index (BMI) were associated with higher self-esteem and body-esteem. The significant differences in the three BMI groups (*underweight/normal, normal/overweight,* and *overweight/obese*) in regard to body-esteem level may indicate that the lower people's weights were in relation to their heights, the better they felt about their appearance. The differences in regard to self-esteem may indicate that the lower people's weights were in relation to their heights, the better they felt about themselves. However, the difference was only significant for the *underweight/normal* and *overweight/obese* categories.

The study is consistent with previous findings regarding self-esteem and physical appearance satisfaction. The evaluation of physical appearance is the strongest predictor

of self-esteem (Crocker, Luhtanen, Cooper, & Bouvrette, 2003). Furthermore, satisfaction levels of physical appearance are positively related to self-esteem (Mendelson, Mendelson, & White, 2001). The current study provided more insight into the relationship because it suggested that college men had higher self-esteem than college women.

According to the self-determination theory (Crocker, Luhtanen, Cooper, & Bouvrette, 2003), self-esteem is always higher when it is based on more abstract variables and distinctive features of self than when it is based on tangible attributes such as appearance. The current study indicated that there was a relationship between physical characteristics and self-esteem. Thus, in accord with the self-determination theory it may be hypothesized that participants would have better self-esteem if they based their attitudes on more abstract attributes. In such a case, a significant correlation between BMI and self-esteem would not be expected (Crocker, Luhtanen, Cooper, & Bouvrette, 2003).

The current study found significant differences between all BMI categories for body-esteem. This adds strength to the findings of Mendelson, Mendelson, and Andrews (2000). They asserted that the relationship between BMI and body-esteem was statistically significant. As shown in the current study, the lower a person's BMI was, the higher their score on the BESAA. Since BMI depends on the weight of participants, the relationship shows that the less individuals weigh in relation to their height, the better they felt about their outward appearance.

The current study extended previous findings of gender differences. Mendelson, Mendelson, and Andrews (2000) claimed that there were significant differences between men and women only in regard to body satisfaction. In the current study there was an additional significant difference between men and women in regard to self-esteem. Men were found to have higher self-esteem and body-esteem than women. Thus, men may be more content with themselves both inwardly and outwardly than women. Some of these differences may result from college men engaging in more body-building activities, which increase self-esteem and body-esteem. Thus, college men may suffer less physical dissatisfaction (Olivardia, Pope, Borowiecki, & Cohane, 2004) and therefore may be happier with themselves than women (Morrison, Morrison, Hopkins, & Rowan, 2004).

Moreover, some of the gender differences may be related to eating disorders in women. Hesse (1992) found that among a sample of college women observed from their sophomore to their senior year, eating disorders were significantly more prevalent initially and decreased by senior year. Furthermore, Rosen and Ramirez (1998) found that people with eating disorders find it difficult to accept their bodies. Thus, they are very distressed and have low self-esteem. In one study, women with eating disorders scored much lower on both self-esteem and body-esteem measures than healthy women (Mendelson, McLaren, Gauvin, & Steiger, 2001). As such, there may be a connection between the predominantly sophomore sample of this study and eating disorders acting as an extraneous variable. Future research should investigate the relationship between self-esteem, body-esteem, and BMI, while accounting for the potential influence of eating disorder symptomology in a college population, especially among female sophomore college students.

One limitation of the current study is the use of a non-probability sample. It is possible that only particular people were attracted to sign up for the study through convenience sampling. As the means of the BMI categories showed, the convenience sample consisted mostly of either underweight/normal participants or overweight/obese ones. This may hint that the study attracted two extremes of participants which, given a larger haphazard sample, could have had a larger, perhaps significant effect on the results. However, this was not the case in the current study. The lack of statistically significant differences

between the convenience sample and the cluster sample showed that any potential extraneous variables did not affect the data or skew the results.

Another limitation of the study is that with the cluster sample participants were contacted only once. The researcher picked nights of the week that were generally free of activities with an expectation that students would be in their rooms. Of course, some students were not in their rooms, and those rooms were skipped. Therefore, data was only collected on those students who were not engaged in extracurricular activities and who had enough time to volunteer (assuming their workload was lighter) for the current study. This diminished the randomization of the sample. However, given the high response rate and the lack of statistically significant differences between the samples, the cluster sample was seen as a strength of the study. Furthermore, it enabled the researcher to gather a less biased sample and to sample almost the same number of males and females, which was essential to the study.

Another strength of the current study is the use of the particular measures of this study. Both the Rosenberg Self-Esteem Scale and the Body-Esteem Scale for Adolescents and Adults are frequently used in the literature and have good psychometric value by maintaining high validity and test-retest reliability (Mendelson, White, & Mendelson, 1997). This likely reduced the amount of random error. Finally, the experimenter ensured that the participants were aware of the anonymity of the measures. Emphasis on this aspect decreased self-promotion effects and helped to obtain unbiased data.

The findings of this study indicate that there may be a treatable problem in society. Although people are healthier when they focus on more abstract attributes (Crocker, Luhtanen, Cooper, & Bouvrette, 2003), their self-esteem and body-esteem appear to be influenced by physical characteristics. To lessen this relationship, Cash and Hrabosky (2004) propose that people undergo self-monitoring and psychoeducation courses. These would include completion of daily diaries and evaluation of the effects of body image on the overall quality of a person's life. Moreover, people would be encouraged to think about their emotions and behaviors in a more critical way and to try to evaluate these attributes. Such treatment has been found to be very helpful in inducing an initial change in people's views of their appearance (Cash & Hrabosky, 2004).

The current study can be supplemented by concentrating on the relationships involving different age groups. As proposed by Mendelson, Mendelson, and White (2001) it would be of interest to study older women and the changes that they experience during pregnancy and childbirth. As suggested, women's perceptions of their physical appearance can change either for the better or for the worse after life-altering events, and such a relationship is worth investigating in the future (Mendelson, Mendelson, & White, 2001).

Another interesting area that could be investigated is ethnic differences. Some researchers have highlighted potential differences in ethnic values regarding physical appearance (Croll, Neumark-Sztainer, Story, & Ireland, 2002). Future research should examine a more diverse sample, both in the college community and a community sample. Given different values regarding physical characteristics, it is likely that there may be different pressures on different ethnic groups regarding BMI. As in the current study, these varying perceptions are also likely to influence self-esteem and body esteem.

The present study investigated the differences between men and women in their attitudes about their appearance and found that men generally have higher self-esteem and body-esteem. It also showed that the lower a person's BMI, the higher self-esteem and body-esteem one is likely to have. Such findings suggest that there may be a problem in our society and that sociocultural factors play a role in people's perceptions (Smolak, Levine, & Thompson, 1999). Indeed,

Mendelson, Mendelson, and Andrews's (2000) findings suggest that if people do not base their self-esteem and feelings of self-worth on weight, they may achieve higher self-esteem. All evidence suggests that society must encourage individuals to evaluate themselves based on more meaningful and less superficial criteria.

REFERENCES

Cash, T. F., & Hrabosky, J. I. (2004). The effects of psychoeducation and self-monitoring in a cognitive-behavior program for body-image improvement. *Journal of Treatment and Prevention, 11,* 255–270.

Coopersmith, S. (1967). *The antecedents of self-esteem.* San Francisco, CA: Freeman.

Crocker, J., Luhtanen, R. K., Cooper, L. M., & Bouvrette, A. (2003). Contingencies of self-worth in college students: Theory and measurement. *Journal of Personality and Social Psychology,* 85, 894–908.

Croll, J. K., Neumark-Sztainer, D. S., & Ireland, M. (2002). Prevalence and risk and protective factors related to disordered eating behaviors among adolescents: Relationship to gender and ethnicity. *Journal of Adolescent Health,* 31, 166–175.

Felker, D. W. (1968). Relationship between the self-concept, body build and perception of father's interest in sport in boys. *Research Quarterly,* 39, 513–517.

Garcia, S. D. (1998). Appearance anxiety, health practices, metaperspectives and self-perception of physical. *Journal of Social Behavior & Personality, 13,* 307–409.

Gleason, J. H., Alexander, A. M., & Somers, C. L. (2000). Later adolescents' reactions to three types of childhood teasing: Relations with self-esteem and body image. *Social Behavior & Personality: An International Journal, 28,* 471–481.

Gray, H. (1977). Social aspects of body image: Perception of normalcy of weight and affect of college undergraduates. *Perceptual and Motor Skills, 45,* 1035–1040.

Hendry, L. B., & Gillies, P. (1978). Body type, body esteem, school and leisure: A study of overweight average and underweight adolescents. *Journal of Youth and Adolescence, 7,* 181–195.

Hesse, B. S. (1992). Report on a panel longitudinal study of college women's eating patterns and eating disorders: Noncontinuum versus continuum measures. *Peer Reviewed Journal, 13,* 375–391.

Hewitt, P. L., Flett, G. L., & Ediger, E. (1995). Perfectionism traits and perfectionistic self-presentation in eating disorder attitudes, characteristics, and symptoms. *International Journal of Eating Disorders,* 18, 317–327.

Hooks, B. (1995). Appearance obsession. *Essence,* 26, 69–71.

Mendelson, B. K., McLaren, L., Gauvin, Lise, & Steiger, H. (2002). The relationship of self-esteem and body esteem in women with and without eating disorders. *Peer Reviewed Journal, 31,* 318–323.

Mendelson, B. K., & White, D. R. (1985). Development of self-body-esteem in overweight youngsters. *Developmental Psychology, 21,* 90–96.

Mendelson, B. K., White, D. R., & Mendelson, M. J. (1996). Self-esteem and body esteem: Effects of gender, age, and weight. *Journal of Applied Developmental Psychology,* 17, 321–346.

Mendelson, B. K., White, D. R., & Mendelson, M. J. (1997). Manual for body-esteem scale for adolescents and adults. Montreal: Concordia University.

Mendelson, B. K., Mendelson, M. J., & White, D. R. (2001). Body-esteem scale for adolescents and adults. *Journal of Personality Assessment, 76,* 90–107.

Mendelson, M. J., Mendelson, B.K., & Andrews, J. (2000). Self-esteem, body esteem, and body-mass in late adolescence: Is a competence importance model needed? *Journal of Applied Developmental Psychology, 21,* 249–266.

Mintz, L. B., & Betz, N.E. (1988). Prevalence and correlates of eating disordered behaviors among undergraduate women. *Journal of Counseling Psychology, 35,* 463–471.

Morrison, T. G., Morrison, M. A., Hopkins, C., & Rowan, E. T. (2004). Muscle mania: Development of a new scale examining the drive for muscularity in Canadian males. *Psychology of Men and Masculinity, 5,* 30–39.

Olivardia, R., Pope, H. G., Borowiecki, J. J., & Cohane, G. H. (2004). Biceps and body image: The relationship between muscularity and self-esteem, depression, and eating disorder symptoms. *Psychology of Men and Masculinity, 5,* 112–120.

Piers, E. V., & Harris, D. B. (1969). *Manual for the Piers-Harris Children's Self-Concept Scale.* Los Angeles: Western Psychological Services.

Rosenberg, M. (1979). *Conceiving the Self.* New York: Basic Books.

Smolak, L., Levine, M. P., & Thompson, J. K. (2001). The use of the sociocultural attitudes towards appearance questionnaire with middle school boys and girls. *Peer Reviewed Journal, 29,* 216–223.

Wade, T. J., Shanley, A., & Imm, M. (2004). Second to fourth digit ratios and individual differences in women's self-perceived attractiveness, self-esteem, and body-esteem. *Personality and Individual Differences, 37,* 799–804.

Methods and Applications of Nuclear and Mitochondrial DNA Typing in Forensics

Erika Howard,
The College of New Jersey
Faculty Sponsor:
Professor W.S. Klug,
Department of Biology

Abstract

The polymerase chain reaction-based profiles of known individuals were generated and interpreted according to the New Jersey State Police (NJSP) protocol manual for short tandem repeats DNA typing. DNA extraction, quantification, and amplification procedures were observed and executed. Capillary electrophoresis was used to separate PCR products, and color fluorescence detection was used to generate electropherograms. Specific programs, including GeneScan and Genotyper, were utilized to size the DNA fragments of each sample and assign genotypes. This report contains detailed information about the molecular basis of the DNA typing procedures used at the NJSP forensic laboratory in Hamilton. An investigation of mitochondrial DNA typing in forensics is also included.

Introduction

The use of forensic DNA typing has become one of the most formidable techniques in the criminal justice system. It is one of the most efficient ways to match known as well as unknown suspects of a crime and has led to the conviction of criminals as well as the acquittal of those who are wrongfully accused of committing crimes. An individual's DNA profile is so useful in crime-solving procedures because the number of genotypes present in the population is extremely vast. Because of this, there is a very high probability that the DNA profiles of any two individuals will be different. Thus, the information gathered from one's DNA profile can be used to exclude falsely accused individuals but can also link a suspect to evidence gathered from a crime scene in the event that a match is found between the suspect's known profile and the profile collected from the sample (Charkraborty and Kidd, 1991). DNA profiling requires only a few human cells that can be found in one drop of blood, a few hairs, saliva, skin, or semen in order to provide sufficient genetic information with which to make a comparison.

The method behind forensic DNA typing is a process known as the polymerase chain reaction (PCR). It focuses on amplifying a specific class of DNA markers known as short tandem repeats (STRs). An STR is a specific region of DNA containing multiple copies of short sequences of bases which contain 2–7 base sequences that are repeated a variable number of times. This overall number of pattern repeats at specific loci varies among individuals (Appenzeller, 1990). Today, 13 markers are assessed in order to provide an extremely high level of discrimination between individuals. Some years ago, it was very difficult for scientists to isolate a specific sequence of DNA from all the genes contained in a sample. However, the discovery of PCR has made this task far easier and has become vital to the work of AIDS researchers, evolutionary biologists, and forensic scientists (Appenzeller, 1990).

The science behind the workings of PCR primarily involves the use of a DNA polymerase as well as two primers. DNA polymerase is an enzyme that can form the com-

plementary strand of a sequence of DNA when given an initial strand (Appenzeller, 1990). It is taken from the bacterium *Thermus aquaticus* that usually lives in hot springs ranging from 70°C to 80°C. The enzyme that it contains is heat stable and is not inactivated by the high temperatures needed for strand separation (Marx, 1988). The primers are added to the mixture after the addition of the DNA polymerase and bind to complementary sequences on either side of the targeted sequence. After this has taken place, the mixture is exposed to a series of heating and cooling cycles which last only a couple of minutes. The purpose of the first cycle is to separate the double-stranded DNA into two single strands of DNA. Once the mixture begins to cool, the primers bind to the complementary sequences and the DNA polymerase begins to extend each primer into a new DNA strand. The next cycle of heating separates the copies from the original strands and creates two sets of templates that can now be synthesized. As a result, the target sequence of DNA multiplies exponentially after each round of heating and cooling (Appenzeller, 1990). One of the advantages of using this method is that it can work with both damaged and intact DNA. As a result, older and possibly more degraded DNA can still be analyzed, which has been helpful in several rape and murder cases (Marx, 1988).

Although PCR has become very useful to forensic science, there are concerns regarding possible contamination. The ability of PCR to create trillions of DNA copies from a template sequence makes it possible for product carryover to occur. This happens when the amplification reacts with a previous PCR reaction product or exogenous DNA. Overall, extreme caution must be taken by pre-aliquoting reagents, using designated pipettes, changing tips, and separating the areas where reaction preparation and reaction product analysis take place. Taking such precautions helps to minimize the possibility of contamination. Other procedures that selectively destroy DNA created from a previous PCR also have been devised in order to help eliminate the possibility of product carryover (Elirich, Gelfand, and Sninsky, 1991).

The internship experience at the New Jersey Forensic Science Technology Complex focused largely on the methods and applications of PCR. However, the steps prior to performing DNA amplification were also heavily emphasized. Hence, the ultimate goal of the internship was to be able to perform the four steps needed to generate a PCR-based DNA profile. These four steps included DNA extraction, quantification, amplification, and typing/data analysis. After becoming proficient with these procedures, one aimed to establish a DNA profile for various individuals and compare the results to a known control sheet. The extraction methods used consisted of Chelex blood and saliva extractions, differential extractions, and Qiagen extractions. After extracting the DNA, a procedure known as a quantiblot was used to quantify the amount of DNA so that the necessary amount could be obtained for the amplification step. ABI 310 and 3100 genetic analyzers were then used to generate the final profile of each sample via a process known as capillary electrophoresis.

Although the DNA profiles generated from the DNA typing process successfully matched the known DNA profiles, there are still various concerns and questions regarding the subject of DNA typing. For example, the collection process of samples is extremely vital to the integrity of the results generated from PCR analysis. If the authenticity of such samples cannot be confirmed, then the value of the results is greatly diminished. Many are concerned about the possibility of samples being switched or otherwise tampered with (Lincoln, 1997). In addition, the complexity of DNA typing increases when there is more than a single source contributing to a sample. Results that are deemed inconclusive are produced very frequently in a forensic laboratory setting. Frequent debate has also occurred on the topic of population

statistics and the reference population from which the sample was derived. Lastly, there may be cases where a suspect's profile is found to match a profile generated from a piece of evidence; however, the person is not guilty of committing the alleged crime. The notion that matching one's profile to a piece of evidence automatically means guilt is not always accurate.

Extraction Methods

The first procedures performed were DNA extractions. The purpose of an extraction is to isolate and enrich the DNA for STR testing purposes. Samples obtained from crime scenes typically contain a variety of substances other than DNA which must be separated out before the sample can be examined. The first technique involved a Chelex extraction which can be used on both saliva and blood samples (Walsh, Metzger, and Higuchi, 1991). This particular extraction method was found to be both rapid and as efficient as phenol-chloroform extraction techniques (Walsh, Metzger, and Higuchi, 1991). Vandenberg, van Oorschot, and Mitchell (1997) compared several DNA extraction techniques and also found that the Chelex extraction was the most efficient method to obtain amplifiable DNA in a quick and proficient manner. Chelex is composed of styrene divinylbenzene copolymers and contains paired iminodiacetate ions. These ions act as chelating groups by binding up polyvalent metal ions such as magnesium. This is a very crucial step because the removal of magnesium inactivates nucleases that would otherwise destroy the DNA (Butler, 2001). However, one must make sure that the resin beads in the Chelex solution are distributed evenly before it is added to the sample. This can be done by vortexing the tube containing the Chelex. In bloodstain samples, an extra wash step is needed to help remove any PCR inhibitors such as heme. After an incubation time of two hours the sample is then boiled for eight minutes. This step aids the chelating of ions and also breaks open the cell so that the DNA is released. It is also important to note that the DNA has now been degraded to a single strand form. The last step in this process involves centrifugation of the sample so that the Chelex resin will be pulled to the bottom of the tube (Protocols Manual, 2004).

Although a Chelex extraction is a relatively simple procedure, there are several PCR guidelines that the New Jersey State Police's forensic laboratory follows. For example, it is necessary to perform DNA extractions from samples containing high levels of DNA separately from samples containing low levels of DNA. So, samples of whole blood and single hairs would not be extracted in the same area. This is done to help minimize the possibility of sample-to-sample contamination. A clean cutting surface must also be used for each sample and any scissors or forceps must be cleaned thoroughly with ethanol and water (Protocols Manual, 2004). All work surfaces are cleaned with a 1:10 dilution of household bleach. When dissecting swabs, it should be cut into two or three pieces of equal sizes so that part of the sample can be kept frozen. Futhermore, only one tube can be uncapped at a time and the scissors/forceps must be cleaned before placing another sample in the tube. Lastly, an extraction blank or control must be included for each set of extractions to demonstrate the quality assurance of the procedure (Protocols Manual, 2004).

Another type of DNA extraction method uses FTA paper. The paper contains four chemical substances that serve to protect the DNA from nucleases that would degrade it. The FTA paper works by adding a spot of blood and allowing it to dry. The cells are lysed upon contact with the paper and the DNA from the individual's white blood cells is preserved in the matrix of the paper (Butler, 2001). A small section of the bloodstain is then taken from the paper and placed in a tube for washing. Again, it is necessary to wash the DNA using a solvent so that inhibitors of the PCR reaction such as heme can be removed. Once washed, the paper is ready for the PCR reaction. However, some

laboratories choose to perform a Chelex extraction on the bloodstain from the card (Butler, 2001).

In addition to Chelex extractions, the NJSP forensic laboratory also uses another method known as a Qiagen (QIAamp) extraction to isolate DNA from blood samples. In contrast to a Chelex extraction, a Qiagen extraction uses more reagents. The first step of a Qiagen extraction is to incubate the bloodstain in ATL buffer which contains sodium dodecyl sulphate and will destroy cell membranes and edetic acid (Protocols Manual, 2004). The addition of proteinase K and AL buffer breaks down proteins such as histones. After several incubation periods, ethanol is added to the mixture to precipitate nucleic acids and create the necessary binding conditions. The remainder of the procedure involves the use of spin columns in which the DNA from the sample is adsorbed onto a silica-gel membrane after being centrifuged. After the initial spin, the spin column is washed with AW1 and AW2. The wash buffers remove any residual contaminants without affecting DNA binding. The final wash step is done using AE with an elution buffer. The purpose of this buffer is to elute nucleic acids and to change the pH from acidic to basic in order to elute the DNA (Protocols Manual, 2004). Scherczinger, Bourke, Ladd, and Lee (1997) compared the effectiveness of one particular Qiagen kit to a phenol-chloroform extraction method. Overall, the researchers found that the Qiagen kit produced a higher yield of DNA. In addition, the DNA was found to be of good quality and suitable for amplification procedures. Greenspoon, Scarpetta, Drayton, and Turek (1998) found that the Qiagen extraction is a useful alternative to the Chelex extraction because it may give better results from certain surfaces, including leather and dyed fabrics.

The last type of extraction method that was performed is known as a differential extraction. A differential is primarily used for mixed samples and serves in the separation of spermatozoa and vaginal epithelial cells. This method has been very helpful in sexual assault cases in which there is a mixture of male and female DNA. The procedure begins by breaking open the female epithelial cells. This is usually done by incubating the sample in an SDS or proteinase K mixture. The liquid containing these cells is called the non-sperm cell fraction and is put into a separate tube (Rudin and Inman, 2002). After doing so, a DTT (SDS/proteinase K/dithiothreitol) mixture is used to lyse the sperm nuclei. The DTT is effective in lysing the sperm cells because it is able to break down the protein disulfide bonds that surround the nuclear membranes of the sperm (Butler, 2001). Hence, the end result is a tube containing all or most of the DNA sperm fraction and a tube containing all or most of the non-sperm cells. This particular extraction method was investigated by Yoshida et al. in 1995. The researchers of this study found that this type of procedure was useful in separating mixed samples containing both sperm and vaginal epithelial cells. However, the separation of sperm from epithelial cells is not always complete because of poor samples. If some of the sperm cells have already popped open, their DNA will have been released into the mixture. As a result, the DNA from the sperm cells may appear in the non-sperm cell fraction. It is also possible to have non-sperm cells present in a sperm-cell fraction. In cases such as these, there are specific guidelines that are used in order to interpret a mixed profile (Rudin and Inman, 2002).

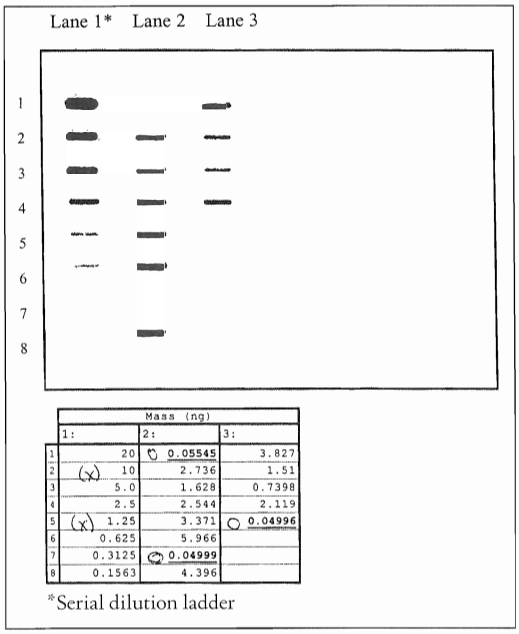
Quantification and Amplification

After the DNA extractions, a process known as DNA quantification was observed and executed. The purpose of quantification is to determine the amount of DNA extracted from a sample and establish whether it is human. This process can be done only after the extraction procedure which isolates the DNA. The ABI Profiler Plus and COfiler STR kits used at the NJSP laboratory specify that 1-2.5 ng of DNA should be used in the

amplification step in order to have successful results. Too much DNA can cause off-scale measurements and too little DNA can result in allele dropout. The method of quantification used at the NJSP laboratory is known as slot-blot quantification (Fig. 1). This procedure involves the use of a probe which hybridizes to the DNA sample and a nylon membrane that captures the genomic DNA (Protocols Manual, 2004).

The first step in slot-blot quantification is putting the extracted DNA in a spotting solution which is a high pH buffer. This serves to denature the DNA and will allow for the visualization of each sample once it is put on the membrane. A slot-blot apparatus is then used to immobilize the single stranded DNA onto the membrane. The DNA is able to stay on the membrane because the membrane has a positive charge which holds the negatively charged DNA (Butler, 2001). Once the DNA is in place, the membrane is washed with hybridization solution and hydrogen peroxide. The hybe solution clears away any heme and prepares the membrane for the addition of the probe. The hydrogen peroxide also cleans up the membrane and oxidizes anything that would interfere with the chemiluminescence reaction (Walsh, Varlaro, and Reynolds, 1992). The probe that hybridizes with the DNA is complementary to a primate specific alpha satellite DNA sequence found on chromosome 17. After the addition of the probe, horseradish peroxidase-strepavidin is added and allows for chemiluminescent detection. Several wash steps are needed in order to remove any excess unbound probe from the membrane. Citrate buffer, which creates an acidic environment reaction, is used to stop the reaction. Once this has been added, oxidation of a luminal-based reagent takes place, resulting in the emission of photons that can be detected on autoradiography film (Walsh, Varlaro, and Reynolds, 1992). The signal intensities of one's samples are always compared against standard samples created through a serial dilution (Fig. 1). The kit used to perform this procedure at the NJSP laboratory is called the QuantiBlot Human DNA Quantification Kit from PE Biosystems (Protocols Manual, 2004).

Figure 1. Slot-blot quantification. Lane 1 corresponds to a serial dilution of 20 ng of DNA which is used as a ladder for comparing unknown quantities of DNA. The signal intensities of the bands found in Lanes 2 and 3 were compared against the standard samples created through the serial dilution in Lane 1. The respective quantities of the DNA samples in each lane can be found in the corresponding table. For example, the quantity of DNA in Lane 2-3 was found to be 1.628 ng. Lane 2-1 typically serves as a negative control for this procedure in which 0 ng of DNA is added.



	Mass (ng)		
	1:	2:	3:
1	20	0.05545	3.827
2	10	2.736	1.51
3	5.0	1.628	0.7398
4	2.5	2.544	2.119
5	1.25	3.371	0.04996
6	0.625	5.966	
7	0.3125	0.04999	
8	0.1563	4.396	

*Serial dilution ladder

Once the DNA samples have been quantified, they must be prepared for the PCR reaction. Based on the quantification results, the necessary volume that must be aliquotted from each sample in order to achieve the target amount of DNA (1 ng) can be determined. In some instances it may be necessary to concentrate or dilute the amount of DNA when there is too little or too much present in the given sample. The pre-amplification mixture also consists of primers and *Taq* polymerase, which serve as the key components of the PCR reaction. In addition to these components, the pre-amplification mixture also contains building blocks composed of the four nucleotides. Deionized water also is added to the mixture to achieve the necessary

volume and concentration of each of the components. The purpose of the primers is to flank each side of the target DNA sequence that is to be copied and the purpose of *Taq* polymerase is to arrange the nucleotide bases complementary to the template sequence (Butler, 2001).

Various controls are used to monitor the integrity of the reaction and the scientist—typically, a negative and a positive control. The negative control includes every component of the pre-amplification reaction mixture with the exception of the DNA sample. The main purpose of the negative control blank is to determine whether any contamination has occurred. A positive control contains a known sample of DNA and is used to determine the success or failure of the reaction components. In addition to these controls, an extraction blank is used to show that reagents used in the extraction procedure are free of any contaminant DNA (Butler, 2001).

In addition to the specific PCR guidelines for DNA extraction, the NJSP laboratory follows established guidelines for DNA amplification. For instance, it is necessary for the pre-amplification mixture to be made within a self-contained hood. Clean work areas covered with disposable bench paper are needed to prepare the reaction mixture. It is also necessary to add the DNA to the PCR reaction mixture last. Microcentrifuge tube racks that are specific for PCR setup and have been autoclaved must be used. It also is preferable to use disposable microcentrifuge racks to transport the amplification tubes to the thermal cycler in the designated amplification area. Lastly, amplified DNA can only be stored in the designated amplification area (Protocols Manual, 2004).

At the NJSP laboratory, PCR is used to amplify a series of STRs. Again, STR markers are highly variable among individuals and are thus useful for identification purposes. The effectiveness of PCR in forensic casework has been thoroughly researched. For example, in 1993, Schneider and Rittner analyzed the reliability of PCR, which was used to amplify the HLA-DQ alpha locus from 42 forensic cases. Overall, 125 samples were used in the study and PCR was effective in 70% of them. The substances that were typed included blood and semen stains, hair, fingernails, and mixed body fluids. In order to have unity within the forensic DNA typing community, a common set of STR loci developed by PE Applied Biosystems and Promega Corporation is used. The 13 loci are currently included in the national DNA database known as CODIS (Combined DNA Index System). These loci include CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, and D21S11 (Collins et al., 2004). The probability of a random match between unrelated individuals is rarer than one in one trillion when all 13 loci are tested for (Butler, 2001). Two kits known as Profiler Plus and COfiler are used to amplify all the necessary loci. Buse et al. (2003) evaluated the overall effectiveness of the Profiler Plus and COfiler typing systems and found that they produced reliable results. The primer sets were found to be human specific and the DNA samples used in this study were effectively amplified. As a control mechanism, both kits contain two overlapping loci to make sure that the same results are produced. Each kit also contains an allelic ladder to allow for accurate genotyping. It is also important that the right blend of oligonucleotide primers is present in each kit. Recent research conducted by Butler et al. (2001) has indicated that high-performance liquid chromatography (HPLC) and time-of-flight mass spectrometry (TOF-MS) can be used to characterize the specific sequences of the primers in commercial STR kits so that their use can be monitored. Hence, possible alteration of the primer sequences used in such kits and the subsequent effect on amplification performance would be identifiable.

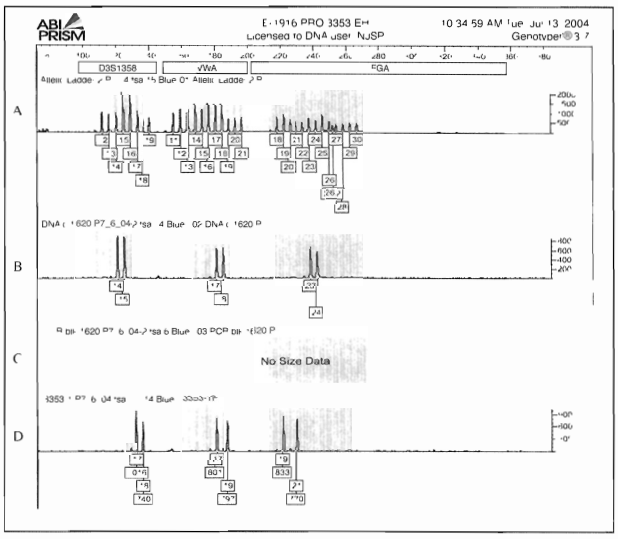
DATA ANALYSIS

A process known as electrophoresis is needed to separate products of PCR in order to decipher one allele from another. The theory behind electrophoresis is that DNA which is negatively charged will move toward a positive electrode when placed in an electrical field. The NJSP laboratory uses a relatively new technique known as capillary electrophoresis, which is much more efficient than other electrophoresis methods (Tagliaro and Smith, 1996). Injection, separation, and detection processes are done automatically and be can unattended. The primary components of a capillary electrophoresis instrument include a capillary, two buffer vials, and two electrodes which are connected to a high-voltage power supply. In addition to this, such instruments are also equipped with a fluorescence detector, tube holders, and a computer to control various actions. The actual separation of DNA is done using a polymer solution that the DNA must pass through once the electrical field has been initiated (Butler, Buel, Crivellente, and McCord, 2004). Smaller molecules can migrate through the gel faster than larger ones. The specific instruments used at the NJSP laboratory include the ABI Prism 310 Genetic Analyzer and the ABI Prism 3100 Genetic Analyzer.

The genetic analyzers at the NJSP laboratory are equipped with multiple color fluorescence detectors to detect STR alleles. This particular process involves detecting light from dye molecules after they have been excited. Fluorophores, which are molecules capable of fluorescence, can be recorded separately from each other. This separation can be done via the use of optical fibers and a fluorophore separation algorithm known as a matrix. This system permits scientists to distinguish various components in a mixture from one another (Butler, 2001). PCR products are fluorescently labeled by attaching the dye to one of the primers. The signals emitted by the fluorophores are amplified and converted to electronic form. They represent the peaks seen in an electropherogram (Fig. 2). The dyes used are provided by ABI and include 5-FAM, JOE, NED, and ROX.

After running samples on the ABI 310 and ABI 3100 genetic analyzers, the data were collected and interpreted. The data are depicted in the form of peaks which represent various STR alleles that have been amplified (Fig. 2). An internal allelic ladder is used to determine the number of repeats of each locus that was amplified during PCR (Fig. 2-A). Thus, a genotype is determined by the comparison of an unknown sample to the known allelic ladder. The first step in interpreting the data involves the use of genotyping software. GeneScan software is used to size the DNA fragments of each sample. The information is then sent to a program called Genotyper which is responsible for determining the genotype of each sample. Buse et al. (2003) demonstrated that Genotyper software was able to make accurate allele calls which may include DNA fragments that differ in length by one base pair. Once all the alleles of each peak have been called, one can begin reviewing the data and making edits if necessary.

Figure 2. Electropherogram data showing allelic peaks at the D3S1358, vWA, and FGA loci. A. Allelic ladder indicating the number of possible base sequence repeats for each locus analyzed. B. Results for a known DNA control sample. This individual is heterozygous at the D3S1358, vWA, and FGA loci. For example, the data indicates that this individual has one allele containing 14 base repeats and another allele containing 15 base repeats at the D3S1358 locus, where one allele is inherited from each parent. C. Results for a PCR blank containing no amplified DNA. This sample serves as the negative control and indicates that no contaminant DNA is present. D. Results for a DNA sample obtained from one of the forensic scientists at the NJSP laboratory. The results were found to match those that were previously on file.



Overall, data for a total of nine extractions were obtained from the genetic analyzers. Eight of the extractions consisted of known samples that were obtained from individuals working at the lab (Fig. 2-D). The other extraction consisted of samples from five family members. After printing the necessary electropherograms, an accompanying case summary sheet was filled out. Each set of data was then reviewed by one of the forensic supervisors. All results were compared against a known control sheet and were found to match.

Significance of Results

In forensic cases, an important matter is to determine the significance of the association between the biological evidence and the reference donor. The ultimate question is to determine what the probability of a random match is. This is usually determined by calculating a profile frequency, which is the number of times the profile is seen in some reference population. In order to do this, a representative number of people from a reference population are tested and the number of times each genotype occurs is recorded (Rudin and Inman, 2002). However, the amount of genetic variability at a particular STR marker can be vast depending on the number of possible alleles. If there are n alleles, there are n homozygous genotypes and $n(n\text{-}1)/2$ heterozygous ones. So, a locus with 12 possible alleles would exhibit $12 + (12 \times 11)/2 = 72$ genotypes (Butler, 2001). Since typing enough people to discover how often each of these types occurs at different loci would be an impossible task, forensic scientists use population genetics theory to estimate frequencies. Two principles known as the Hardy-Weinberg equilibrium (H-W) and the linkage equilibrium (LE) are used to estimate genotypes based on individual alleles rather than observed genotype frequencies. According to the Hardy-Weinberg model, "There is a predictable relationship between allele frequencies and genotype frequencies at a single locus" (Rudin and Inman). Based on this mathematical relationship, genotype frequencies in a population can be estimated without having been seen in an actual survey. Linkage equilibrium is defined as the "steady-state condition of a population where the frequency of any multi-locus genotypic frequency is the product of each separate locus" (Rudin and Inman, 2001). Thus, a DNA profile can be estimated over several loci whether or not it has been seen in an actual population survey.

In general, population data are collected for different population groups, which usually include Caucasian, African American, Hispanic, and Asian. Allele frequencies at each locus of interest are calculated and results are tested for departure from Hardy-Weinberg and linkage equilibrium. If the magnitude of departure is too great, then a decision might be made to reject the frequency data. However, if the data can be used, then the allele frequencies will be used to calculate the frequency of a genotype at any one locus using Hardy-Weinberg proportions. The product rule is then used to determine the frequency of a complete DNA profile. This is done by multiplying the frequencies of each locus together and expressing the calculation as a percentage (Rudin and Inman, 2001). For example, if five independent loci with genotype frequencies of 0.05, 0.1, 0.15, 0.3, and 0.4 are tested, the probability of a match is $(0.05)(0.1)(0.15)(0.3)(0.4) = 0.00009$ (Kirby, 1990). Thus, the more DNA markers examined and compared, the greater the chance that two unrelated individuals will have different genotypes. The extremely small probabilities of a random match serve as compelling evidence in a court of law because the calculations are mathematically irrefutable.

Investigation of Mitochondrial DNA Typing

Although the NJSP forensic laboratory does most of its DNA profiling using STR typing, the lab was recently equipped with the technology to perform mitochondrial DNA analysis. Because of its resistance to extreme environmental conditions, the use of

mitochondrial DNA in forensic case work has become extremely valuable in instances where the sample DNA is degraded. Mitochondrial DNA (mtDNA) differs from nuclear DNA in that it is only maternally inherited. Hence, mtDNA does not undergo recombination and therefore represents only the maternal ancestry of the individual (Butler and Levin, 1998). The mitochondrial genome contains a noncoding region known as the control region which is made up of approximately 1,100 base pairs. mtDNA has proven to be useful in forensic typing because each cell can contain up to 1,000 copies of it. Hence, it is easier to recover and sequence than the single copy of nuclear DNA that exists in each cell. This is especially true when the sample DNA being sequenced is either small or degraded. mtDNA also has a much higher mutation rate than nuclear DNA because of less efficient repair mechanisms and polymerases. As a result of this, some regions of the mtDNA are of interest because of their hypervariability between individuals. Most of the sequence variation between individuals is found in two specific areas of the control region, which are referred to as hypervariable region 1 (HV1) and hypervariable region 2 (HV2) (Budowle, Allard, Wilson, and Chakraborty, 2003).

Research regarding the specific techniques used to amplify and identify human mtDNA efficiently is currently being investigated. Matsuda et al. (2004) designed a new primer pair for use in identifying human DNA via PCR. Another study conducted by Chong et al. (2004) investigated the use of a duplex primer set for amplification of the HV1 and HV2 control regions. Unlike singleplex amplifications, duplex amplifications involve the co-amplification of the HV1 and HV2 regions rather than two separate amplifications. Thus, these regions are amplified using a specific primer blend and mtDNA reaction mix. This method seems to be more useful than singleplex amplification because it requires fewer pipetting steps. Because of this, there is less risk for contamination and human error. In addition, this method may be more convenient in cases where a DNA sample is limited because less product is required to amplify both of the hypervariable regions (Chong et al., 2004).

Although there has been a great deal of improvement in the amplification process and discriminatory power of mtDNA analysis, there are still several technical challenges and genetic issues. One major concern is a phenomenon known as heteroplasmy, in which more than one mtDNA type can be found in a single individual. The two kinds of heteroplasmy include length and point substitution (Budowle et al., 2003). In point heteroplasmy, the mtDNA types of an individual usually differ at a single base. Although the difference can occur at more than one site, it is extremely rare. There are three possible scenarios in which an individual can display heteroplasmy. First, an individual might exhibit more than one mtDNA type in a single tissue. Second, an individual could display differing mtDNA types in different tissues. Third, an individual might be heteroplasmic in one tissue and homoplasmic in another (Budowle et al., 2003).

Although heteroplasmy is a serious concern to the forensic community, recent research has indicated that new techniques are available to help solve the problem of identifying heteroplasmy at all levels (Young et al., 2004). Indeed, knowledge that is gained about the heteroplasmic nature of mtDNA might become an asset in improving the discriminatory power of mtDNA typing. At present, mtDNA typing can serve as another means of possible identification when nuclear DNA typing fails. The fact that mtDNA typing can be successful when sample DNA is highly degraded or old holds a great deal of promise for forensic casework that could not be solved using nuclear DNA typing. Overall, the NJSP forensic laboratory is on the cutting edge of DNA technology profiling. Both the nuclear and mitochondrial DNA sections of the lab will soon be

working in a cooperative effort to solve a variety of cases.

ACKNOWLEDGMENTS

The nuclear DNA-typing methods described in this report were completed at the New Jersey Forensic Science Technology Complex at Hamilton during the summer of 2004. Many individuals contributed to the success of this internship program as well as the completion of this paper. First, I would like to thank Dr. Tom Brettell for giving me the opportunity to intern at the laboratory. I would also like to thank Ed Larue and Lyn McBride for directly supervising our efforts to become acquainted with each of the DNA typing procedures and data analysis methodologies. In addition, I would like to recognize all the forensic scientists in the DNA unit who allowed me to observe their performance techniques and interpretation of the protocol manual. They provided a friendly and encouraging environment to which the success of the pilot forensic internship program can be directly attributed. Lastly, I would like to acknowledge Professor W. S. Klug's efforts in coordinating the Biology 397 internship credit course and the mentoring of this paper.

REFERENCES

Appenzeller, T. (1990). Democratizing the DNA sequence. *Science*, 247, 1030–1032.

Budowle, B., Allard, M.W., Wilson, M.R., and Chakraborty, R. (2003). Forensic and mitochondrial DNA: Applications, debates, and foundations. *Annu. Rev. Genomics Hum. Genet.*, 4, 119–141.

Butler, J.M. (2001). *Forensic DNA typing: Biology and technology behind STR markers.* San Diego, CA: Academic Press.

Butler, J.M., Buel, E., Crivellente, F., and McCord, B.R. (2004). Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. *Electrophoresis,* 25, 1397–1412.

Butler, J.M, Devaney, J.M., Marino, M.A., and Vallone, P.M. (2001). Quality control of PCR primers used in multiplex STR amplification reactions. *Forensic Science International,* 119, 87–96.

Butler, J.M., and Levin, B.C. (1998). Forensic applications of mitochondrial DNA. *Tibtech,* 16, 158–162.

Buse, E.L., Putinier, J.C., Hong, M.M., Yap, A.E., and Hartmann, J.M. (2003). Performance evaluation of two multiplexes used in fluorescent short tandem repeat DNA analysis. *The Journal of Forensic Sciences,* 48, 348–357.

Chakraborty, R. and Kidd, K.K. (1991). The utility of DNA typing in forensic work. *Science,* 254, 1735–1739.

Chong, M.D., Calloway, C.D., Klein, S.B., Orrego, C., and Buoncristiani, M.R. (2004). Optimization of a duplex amplification and sequencing strategy for the HV1/HV2 regions of human mitochondrial DNA for forensic casework. *Forensic Science International,* 30, xxx–xxxv.

Collins, P.J., Hennessy, L.K., Leibelt, C.S., Roby, R.K., Reeder, D.J., and Foxall, P.A. (2004). Developmental validation of a single-tube amplification of the 13 CODIS STR loci, D2S1338, D19S433, and amelogenin: The AmpFlSTR Identifiler PCR Amplification Kit. *The Journal of Forensic Sciences,* 49, 1265–1277.

Erlich, H.A., Gelfand, D., and Sninksy, J.J. (1991). Recent advances in the polymerase chain reaction. *Science,* 252, 1643–1650.

Greenspoon, S.A., Scarpetta, M.A., Drayton, M.L., and Turek, S.A. (1998). QIAamp spin columns as a method of DNA isolation for forensic casework. *The Journal of Forensic Sciences,* 43, 1024–1030.

Kirby, L.T. (1990). *DNA Fingerprinting.* New York, NY: Stockton Press.

Lincoln, P.J. (1997). Criticisms and concerns regarding DNA profiling. *Forensic Science International,* 88, 23–31.

Marx, J.L. (1988). Multiplying by leaps and bounds. *Science,* 240, 1408–1410.

Matsuda, H., Seo, Y., Kakizaki, E., Kozawa, S., Muraoka, E., and Yukawa, N. (2004). Identification of DNA of human origin based on amplification of human-specific mitochondrial cytochrome b region. *Forensic Science International,* xxx, xxx-xxx.

New Jersey State Police Office of Forensic Sciences. (2004). STR Protocols Manual.

Rudin, N. and Inman, K. (2002). *An Introduction to Forensic DNA Analysis.* Boca Raton, Fla.: CRC Press.

Scherczinger, C.A., Bourke, M.T., Ladd,C., and Lee, H.C. (1997). DNA extraction from liquid blood using QIAamp. *The Journal of Forensic Sciences,* 42, 893–896.

Schneider, P.M., and Rittner, C. (1993). Experience with the PCR-based HLA-DQ alpha DNA typing system in routine forensic casework. *Int J Legal Med,* 105, 295–299.

Tagliaro, F., and Smith, F.P. (1996). Forensic capillary electrophoresis. *TrAC Trends in Analytical Chemistry,* 15, 513–525.

Vandenberg, N., van Oorschot, R.A., and Mitchell, R.J. (1997). An evaluation of selected DNA extraction strategies for short tandem repeat typing. *Electrophoresis,* 18, 1624–1626.

Walsh, P.S., Metzger, D.A., and Higuchi, R. (1991). Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques,* 10, 506–513.

Walsh, P.S., Varlaro, J., and Reynolds, R. (1992). A rapid chemiluminescent method for quantitation of human DNA. *Nucleic Acids Research,* 20, 5061–5065.

Yoshida, K., Sekiguchi, K., Mizuno, N., Kasai, K., Sakai, I., Sato, H., and Seta, S. (1995). The modified method of two-step differential extraction of sperm and vaginal epithelial cell DNA from vaginal fluid mixed with semen. *Forensic Sci Int,* 50, 25-33.

Young Lee, H., Chung, C., Yoo, J., Jin Park, M., and Shin, K. (2004). Quantitative and qualitative profiling of mitochondrial DNA length heteroplasmy. *Electrophoresis,* 25, 28-34.

T-Cell Exacerbation of Spinal Cord Injury through Regulation of DNA-Binding Proteins

Mark Strohmaier,
The College of New Jersey
Faculty Sponsor:
Professor W.S. Klug,
Department of Biology

Abstract

The immune system is known to have effects on the actions of nearly all parts of the body. However, the exact nature of these effects is, at least for the spinal cord, mostly unknown, though some studies have recently been done to look at relationships between the immune system and neurological damage in rodents. It has been hypothesized that the actions of a portion of the immune system, namely T-cells, are detrimental when it comes to recovery from spinal cord injury (Dr. Phillip Popovich, pers. comm.). The current work has focused on analyzing previously generated microarray data in an attempt to determine whether this was the case. While it is not possible at this time to determine whether T-cells are actually detrimental to the recovery from spinal cord injury, current findings demonstrate that there are significant differences in gene expression as a result of the presence or absence of T-cells at the site of spinal cord injury.

Introduction

Early experiments looking at the interaction of T-cells and spinal cord injury have focused on a mutant strain of rats known as the nude rat. In addition to having no fur, the nude rat also lacks the genes necessary to produce T-cells, as their thymus fails to develop. Thus, nude rats are the perfect mechanism for studying effects of T-cells on spinal cord injury. Lacking an effective immune system, nude rats can be compared to rats with normal immune systems, and any differences in their recovery from induced spinal cord injury can be attributed to the lack of T-cells in the nude rats.

Previous studies have shown that T-cells can have neuroprotective effects on injury recovery. One such study was Moalem et al.[1], which showed that autoimmune T-cells increased the rate of functional retinal ganglion cell survival over 300% after injury to optic nerves in rats. However, unpublished studies examining the scores of nude rats (*NU/NU*) and control rats (*NU/+*) on open field locomotor tests following spinal cord injury have a different result (Popovich, pers. comm.). These studies have shown that the nude rats recover better on the tests in the first 10 days post injury (Figure 1), which is the time the largest concentration of T-cells is present at the site of injury (Figure 2).

Figure 1. Open field locomotor scores for (Nu/Nu) versus (Nu/+) rats for 50 days post injury.

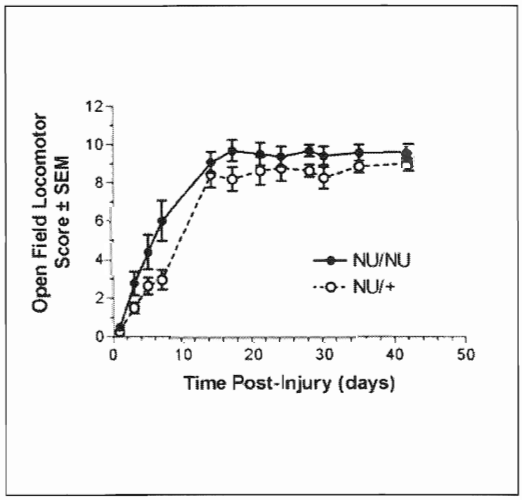
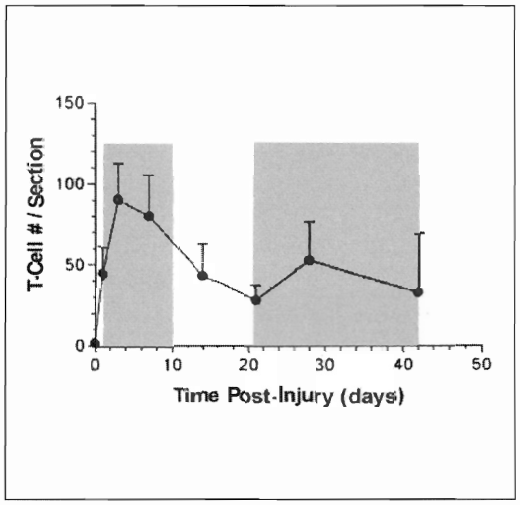


Figure 2. T-cell concentration for the same rats as in Figure 1, versus days post injury.



Because of this apparent contradiction in the data, an attempt is underway to determine whether the protective or harmful effects of T-cells are more predominant in cases of damage to the nervous system. The hypothesis is that the pathological effects outweigh the neuroprotective effects and the presence of T-cells serves to exacerbate spinal cord injury.

The experiment performed involved a total of 32 rats—16 homozygous nude rats (*NU/NU*), and 16 heterozygous control rats (*NU/+*). Each rat was given a uniform spinal cord injury using an electromagnetic spinal contusion device developed at Ohio State University.[2-4] Four rats from each group were terminated at each of four time points—immediately after injury, one day after injury, three days after injury, and seven days after injury—and their spinal cords were removed.

The tissue samples were sent to the Hart Laboratory at the W. M. Keck Center for Collaborative Neuroscience and had RNA extracted from them using a Qiagen RNeasry protocol, described in Carmel et al.[5] The RNA was reverse transcribed, and hybridization targets were prepared using Genisphere 3DNA dendimer systems. The resultant products were automatically hybridized with microarray chips that had probes containing 7793 rat genes spotted on them, using a Ventana Discovery System, following a protocol developed by the Hart lab.5 (Each chip was hybridized with a sample of pooled uninjured spinal cord as control and one sample.) The current investigation focused on analyzing the data of the microarrays, and then confirming the data for genes of interest using Real Time PCR.

Methodology

Filtering by P-Value

The first step performed in the analysis of the microarray data was filtering out significantly different genes. T-tests were performed comparing the average expression level of each gene in the injured sample to the average expression of that gene in the uninjured control. A gene was marked as significantly different if it had a *p*-value of .001 or less at four out of eight data points (*NU/NU* and *NU/+* at t=0, 1, 3, or 7). While in most experiments *p*-values of .01 or .05 are considered significant, a *p*-value of .001 was chosen to minimize the possibility of false discovery. False discovery occurs when a sample (or in this case a gene) is chosen as significantly different by chance, when it is in fact not different. A *p*-value of .05 will have a false discovery rate of about 5/100, while changing the *p*-value to .001 means approximately only 1/1000 genes will show up via false discovery. Once the genes with *p*-values greater than .001 were removed, the sample size was reduced from 7,793 unique genes to a total of 494 genes.

Principal Components Analysis

The second step performed was principal components analysis (PCA). The purpose of PCA is to reduce the number of dimensions of the data by looking at underlying statistical patterns. In this case, there are eight dimensions for the data (from four time points, and two samples per time point). By locating the principal component vectors, the number of dimensions of the data can be reduced to two or three dimensions that will account for most of the variability. The scores for each gene are then

converted from the expression level at each time point to a score of how closely they match each of these underlying vectors.

Principal components analysis allows true similarity between genes to be recognized more easily. When looking at the normalized gene expression levels, two genes may have similar expression levels at one time point, but be completely different at all the other time points. By contrast, in this study the first score on the first PCA vector (the one that holds most of the variability) accounted for 91 percent of the variability, and the first two vectors combined accounted for 96 percent of the variability. Thus, if there were two genes that had similar scores on both of the first two vectors, then they most likely are very similar to each other in their overall expression pattern.

K-Means Clustering

Clustering is a statistical tool designed to place the items in question in groups where the group members are more closely related to each other than to the elements outside the group. One common clustering algorithm is called K-means. The basic steps of the K-means algorithm are as follows: The researcher specifies a number of clusters into which the data is to be broken. One element is randomly chosen as starting center of each of the clusters. Each additional data point is then placed in the cluster whose center is closest to the data point's value. Then the new centers of the clusters are calculated, and the clusters are re-evaluated. This process repeats for a user-specified number of cycles, or until the clusters do not change.

For example, if the data set is 1.1, 1.3, 1.5, 2.5, 2.7, 2.8 and two different clusters are to be made, the first two elements would be randomly selected (say 1.5, 2.5). Then the elements would be placed in a group with the center they would be closest to—1.1, 1.3, 1.5 and 2.5, 2.7, 2.8. Now the new centers are calculated (note that the centers do not need to be an actual element)—1.3, 2.67. The elements are then placed in their new clusters—1.1, 1.3, 1.5 and 2.5, 2.7, 2.8. Since the clusters did not change, the algorithm is completed.

Clustering Consistency

Because of the randomness inherent in the first stage of K-means clustering, the possibility exists that different runs of the algorithm could produce different cluster sets. If this was the case, then the clusters found would be invalid. To test whether the K-means algorithm was consistent, 20 different clusterings of a set of genes were computed using the statistical analysis package R. The resultant clusterings were checked for consistency using a program written in the programming language Java. Cluster consistency was checked chiefly by looking for similar patterns in the data, as the actual clusters to which the genes are assigned do not matter, as long as the same genes show up together. This was done by generating all possible clusterings and then comparing them.

For example, consider a set of 20 genes. In the first clustering, genes 1–5 and 18–20 are in cluster 1; genes 6–7 and genes 11–14 in cluster 2; and genes 8–10 and 15–17 in cluster 3.

In the second clustering, genes 8–10 and 15–17 are in cluster 1; genes 1–5 and 18–20, in cluster 2; and genes 6–7 and 11–14, in cluster 3. These two clusterings are equivalent.

Once a complete set of n! mappings was generated (where n is the number of clusters), it was then possible to calculate the similarities between two cluster sets. The program calculated the cluster set resulting from the map, and then scanned through the list of clusters, comparing the mapped value to the value in the second set. If the numbers were not the same, the number of differences found would be incremented. If a map with no differences was found, the program would terminate. Otherwise, the program would continue to cycle until all possible maps were tried. If all maps were tried, the smallest number of differences was returned, so that a percentage similarity between the clusters could be calculated.

In the tests conducted on the 20 K-means sets, two main cluster groups were found. These cluster groups were relatively similar, with 72 percent of the genes showing up in the same clusters. But even more encouraging was the

result that each of the 20 cluster sets showed up in one of the two groups. Both of these results suggest that K-means clustering is relatively consistent.

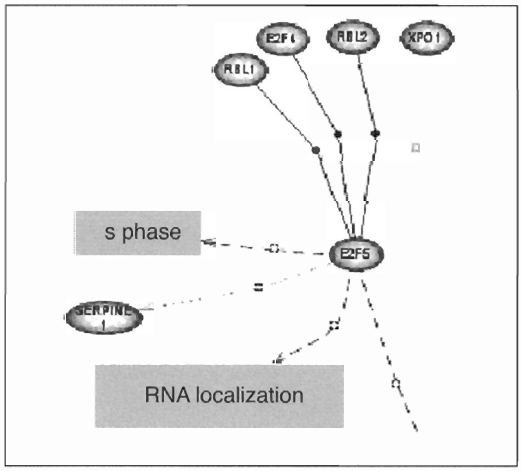
Pathway Assist

Having clustered the genes, the next step was to select a particular subset of genes to examine closely. Genes coding for DNA binding proteins were chosen as the main area of focus, for several reasons. First, it was decided that choosing genes in one general functional category would be useful, as their effects would be similar. Additionally, by choosing DNA binding genes, studies of proteins produced by the genes would be simplified, as they would tend to be located mainly in nuclei.

The relevant genes extracted via *P*-value filtering and clustering were input into a program called Pathway Assist, developed by Stratagene. Pathway Assist is, at its core, a natural language processing tool, which searches through online databases and catalogs for all known information about particular proteins. This information is displayed graphically, telling the user which biological processes and molecules affect or are affected by the protein(s) in question (Figure 3).

Using Pathway Assist, the proteins produced by the genes in question were studied, and the genes with the most known information about them were chosen for Real Time PCR to confirm their microarray results.

Figure 3. Illustration of Pathway Assist program applied to the protein E2F5.



Real Time Polymerase Chain Reaction

Real Time Polymerase Chain Reaction (RT-PCR) is a method of amplifying RNA and quantitatively measuring the amount that is produced. It can be used to measure gene expression levels and can also be useful for confirming microarray results. The general RT-PCR protocol used involves three main steps. First, cDNA is created from the sample RNA under investigation. Then, the cDNA is combined with primers for the gene(s) in question and with a light-reactive dye. Then the combination is placed in a Real Time PCR machine, which hybridizes the primers and cDNA and measures the amount of RNA that is produced. By comparing the amount for the normal and nude rats, another measure of the change in gene expression can be produced, thus providing a way of confirming or denying the microarray results for the same gene.

The RT-PCR assay was used to measure the expression level of the genes coding for the DNA-binding proteins in question. The assay was most successful for two genes—the gene coding for the protein E2F5, and the gene coding for IRF1.

Results

Of the isolated genes, the RT-PCR assay produced sufficient data to confirm the microarray expression results for the genes coding for the proteins E2F5 and IRF1. For E2F5 (Figure 4), RT-PCR data was obtained only for one time point—nude rat RNA one day after injury. However, this result still seems to support the microarray data, as the nude rat day 1 microarray score was more than double that of any other point, and all other RT-PCR samples failed to produce enough RNA to be accurately measured.

The RT-PCR run for IRF1, however, had all samples react measurably. In this case, the microarray data (Figure 5) was not totally confirmed, as the RT-PCR data (Table I) did not totally agree with the microarray data for day 0 (the square). However, the RNA samples that corresponded to days 1 and 3 did agree

with the microarray data, suggesting that, for the most part, the data is correct.

Figure 4. Microarray data for E2F5, with the RT-PCR confirmed point marked as x.

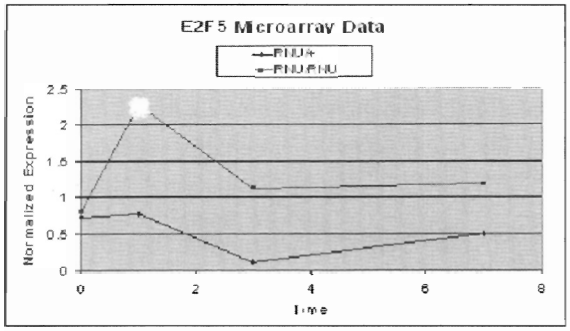


Figure 5. Microarray data for IRF1, with the RT-PCR confirmed points marked with x's, non-confirmed points marked with squares.

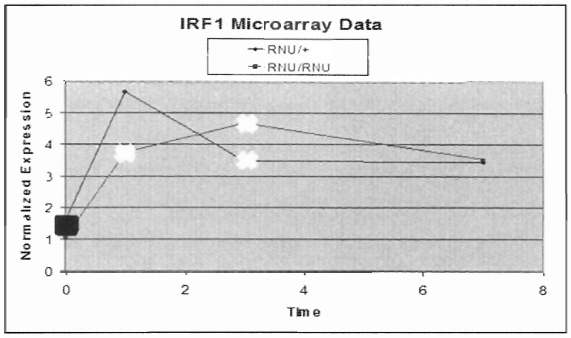


Table I.

Day	RT-PCR Data	
0	1	1
1	3.92	0
3	4.36	7.24

RT-PCR results, scaled to have a time zero value of 1.

Discussion

Current Directions in Spinal Cord Injury Treatment

Traditionally, spinal cord injury treatment research has focused on trying to repair the injury artificially through stimulation, transplanting cells to repair the site of injury, or using physical therapy to help the victim adjust to the effects of the injury. The effectiveness of physical therapy largely depends on the extent of the injury and the overall health of the person. The effectiveness of the other treatments varies from situation to situation, though there has been a general trend toward more effective treatments, particularly through recent studies using neuronal precursor cells in treatment[8-10] and attempts to help the surviving neurons overcome the glial scar resulting from the injury.[11,12]

Moreover, there has been pioneering work studying immune system effects on recovery from nerve damage, including spinal cord injury. A 1997 article by Popovich, Wei, and Stokes observed differences in the inflammatory response following spinal cord injury and the recovery rates in two different rat variants (Sprague-Dawley and Lewis).[13] Moalem et al., in 1999, conducted a study which showed that differing immune system responses in the central and peripheral nervous systems led to variant rates of recovery and markedly different injury sites after the injury,[14] in addition to the previously mentioned study demonstrating that T-cells greatly improved the survival of retinal ganglion cells after optic nerve injury.[1] Hauben et al., in a September 2000 article in the *Journal of Neuroscience*,[15] seemed to verify the findings from Moalem's optic nerve study, demonstrating that the presence of T-cells, combined with manipulation of Myelin Basic Protein, can promote recovery from spinal cord contusion damage. A second paper by Hauben et al., a year later, demonstrated a similar effect, this time showing that the inhibition of the myelin-associated protein Nogo-A produces a T-cell response which aids in recovery from injury.[16] However, the recent unpublished studies cited above seem to indicate that T-cells may have a detrimental effect on recovery from spinal cord injury in rats, all other factors being equal.

Significance of this Study

As previously mentioned, the analysis done in this study was mainly focused on identifying significant differences in gene expression between the (*NU/NU*) and (*NU/+*) rats. A large number of variant genes was discovered, as the *p*-value tests (p=.001) identified 494 out of 7,793 genes, or 7 percent, as significantly different, while a p-value of .001 would suggest only eight genes would be different based

on chance alone. Of the 494 genes, 418 were grouped together by K-means clustering, indicating the presence of a large number of variable genes, all of which have similar expression patterns. It is still too early to make conclusive statements about the effects that the observed changes in gene expression may have on spinal cord recovery. Furthermore, in the case of the gene coding for the DNA binding protein IRF1, it is very difficult to begin to draw conclusions about how it may be affecting the recovery, as IRF1 has a huge body of information known about it, making it difficult to guess which aspects of the cell it is influencing. Also, as IRF1 is upregulated in the control sample on day 1, and downregulated on days 3 and 7, it is even harder to begin to know what effects it is having.

In the case of E2F5, however, we have preliminary evidence that supports the hypothesis that T-cells have a detrimental effect. E2F5 is known to weakly promote cell entrance into the S-phase,[6] which can lead to cell proliferation. We hypothesize that this increase in cell proliferation could be occurring in macrophages, which T-cells are known to regulate, and which could benefit injury recovery through destroying harmful or badly damaged cells. A previous study conducted at the Hart Laboratory at the W. M. Keck Center for Collaborative Neuroscience that looked at mRNA levels in isolated macrophages did show a small presence of E2F5, but the data on this is not conclusive.

Future Directions

The next step in this project is to try to determine what possible effects the upregulation or downregulation of the significantly different genes would have on injury recovery and to look at how T-cells may cause the differences in expression levels. It is also possible to conduct studies of the genes in question to see what relationships they may have in spinal cord injury and what interactions they have with T-cells.

Perhaps the best way to proceed, however, is to stain tissue samples of rats for T-cells and proteins produced by the genes in question. This will, one hopes, furnish information on how T-cells change protein levels throughout the sample, showing which proteins are directly affected by the T-cells. Looking at how these proteins change cells will then show how the T-cells modify spinal cord injury recovery.

Acknowledgments

I would like to acknowledge Dr. Ron Hart, Rutgers University, for mentoring this study and the W. M. Keck Center for Collaborative Neuroscience, Rutgers University, where the research was conducted; Dr. Cheryl Dreyfus and the Neuroscience Summer Undergraduate Research Program for New Jersey Students, University of Medicine and Dentistry of New Jersey, for creating this research opportunity; Professor Phillip Popovich, Ohio State University, for creating the experiment underlying this study and gathering the RNA samples; and Professor W. S. Klug and the Biology Internship program, The College of New Jersey, for assisting with presentation of this work and the creation of this paper.

References

[1] Moalem G, Leibowitz-Amit R., Yoles E., Mor F., Cohen I., Schwartz M. (1999). Autoimmune T-cells protect neurons from secondary degeneration after central nervous system axotomy. *Nature America,* 5, num 1, 49–55.

[2] Bresnahan J., Beattie M., Stokes B., Conway K. (1991). Three-dimensional computer-assisted analysis of graded contusion lesions in the spinal cord of the rat. *J. Neurotrauma* 8(2), 91–101.

[3] Stokes B., Noyes D., Behrmann D. (1992). An electromechanical spinal injury device with dynamic sensitivity. *J. Neurotrauma* 9, 187–195.

[4] Jakeman L., Guan Z., Wei P., Ponnappan R., Dzwonczyk R., Popovich P., Stokes BT. (2000). Traumatic spinal cord injury produced by controlled contusion in mouse. *J Neurotrauma* 17, 299–319.

[5] Carmel J, Kakinohana O, Mestril R, Young W, Marsala M, Hart R. (2004). Mediators of ischemic preconditioning identified by microarray analysis of rat spinal cord. *Experimental Neurology,* 185, 81–96.

[6] DeGregori J, Leone G, Miron A, Jakoi L, Nevins J. (1997). Distinct roles for E2F proteins in cell growth control and apoptosis. *Biochemistry,* 94, 7245–7250.

[7] Sroga J, Jones T, Kigerl K, McGaughy V, Popovich P. (2003). Rats and mice exhibit distinct inflammatory reactions after spinal cord injury. *Journal of Comparative Neurology,* 223–240.

[8] Chopp M, Zhang X, Li Y, Wang L, Chen J, Lu D, Lu M, Rosenblum M. (2000). Spinal cord injury in rat: Treatment with bone marrow stromal cell transplantation. *Neuroreport,* 3001–5.

[9] Hofstetter C, Schwarzdagger E, Hess D, Widenfalk J, El-Manira A, Prockopdagger D, Olson L. (2002). Marrow stromal cells form guiding strands in the injured spinal cord and promote recovery. *Proceedings of the National Academy of Sciences: Medical Sciences,* 99, no. 4, 2199–2204.

[10] Tamura T, Nakagawa T, Iguchi F, Tateya I, Endo T, Kim T, Dong Y, Kita T, Kojima K, Naito Y, Omori K, Ito J. (2004). Transplantation of neural stem cells into the modiolus of mouse cochleae injured by cisplatin. *Acta Oto-Laryngologica,* 124, Supplement 551, 65–68.

[11] Teng Y, Lavikdagger E, Qu X, Park K, Ourednik J, Zurakowski D, Langer R, Snyder E. (2005). Functional recovery following traumatic spinal cord injury mediated by a unique polymer scaffold seeded with neural stem cells. *Proceedings of the National Academy of Sciences,* 99, no. 5, 3024–3029.

[12] Coumans J, Lin T, Dai H, MacArthur L, McAtee M, Nash C, Bregman Band Barbara S. (2001). Axonal regeneration and functional recoverv after complete spinal cord transection in rats by delayed treatment with transplants and neurotrophins. *The Journal of Neuroscience* 21(23), 9334–9344.

[13] Popovich P, Wei P, Stokes B. (1997). Cellular inflammatory response after spinal cord injury in Sprague-Dawley and Lewis rats. *Journal of Comparative Neurology* 377, issue 3, 443–464.

[14] Moalem G, Monsonego A, Shani Y, Cohen I, Schwartz M. (1999). Differential T-cell response in central and peripheral nerve injury: Connection with immune privilege. *The Federation of American Societies for Experimental Biology Journal,* 13, 1207–1217.

[15] Hauben E, Butovsky O, Nevo U, Yoles E, Moalem G, Agranov E, Mor F, Leibowitz-Amit R, Pevsner E, Akselrod S, Neeman M, Cohen I, Schwartz M. (2000). Passive or active immunization with Myelin Basic Protein promotes recovery from spinal cord Contusion. *Journal of Neuroscience,* 20 (17), 6421–6430.

[16] Hauben E, Ibarra A, Mizrahi T, Barouch R, Agranov E, Schwartz M. (2001). Vaccination with a Nogo-A-derived peptide after incomplete spinal-cord injury promotes recovery via a T-cell-mediated neuroprotective response: Comparison with other myelin antigens. *Proceedings of the National Academy of Sciences: Medical Sciences,* 98, 15173–15178.

Investigation of μ/δ Opioid Receptor Heterodimers *In Vivo*

Haroon Rahimi,
The College of New Jersey
Faculty Sponsor:
Professor W.S. Klug,
Department of Biology

Abstract

There are at least three known classes of opioid receptors: μ, δ, and κ. The complicated and often unpredictable ligand-receptor interactions of many subtypes of these receptors are manifested in the severe and little understood dependency and tolerance afflictions of opiate drug abusers. The seriousness of this problem has stimulated current research into the physical structure of the opioid receptors in order to better understand their function. In this regard, one of the more credible theories postulates that opioid receptors are often found as complexes, the most simple of which are dimers that require stimulation by two ligands to elicit receptor activation. Evidence for opioid receptor heterodimerization has been shown previously. One specific study used varying concentrations of a μ receptor agonist and a δ receptor antagonist on mouse cell membranes from spinal cord tissue, which is known to co-localize μ and δ opioid receptors. When the ligands were used in combination, receptor stimulation was potentiated, a finding that is consistent with the presence of opioid receptor heterodimers in these cells. The present study performed the converse of this experiment. The δ receptor agonist Deltorphin II and the μ receptor antagonist CTOP were used to treat mouse spinal cord tissue cell membranes using a GTPγ^{35}S binding assay. The effect elicited by CTOP was potentiated when used in combination with Deltorphin II to treat membranes, corroborating the results of the previous experiment and lending support to the dimerization theory of opioid receptors. Future studies will be necessary to firmly establish the more general idea that a wide range of opioid receptor dimers underlies the diversity of responses to opioid receptor activation.

Introduction

In recent years, inadequate pain relief as a result of chronic diseases has developed into a serious crisis (Kosten, 2002). The problem has become so drastic, particularly in minority groups in America, that in July 2002, the National Medical Association sponsored a meeting in Washington, D.C., entitled "Managing Pain: The Challenge in Underserved Populations: Appropriate Use versus Abuse and Diversion" (Primm et al., 2004). Experts on pain management and abuse of substances highlighted a variety of reasons for the problem—lack of physician training, denial of sufficient medication to patients, and even funding problems within the Medicaid system, which have prevented access to the best treatment for pain management (Primm et al., 2004). Moreover, pain control problems are seen most commonly in individuals afflicted with certain diseases. For example, in patients with rheumatologic disorders, the most common complaint is the pain associated with the disease. However, medicine administered to help alleviate pain and discomfort is not always effective. A recent survey of rheumatoid arthritis patients in Europe revealed that 67 percent of patients consider their current regimen of pain relief to be unsatisfactory (Borenstein,

2005). While ineffective pain management may have several causes, the most fundamental is an incomplete understanding of the underlying cellular and molecular processes that are altered with medication. Specifically, the mechanism of action of most pain medications involves an effect, at least in part, on the opioid system (Borenstein, 2005).

The opioid system consists of endogenous ligands and several receptors (μ, δ, and κ), which are found primarily in the central nervous system. The first opioid ligand to be discovered was morphine, originally isolated from opium poppies by Serturner in 1803. To this day, morphine remains the most potent opioid agonist ever discovered or synthesized (Schumacher et al., 2002). Three different types of ligands can associate with any particular opioid receptor: agonists, partial agonists, which elicit at least some degree of receptor activity, and antagonists, which bind to but do not stimulate function of the receptor. Morphine, for instance, is an agonist; codeine, a commercially synthesized substance derived from morphine, is a partial agonist; and naloxone is an antagonist. Some opioid ligands have multiple effects because they can bind to different types of opioid receptors. Nalbuphine, for instance, acts as an agonist in association with the κ opioid receptor but acts as an antagonist when bound to the μ opioid receptor. While morphine and other exogenous opioid peptides were discovered first, the existence of naturally produced exogenous opioid ligands has since been established. Endogenous ligands are all derived from precursor proteins, which must be modified to form the functional proteins (Schumacher et al., 2002). These zymogens include prepro-opiomelanocortin, preproenkephalin, and preprodynorphin, forming β-endorphin, the enkepalins, and the dynorphins, respectively (Pintar and Kieffer, 2003). Each of these opioid peptides is relatively selective for a particular receptor subtype; however, there is a great deal of cross reactivity of each ligand with more than one receptor type (Pintar and Kieffer, 2003).

Opioid receptors belong to a large class of receptors that are linked to G-proteins. G-protein-coupled receptors (GPCRs) make up only 5 percent of all cellular proteins (Alves et al., 2004). Because of this low natural abundance of GPCRs in general, little is known about the physical structure and function of opioid receptors. There is great interest in learning more about opioid receptors because they are one of the most important physiological targets of many pharmacologic agents used in medicine. There are three main classes of opioid receptors: μ (MOR-1), δ (DOR-1), and κ (KOR-1). Within each main class of receptors are receptor subtypes, based on ligand specificity. Currently established opioid receptor subtypes include μ_1, μ_2, δ_1, δ_2, κ_1, and κ_2. All of the opioid receptors produce analgesia at both spinal and supraspinal levels. In addition, opioid receptors have been linked to other physiologic functions. For example, members of the MOR-1 class have been shown to modulate neurotransmitter and hormone release, decrease respiration, and induce sedation, while the DOR-1 subtypes also influence hormone and neurotransmitter release, and the KOR-1 subtypes are involved in slowing gastrointestinal transit (Schumacher et al., 2002).

Current evidence indicates that opioid receptors can dimerize, at least, *in vitro*. Using transfected cell lines, a combination of specific antagonists and agonists have been shown to potentiate receptor activity *in vitro* (Gomes et al., 2002; Gomes et al., 2004), consistent with the presence of opioid receptor dimers in the cells. George et al. (2000) performed immunoprecipitation studies in cell lines that co-express DOR-1 and MOR-1 and found the presence of not only dimers but also higher order oligomers. Additionally, functional properties unique to the heteroreceptor complexes were observed, such as a lowered affinity for the MOR-1-selective agonist DAMGO and the DOR-1-selective agonist DPDPE, suggesting the creation of an emergent binding profile that is unique to the complex (George et al., 2000). Based on evidence

suggesting the presence of opioid dimers *in vitro*, it has been theorized that opioid receptors also dimerize *in vivo*. Previously in Dr. John Pintar's laboratory, Ansonoff (unpublished observations) studied adult wild type and various opioid receptor knockout mouse membranes using GTPγ^{35}S binding assays. Specifically, the binding effects on samples of each type of membrane were studied under two different conditions: treatment with DAMGO alone and treatment with DAMGO and TIPPψ, a DOR-1 antagonist. This initial work provided evidence for dimerization of DOR-1 and MOR-1 *in vivo* using this analytical technique (unpublished observations). In conjunction with Ansonoff's work, the present study focused on a set of converse experiments. The GTPγ^{35}S binding assay was used to assess adult wild type and MOR-1 knockout mouse spinal cord membranes under two conditions: treatment with Deltorphin II, a DOR-1 agonist, alone or co-treatment with CTOP, a MOR-1 antagonist. In both the original and converse experiments, antagonist treatment, in combination with agonist, was shown to elicit a potentiative effect, increasing the total amount of receptor stimulation compared to treatment with an agonist alone.

Materials and Methods

Tissue Preparation

Adult wild-type mice and MOR-1 knockout mice were sacrificed via carbon dioxide anesthesia. The heads of the mice were removed, and intrathecal saline injections were performed to eject the spinal cord from the isolated body. The individual spinal cords were collected, and three spinal cords from the same mouse genotype were suspended in 2 mL of homogenization buffer (50 mM Tris-Cl/1 mM ethylenediaminetetraacetic acid (EDTA)/10% sucrose, pH = 7.4 at 0°C) and stored at 0°C until use. Spinal cord cell membranes were used since this tissue is known to co-localize MOR-1 and DOR-1 (Cheng, et al., 1997). The tissue was gently homogenized by hand (Dounce homogenizer) to disrupt the spinal cord tissue such that the cell membranes formed micelles containing the opioid receptors of interest. Generation of micelles in this **way** was crucial to the experimental design since previous attempts with electronic homogenizers did not produce micelles with functional G-proteins that could be activated via extracellular receptor stimulation.

The homogenized tissue was suspended in 25 mL of homogenization buffer and centrifuged for 15 minutes (35,000xg) at 4°C. The supernatant was discarded and the pellets were re-suspended in 10 mL of a different buffer (50 mM Tris-HCl/1 mM EDTA, pH = 7.4), incubated on ice for 30 minutes, and centrifuged again for 20 minutes (35,000xg) at 4°C. After pouring off the supernatant, the pellets were re-suspended in 0.5 mL of a third buffer (50 mM Tris-HCl, pH = 7.4) and the mixture was passed through a 26-gauge needle to ensure a homogeneous suspension by breaking up any clots of tissue that may have formed after the pellet was re-suspended.

Lowry Assay

To assess the total protein concentration of the spinal cord tissue samples, bovine serum albumin (BSA) was used as the standard in a Lowry assay. Five test tubes were prepared, each containing 100 µl of 1 M NaOH. The test tubes received 0, 25, 50, 100, and 200 µg of BSA, respectively. Two additional test tubes were prepared by adding 90 µl of 1 M NaOH to each and then 10 µl of either the wild-type mouse spinal cord mixture or 10 ml of the MOR-1 knockout mouse spinal cord mixture. Thus, all seven tubes contained the same total volume, which was important for an accurate determination of protein concentration. Following these initial preparations, 2 mL of the Lowry mixture (composition: 98 mL 2% Na_2CO_3/1 mL 1% $CuSO_4$/1 mL 1.6% $Na_2C_4H_4O_6$) were added to each tube and incubated for 10 minutes to allow a colorimetric reaction to develop. The Cu^{2+} ions react with the bonds in proteins to turn the yellow mixture a dark blue color. Next, 200 µl of Folin phenol reagent (composition:

1 mL Folin/2 mL dH_2O) were added to each tube, the tubes were vortexed immediately, and incubated for three minutes. The Folin phenol reagent contains a mix of inorganic salts that reacts specifically with the amino acid residues tryptophan and tyrosine to produce a blue-green color. Finally, 200 µl of a dithiothreotol solution (0.30848 g dithiothreotol/100 mL dH_2O), a reducing agent, were added to each tube and again vortexed. Spectrophotometric absorbance values (at 740 nm) were obtained for each of the tubes, and the average protein concentrations for both wild-type and MOR-1 knockout spinal cord homogenate were determined. The membranes were subsequently stored at -70°C at a concentration of 5 µg/µL in buffer. The total protein concentrations were used to ensure that equivalent amounts of wild-type and MOR-1 knockout spinal cord membranes were used in the GTPγ^{35}S binding assay, described below.

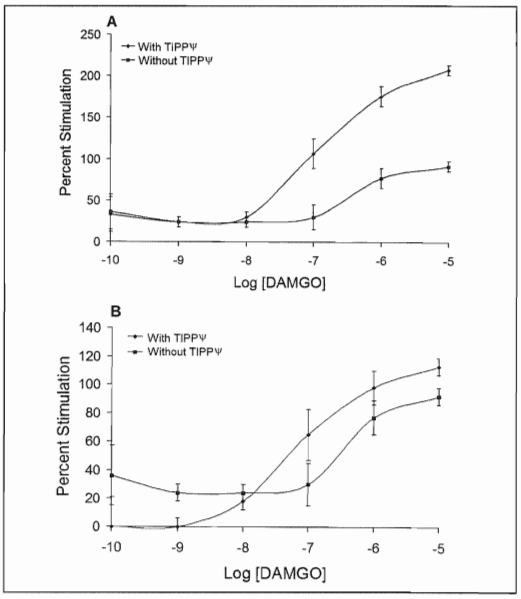
GTPγ^{35}S Binding Assay

After thawing for 30 minutes at room temperature, the wild-type and MOR-1 mouse spinal cord cell membranes were used in the GTPγ^{35}S binding assay protocol established by Nitsche and Pintar (2003). The membranes (2 µg/5 µL) were incubated in assay buffer (50 mM Tris-Cl/3 mM $MgCl_2$/0.2 mM EGTA/100 mM NaCl, pH = 7.4) with GDP (2 mM), GTPγ^{35}S (0.5 nM), GTPγS (2 nM), and varying concentrations (1 nM, 10 nM, 100 nM, 1 µM, 10 µM, 100 µM) of either Deltorphin II alone or co-treatment with a constant CTOP concentration, for 60 minutes at 30°C. The GTPγ^{35}S (specific activity 1250 Ci/mmol) used in the experiments was purchased commercially. Additional tubes were set up to measure basal GTPγ^{35}S binding (buffer, membrane, GDP, GTPγ^{35}S in the absence of exogenous ligand, and any non-specific binding (buffer, membrane, GDP, GTPγ^{35}S, and GTPγS) in the presence of excess non-radiolabelled GTPγS. Incubations of membrane preparations were ended by vacuum filtration onto Whatman GF/B filters using a cell harvester. The filters were washed with 50 mM Tris-Cl (pH = 7.4) at 0°C, extracted into toluene scintillation fluid overnight, and opioid receptor activation was measured using a liquid scintillation counter. The experiments were performed in triplicate, and data were expressed as percent net stimulation using the formula: (Agonist stimulation—Nonspecific stimulation)/(Basal stimulation—Nonspecific stimulation) x 100.

RESULTS

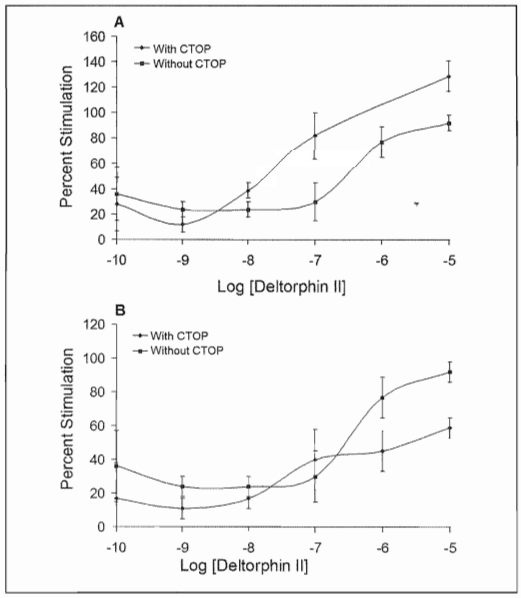
Compared to wild-type mouse membranes treated with the MOR-1 agonist (DAMGO) alone, wild-type membranes treated with both DAMGO and the DOR-1 antagonist (TIPPψ) showed an increase in receptor stimulation, indicating a potentiative effect (Figure 1A). In the DOR-1 knockout mouse membranes, where MOR-1/DOR-1 heterodimers are not possible, there was no difference in receptor stimulation between the membranes treated with DAMGO alone and the membranes treated with DAMGO and TIPPψ (Figure 1B), as expected.

Figure 1. Dose/response curves for G-protein binding in the presence of the MOR-1 agonist DAMGO either alone or in combination with the DOR-1 antagonist TIPPψ in mouse spinal cord membranes isolated from adult wild-type (A) or DOR-1 knockout (B) mice. (A) DAMGO stimulation of G-protein binding was increased in the presence of TIPPψ in wild-type spinal cord membranes. (B) In contrast to wild-type membranes, no difference in DAMGO stimulation of G-protein binding in the presence of TIPPψ was observed in membranes isolated from DOR-1 mutants. All values are expressed as means ± SD.



In the converse set of experiments, wild-type mouse membranes treated with both Deltorphin II and CTOP exhibited a greater percentage of receptor stimulation compared to membranes treated with Deltorphin II alone (Figure 2A). Moreover, in the MOR-1 knockout mouse membranes, where MOR-1/DOR-1 heterodimers cannot exist, there was an equivalent amount of receptor stimulation in membranes treated with Deltorphin II alone and membranes treated with Deltorphin II and CTOP (Figure 2B).

Figure 2. Dose/response curves for G-protein binding in the presence of the DOR-1 agonist Deltorphin II alone or in combination with the MOR-1 antagonist CTOP in spinal cord membranes isolated from adult wild-type (A) or MOR-1 knockout (B) mice. (A) Deltorphin II stimulation of G-protein binding was increased in the presence of CTOP. (B) In contrast to wild-type membranes, no difference was observed in Deltorphin II stimulation of G-protein binding in the presence of CTOP. All values are expressed as means ± SD.



Discussion

The present study demonstrates that a DOR-1 antagonist can potentiate μ-agonist stimulation of GTPγ^{35}S binding in membranes isolated from wild-type mice, and in the absence of functional DOR-1 receptors in the DOR-1 knockout mouse membranes, a δ antagonist cannot potentiate δ-agonist stimulation of GTPγ^{35}S binding. Similarly, a MOR-1 antagonist can potentiate d-agonist stimulation of GTPγ^{35}S binding in membranes from wild-type mice, while a μ antagonist cannot potentiate δ-agonist stimulation of GTPγ^{35}S binding in membranes isolated from MOR-1 knockouts, which lack functional MOR-1 receptors. These data therefore support the hypothesis that MOR-1/DOR-1 heterodimers exist in mouse spinal cord membranes *in vivo*. Spinal cord membranes from wild-type mice consistently responded pharmacologically as if MOR-1/DOR-1 heterodimers existed *in vivo*, while mutant mouse membranes lacking either the MOR-1 or the DOR-1 receptor (in which MOR-1/DOR-1 heterodimer formation is not possible) displayed none of the pharmacological responses that would be expected if MOR-1/DOR-1 heterodimers were present.

Although evidence has accumulated for the existence of opioid receptor dimers, functional characteristics, such as their specific binding properties, are not well known. Nevertheless, some insight has been provided in recent studies regarding the functionality of certain opioid receptor dimers. For example, Cvejic and Devi (1997) used cell lines to demonstrate that agonist binding to a δ/δ homodimer first induced the dimer to monomerize, which then required subsequent interactions with adapter proteins. Differential receptor dimerization is not limited to the opioid receptors. George et al. (2000) have shown that other GPCRs, including the serotonin 5HT-1B and 5HT-1D and the γ-aminobutyric acid B1 and B2 receptors, dimerize to elicit novel functions. In both cases, not only are different binding properties generated, but many of the emergent properties resulting from dimerization appear to serve important regulatory functions for the cell. For instance, receptor internalization by the cells expressing these receptors only occurs when the receptors have dimerized (George et al., 2000). Continued research of opioid receptor dimers may produce similar findings.

Dimerization has also been recognized as occurring between opioid receptors and other types of GPCRs. Jordan et al. (2001)

studied the oligomerization of β_2-adrenergic receptors with opioid receptors. Whether as a dimer with other types of β_2-adrenergic receptors or as a monomer, β_2-adrenergic receptors are known to couple to stimulatory G-proteins, while opioid receptors couple solely to inhibitory G-proteins and thus have different types of secondary effects. In the case of dimerization between opioid receptors and β_2-adrenergic receptors, novel pharmacologic and binding properties were not detected, in contrast to the differential effects observed in dimers of receptors of one receptor type. Whether KOR-1 or DOR-1 complexed with the β_2-adrenergic receptors, in all cases the resulting binding properties were as if the individual receptors existed as monomers of the β_2-adrenergic receptor and the KOR-1 or DOR-1 receptor, using a variety of agonists and antagonists. Thus, the generation of novel functional properties as a result of dimerization seems to be restricted to dimerization within one receptor type, as demonstrated within the opioid receptors.

Establishing the functional properties of specific opioid receptor dimerizations may aid in drug design, allowing the production of drugs that would selectively target receptor complexes associated with specific physiologic effects. This would help to minimize the problem of side effects caused by nonspecific cross-reactivity with other receptor complexes. Indeed, evidence that this will be possible is beginning to emerge. For example, Fujii et al. (2004) modified a receptor agonist for the putative ε opioid receptor TAN-821 and created a specific antagonist for the receptor that demonstrated little cross-reactivity with other opioid receptors. Similar techniques have been used to create synthetic MOR-1 agonists (Page et al., 2003). Such drug design would prove invaluable in creating synthetic ligands capable of stimulating these opioid receptor dimers to capitalize specifically on their novel functions.

Differences in affinity for agonists and antagonists by monomers and dimers of the various opioid receptors may explain the differential physiological responses exhibited by different receptor complexes. Charles and Hales (2004) studied the generation of novel functional properties in cells co-expressing MOR-1 and DOR-1 as a result of dimerization. Such dimers were found to couple an excitatory G-protein leading to new functional effects. With more research into the specific cause-and-effect relationships between receptor dimers and their functional responses, pharmacological approaches aimed at providing more effective analgesia and a better understanding of how to control the dependence and tolerance problems currently associated with many medicinal drugs would likely result.

REFERENCES

Alves, I. D., Ciano, K. A., Boguslavski, V., Varga, E., Salamon, Z., Yamamura, H. I., Hruby, V. J., and Tollin, G. (2004). Selectivity, cooperativity, and reciprocity in the interactions between the δ-opioid receptor, its ligands, and G-proteins. *The Journal of Biological Chemistry.* 279, 44673–44682.

Borenstein, D. (2005). Opioids: To use or not to use? That is the question. *Arthritis & Rheumatism.* 52, 6–10.

Charles, A. C., and Hales, T. G. (2004). From inhibition to excitation: Functional effects of interaction between opioid receptors. *Life Sciences.* 76, 479–485.

Cheng, P. Y., Liu-Chen, L., and Pickel, V. M. (1997). Dual ultrastructual immunocytochemical labeling of μ and δ opioid receptors in the superficial layers of the rat cervical spinal cord. *Brain Research.* 778, 367–380.

Cvejic, S., and Devi, L. A. (1997). Dimerization of the d opioid receptor: Implication for a role in receptor internalization. *The Journal of Biological Chemistry.* 272, 26959–26964.

Fujii, H., Narita, M., Mizoguchi, H., Hirokawa, J., Kawai, K., Tanaka, T., Tseng, L. F., and Nagase, H. (2004). Rational drug design and synthesis of a selective ε opioid receptor antagonist on the basis of the accessory site concept. *Bioorganic & Medicinal Chemistry Letters.* 14, 4241–4243.

George, S. R., Fan, T., Xie, Z., Tse, R., Tam, V., Varghese, G., and O'Dowd, B. F. (2000). Oligomerization of μ- and δ-opioid receptors. *The Journal of Biological Chemistry.* 275, 26128–26135.

Gomes, I., Filipovska, J., Jordan, B. A., and Devi, L. A. (2002). Oligomerization of opioid receptors. *Methods.* 27, 358–365.

Gomes, I., Gupta, A., Filipovska, J., Szeto, H. H., Pintar, J. E., and Devi, L. A. (2004). A role for heterodimerization of μ and δ opiate receptors in

enhancing morphine analgesia. *Proceedings of the National Academy of Sciences.* 101, 5135-51391-5.
Jordan, B. A., Trapaidze, N., Gomes, I., Nivarthi, R., and Devi, L. A. (2001). Oligomerization of opioid receptors with β_2-adrenergic receptors: A role in trafficking and mitogen-activated protein kinase activation. *Proceedings of the National Academy of Sciences.* 98, 343–348.
Kosten, T. R. (2002). Drugs of abuse. 517–528.
Nitsche, J. F., and Pintar, J. E. (2003). Opioid receptor-induced GTPγ^{35}S binding during mouse development. *Developmental Biology.* 253, 99–108.
Page, D., Nguyen, N., Bernard, S., Coupal, M., Gosselin, M., Lepage, J., Adam, L., and Brown, W. (2003). New scaffolds in the development of mu opioid-receptor ligands. *Bioorganic & Medicinal Chemistry Letters.* 13, 1585–1589.
Pintar, J. E., and Kieffer, B. L. (2003). Genetic analysis of the endogenous opioid system. 313–41.
Primm, B. J., Perez, L., Dennis, G. C., Benjamin, L., Clark, W., Keough, K., Leak, W. D., Payne, R., Smith, D., and Sullivan, L. W. (2004). Managing pain: The challenge in underserved populations: Appropriate use versus abuse and diversion. *Journal of National Medicine Association.* 96, 1152–1161.
Schumacher, M. A., Basbaum, A. I., and Way, W. L. (2002). Opioid analgesics & antagonists. 497–516.

ACKNOWLEDGMENTS

I would like to thank Dr. Cheryl Dreyfus for coordinating the Neuroscience Undergraduate Summer Research Program for New Jersey Students, Dr. John E. Pintar for providing the opportunity to work in his laboratory, and Dr. Michael Ansonoff and Dr. Tracy Czyzyk for mentoring through the research experience. Additional thanks to the professors at The College of New Jersey: Dr. W. S. Klug, for coordinating the research internship program, and Dr. Jeffery Erickson, for his guidance through the analysis and understanding of the research performed.

Temperature Dependence of the Raman Spectra of Silica Glass

Richard Ottens,
The College of New Jersey
Faculty Sponsor:
Professor Romulo Ochoa,
Department of Physics

Abstract

When laser light is shined onto a sample, the scattered light can have energies that differ from those of the incident light. This scattered light produces a Raman spectrum that is temperature dependent. These studies were focused on obtaining the Raman spectra of Dynasil glass at different temperatures and then comparing our measurements with theoretical predictions. The technique developed could be applied to non-contact thermometers.

Introduction

Raman scattering involves the inelastic collisions of light with the atoms or molecules in a material. These inelastic collisions are caused by quasi-excitations, that is, vibrational modes of a lattice in a solid (phonons). Raman spectroscopy is the measure of these vibrational modes.

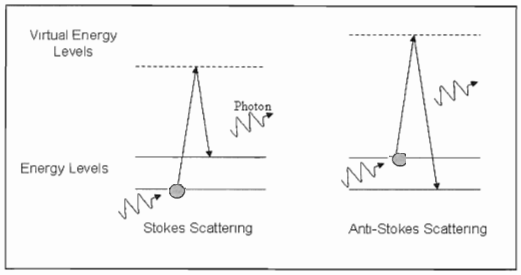
Classically, Raman scattering can be explained by the time-varying polarizability of a molecule. When a molecule is subjected to an electric field of a beam of electromagnetic radiation (light), the dipole moment oscillates. If the molecule has small vibrational modes, the polarizability equation of the molecule will contain many parts that contribute to its polarization. The most important parts in the polarizability equation are Rayleigh scattering, infrared absorption, and Raman scattering. If there is a change in the polarizability tensor with respect to the normal coordinates of the vibrating molecule, then the Raman scattering part will become nonzero and Raman scattering will occur.

These inelastic collisions can be explained another way. As photons (light) hit a molecule, an electron-radiation interaction occurs. This means that as photons of a certain frequency hit the molecule and the energy of the photons is absorbed by the molecule, the energy will cause two things to occur. The first is that the molecule will vibrate through an electron-lattice interaction. The second is that some of that energy will create another photon of a slightly different frequency from the original photon through another electron-radiation interaction. Sometimes the incident light will cause the molecule to vibrate more than its original state, and as a consequence the scattered light will be of a lower frequency. When this occurs it is called Stokes scattering. Other times the incident light will cause the molecule to vibrate less than it originally was; this will cause the scattered light to be higher in frequency than the incident light. When this occurs it is called anti-Stokes scattering.

Quantum mechanically, Raman scattering can be thought of as the excitation of electrons. When a photon hits an electron, the energy of the photon will get transferred into the electron, causing the electron to go into a higher energy level called a virtual energy level. The electron tends to lose that energy quickly, but in the case of Raman scattering the electron does not return to its original energy level (hence Raman scattering is an inelastic process since energy is gained or

lost). When the electron decays it will release another photon that will have a slightly different frequency from the incident photon. If an electron is excited to a virtual energy level and changes to a higher energy level than it was in originally, this is called Stokes scattering. If it were to drop down to a lower energy level than it was in originally it would be called anti-Stokes scattering. This process can be seen in Figure 1.

Figure 1. This is a quantum mechanical model of the Raman scattering process. As a photon hits an electron it will cause it to jump to a virtual energy level. It will then move to another energy level (either slightly higher or slightly lower than the original), causing another photon to be produced.



No matter which way one conceives of the operation of Raman scattering, classically or quantum mechanically, the resulting effect is the same. When one shines a laser of a specific frequency at an object, some of the light detected from the surface is slightly different in frequency. If one were to plot the light intensity versus the wavelength of the light detected from the surface, then a graph similar to Figure 2a would be seen. The only difference between Figure 2a and Figure 2b is that the wavelength is converted to more appropriate inverse centimeter units to convey the independence of the Raman spectra on the incident laser light. Additionally, by convention, the Stokes scattering will be on the positive side (right side) and the anti-Stokes scattering will be on the negative side (left side) of the graph. The peaks observed correspond to different vibrational states of the molecules in the material investigated.

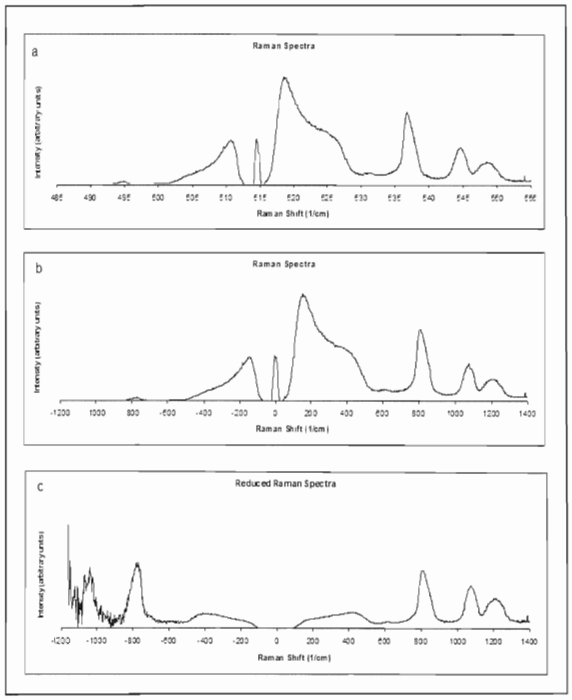
EXPERIMENT

2.1 Setup

The purpose of these studies is to confirm the temperature dependence when measuring Raman spectra. In this experiment, the Raman scattering spectra of Dynasil glass (a highly pure fused form of silica glass) is found. To do this, two stages of the experiment have to be performed. The first stage is to set up the experiment to measure Raman spectra at room temperature. The second stage is to find a reliable way to heat the glass to a constant temperature so that heating will not interfere with measuring the Raman spectra.

These experiments used a laser, lenses, mirrors, two notch filters, and a spectrometer. The laser used in this experiment was a frequency-varying Argon-ion laser. It was used at approximately 300 mW with a wavelength of 514 nm, which is green light. The laser beam was directed into a lens that focused the beam at the sample of Dynasil glass. The glass scattered light in all directions. Some of that light went through two notch filters. Notch filters are designed to filter out the light of the laser (514 nm in wavelength) so that only the relatively low intensity Raman-scattered spectra pass through and not the laser light that reflected off the glass. The filtered light was then sent into a spectrometer. Spectrometers diffract light into all the wavelengths of light, like a prism diffracts white light into a rainbow. The diffracted light is sent into a charge-coupled device (CCD) camera (similar to a digital camera) that measures the intensity of scattered light within a specific range of wavelengths and plots these data points on a computer.

Figure 2. These three graphs show the Raman spectra in three different forms. Graph a is a graph of the data points received by the spectrometer without any manipulations. It is a plot of photon counts versus wavelengths of light. Graph b is the same as graph a except the horizontal axis is in inverse centimeters, which is the conventional way of representing Raman spectra. Graph c is a reduced Raman spectrum where both sides of the graph are made to look like mirror images of one another.



2.2 Viewing Temperature Dependence

Figure 2b illustrates that peaks on the Stokes scattering side (right side) occur nearly the same distance from the laser line (zero) as peaks on the anti-Stokes scattering side (left side). This occurs because these peaks correspond to the same change in energy. Stokes scattering occurs more often than anti-Stokes scattering at room temperature because many of the electrons are at their lowest energy level and can only go to higher energy levels. Using an equation to balance out the intensities of the Stokes and anti-Stokes manipulates the data points so that both sides of the graph resemble mirror images of each other. This equation is called the Reduced Raman Spectra Equation and can be seen as Equation 1 in Figure 3. In this equation the Bose-Einstein Occupation Number is temperature dependent as can be seen in Equations 2 through 5 in Figure 3, where T represents temperature. The Stokes scattered light and the anti-Stokes scattered light have different equations for their Bose-Einstein Occupation Number as can been seen in Equations 2 and 4 of Figure 3. This causes the Reduced Raman Spectra Equation to be different for Stokes and anti-Stokes scattering as can be seen in Equations 3 and 5 in Figure 3. When these equations are applied to a Raman spectrum graph (Figure 2b) a new reduced Raman spectrum graph is formed (Figure 2c). The two sides of the graph are mostly mirror images of one another. If the reduced Raman spectra graphs are found for a sample at different temperatures, the graphs should line up and prove that there is a temperature dependence.

Figure 3. This is the Reduced Raman Spectra Equation. Equation 1 is the general form, and each part is labeled. Equation 2 is the Bose-Einstein Occupation Number for Stokes scattering. Equation 3 is the compete Reduced Raman Spectra Equation for Stokes scattering. Equation 4 is the Bose-Einstein Occupation Number for anti-Stokes scattering. Equation 5 is the compete Reduced Raman Spectra Equation for anti-Stokes scattering.

Reduced Raman Spectra / Raman Scattering Proportionality

1) $$I^{p}_{red}(\omega) = I^{p}(\omega_L, \omega_S)(\omega_L + \Omega)^{-4}\left[\frac{n(\Omega)}{\Omega}\right]^{-1}$$

Measured Raman Spectra / Bose-Einstein Occupation Number (Temperature Dependent)

For Raman Shift (Ω) < 0 (Stokes)

2) $$\frac{n(\Omega)}{\Omega} = \omega^{-1}\left[\left[e^{\frac{\hbar\omega}{k_B T}} - 1\right]^{-1} + 1\right]$$

3) $$I^{p}_{red}(\omega) = I^{p}(\omega_L, \omega_S)(\omega_L - \omega)^{-4}\omega\left[\left[e^{\frac{\hbar\omega}{k_B T}} - 1\right]^{-1} + 1\right]^{-1}$$

For Raman Shift (Ω) > 0 (Anti-Stokes)

4) $$\frac{n(\Omega)}{\Omega} = \omega^{-1}\left[e^{\frac{\hbar\omega}{k_B T}} - 1\right]^{-1}$$

5) $$I^{p}_{red}(\omega) = I^{p}(\omega_L, \omega_S)(\omega_L + \omega)^{-4}\omega\left[e^{\frac{\hbar\omega}{k_B T}} - 1\right]$$

2.3 Heating Up the Sample

Three ways of heating the Dynasil glass were attempted to determine which one would work best. The first was to use a Peltier cooler. A Peltier cooler has one side that becomes hot and another that becomes cool when a current is passed through it. Although a Peltier cooler was used, more conclusive results were obtained using other methods of heating that provided higher temperatures.

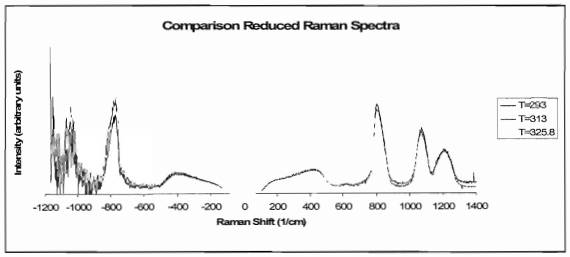
The second way of heating the Dynasil glass was to boil it in water. However, this was problematic because the temperature of the glass decreased too quickly when taken out of the water and could not be used.

The third way of heating involved using a soldering iron that touched the side of the glass. The soldering iron used two different temperature settings and raised the glass by ~16°K and ~28°K. This turned out to be the best heater of the three; it did not interfere with measuring the Raman spectra, it caused the greatest increase in temperature, and it provided two different temperatures without a change in the setup.

Results

When the reduced Raman spectra taken at room temperature and the two different temperatures provided by the soldering iron were overlapped, Figure 4 was the result. These three graphs virtually overlap, thus showing the equations are true and that there is a temperature-dependence in the Raman spectra.

Figure 4. This graph compares three reduced Raman spectra taken at different temperatures. The three virtually overlap one another, showing that the Reduced Raman Spectra Equation holds.



There were many factors that led to the success of these studies. One such factor was the ability to use a high-powered laser. The power of the laser is much greater than that of a laser pointer. Without such power output, a Raman spectrum would be difficult, indeed nearly impossible to distinguish from background light. Also with higher power, less exposure time is needed for the spectrometer, since more power means more photons.

Another factor leading to the success of these experiments was that a way was found to block out laser light that was reflected off many objects in the room, such as the table, walls, and metal poles. If this light had not been blocked, only short exposure times of the spectrometer could have been done and the CCD would have become oversaturated with light and possibly damaged. By using black cardboard and black cloth, much of the undesirable reflected light did not go into the spectrometer and therefore longer exposures could be achieved.

Finding a reliable way to heat the sample was one of the most important factors to the success of these experiments. Of the several methods attempted, using a simple soldering iron, which is a common laboratory item, led to a cheap and effective success.

Summary

The temperature dependence of Dynasil glass was measured and the theoretical equation was verified for the temperatures used in this experiment. Future experiments should be done to verify whether the equation holds at higher and lower temperatures. Experiments can be done that use Raman spectroscopy to measure the temperature of objects at remote locations or under working conditions.

Bibliography

[1] "Dynasil Glass," Fused Silica, Dynasil Corporation, West Berlin, New Jersey, http://www.dynasil.com.
[2] Long, D. A. "Raman Spectroscopy." 1977.
[3] Prevot, B. and Wanger, J. Prog. "Raman Characterization of Semiconducting Materials and Related Structures." *Crystal Growth* and *Characterization*, Vol. 22, pp. 245–319, 1991.

Conference Proceedings

Listed below are abstracts of student-faculty collaborative work presented at regional, national, and international conferences.

Stacey Van Metre, JenniLee Groegler, Laura Munice, Jaclynn Heefner, Danielle McConnell, Elizabeth Kuhn, Jackie Kotler, Marisa Leston, Jean Walker, and Kristen LePage
The College of New Jersey
(Blythe Hinitz, Faculty Sponsor)
Native American Literature and Arts: Their Importance in the Conceptual Development of Young Children in New Jersey

Presented at the Annual Conference of the New Jersey Association for the Education of Young Children, New Brunswick, NJ, October 18, 2004

This 10-person presentation was designed to inform teachers of young children about elements of Native American culture and to encourage them to avoid Native American stereotyping. Various stories, such as *The Mud Pony* and *How Turtle Flew South for the Winter,* were introduced and discussed. In addition, the group demonstrated how to work on such craft projects as a rain stick, pinch pots, and a turtle shell rattle. Rainmaking and the use of the talking stick were also taught. From start to finish, the presentation emphasized the uniqueness of each Native American tribe's culture and rituals.

Because learning styles vary from child to child, the presentation incorporated insights and activities derived from Howard Gardner's work on multiple intelligences. The presenters stressed the vast number of ways students learn and how to utilize students' different strengths and learning skills. They then provided examples of how to integrate students' different forms of intelligence into lessons. For example, a rain stick was incorporated into a music/movement activity to give students freedom of expression. Finally, the presenters discussed the Native American belief in an animated Nature, that is, a natural world informed by spirits.

Nicole Hartman and Valerie Episcopo
The College of New Jersey
(Diane Bates, Faculty Sponsor)
Chronic Stress and Social Location

Presented at the American Sociological Association annual meeting, Philadelphia, PA, August 15, 2005

The Hierarchical Stress Model suggests that socioeconomic status is responsible for health inequality. Health problems have changed as scientists and public health officials have learned about diseases, but the inequality of who experiences illnesses and who lives the longest has endured. This session examines different social locations and how they create chronic stress. We will address control in the work environment, social factors that influence women's entrance into the work force, policies that create unequal access to health care, the effect of education on one's sense of well-being, and how marriage changes a person's outlook on life.

Richard Ottens
The College of New Jersey
(Malcolm Boshier, Faculty Sponsor)
Electronic Current Control Circuit for a Direct Current Electromagnet
LA-UR-05-5993
Highlighting Student and Postdoctoral Research

Presented at the Los Alamos National Laboratory Science and Technology Base, Education Program Office Symposium, University of New Mexico, Los Alamos Branch Campus, Los Alamos, NM, August 2, 2005

Keeping conditions constant in scientific experiments is very important, whether it be pressure, temperature, or electric current. If these conditions are left unchecked, instabilities tend to develop. Such instabilities can be controlled and reduced by using feedback loops. In a Bose-Einstein condensation experiment the use of electromagnets is a very important part of trapping and cooling atoms. These electromagnets need to be kept at a constant current for long periods of time in order for the experiment to work properly. Therefore, I set out to design, construct, deploy, and document a control circuit for a direct current electromagnet used on a Bose-Einstein condensation experiment.

Dennis Chin
The College of New Jersey
(Elizabeth Borland, Faculty Advisor)
Racial and Religious Reconciliation: A General Discussion of Minority Identity Transformation among Second Generation Asian-American Christians

Presented at the Annual Conference of the American Sociological Association, Philadelphia, PA, August 14, 2005

This paper focuses on the racial and religious identities of second-generation Asian-American Christian students. It draws upon the literature on religious involvement and applies racial formation theory and the racial reconciliation perspective to data collected in personal interviews with Asian-American undergraduates. Through integration and application of these concepts and a consideration of ethnic stereotypes, the author concludes that second-generation Asian-American Christian students undergo a transformation of their minority identity through reconciliation of their racial and religious identities.

Mitchell Kleiman and
Dr. Orlando Hernández
The College of New Jersey
(Orlando Hernández, Faculty Sponsor)
Face Recognition Using Multispectral Random Field Texture Models, Color Content, and Biometric Features

Presented at the 34th Applied Imagery Pattern Recognition Workshop, Washington, D.C., October 19, 2005

Most of the available research on face recognition has been performed using grayscale imagery, thus ignoring the texture and color characteristics of human faces. This paper describes a novel two-pass face recognition system that uses a Multispectral Random Field Texture Model, specifically the Multispectral Simultaneous Auto Regressive (MSAR) model and illumination-invariant color features. During the first pass, the system detects and segments a face from the background of a color image and confirms the detection based on a statistically modeled skin pixel map and the elliptical nature of human faces, then reorients and resizes the face to a standard size. In the second pass, the face regions are located using the same image segmentation approach on a subspace of the original image, biometric information, and spatial relationships. The determined facial features are then assigned biometric values based on color information and anthropometrics, and a set of vectors is created to determine similarity in the facial feature space. The performance of the system is tested on two databases.