Virtue and Animal Ethics: An Interdisciplinary Study  
Alison K. Oliver, University of San Diego  
Mentor: Christopher Carter

Paradise Uncaged: John Milton’s Subtle Fixation with Constraints  
Jack H. Wolfram, Emory University  
Mentor: Sarah Higinbotham

Musical Hybridization and Political Contradiction: The Success of Arthur Honegger’s Antigone in Vichy, France  
Emma K. Schubart, University of North Carolina, Chapel Hill  
Mentor: Sharon James

The Stratification of Justice: Evaluating the Relationship between Gender, Race, and Crime  
Catherine Bain, Butler University  
Mentor: Krista Cline

“OK Boomer!”: How a Pejorative can Promote Societal Change through Humor, Impoliteness, and Carnivalesque  
Gwendolyn M. Goodhead, Northwest University  
Mentor: Jacob Witt
Mulier est Urbis Confusio? —The Function of Rhea Silvia, Tarpeia, and Horatia as Exempla in Livy’s *The Early History of Rome*

Bhadrajee S. Hewage, Princeton University
Mentor: Andrew Feldherr

Privilege & Voting for Predators: Exploring the Relationship between Privilege and Voting for Sex Offenders

Reilly Simmons, Butler University
Mentor: Gregory Shufeldt

Rockin’ the GRE: The Effects of Preferred, Non-preferred, and Classical Music on College Students’ Cognitive Test Performance

Hannah B. Bolander, Sean Callahan, Butler University
Mentor: Tara Lineweaver

The Campus as Commons: Educational Anchor Institutions and their Implications for Political Economy

Cameron N. Conner, Whitman College
Mentor: Keith Farrington

The Impact of Nurse Stress, Job Distress, Job Satisfaction, and Mood on Patient Outcomes

Madelynne Wright, Abbey Collins, Butler University
Mentor: Tara Lineweaver

On the Hybrid Nature of Brown Dwarfs

Karla M. Rojas, Amsterdam University College
Mentor: Lotte Tavecchio

Failures of Power-Balancing: The Ottomans and British in Iraq and Kuwait

Amna Abudyak, American University of Sharjah
Mentor: Isa Blumi
Effects of the COVID-19 Pandemic on U.S. Employment Rates

Jordan Neese, Olivet Nazarene University
Mentor: Kayla Bruce

The Political and Economic Impact of Media Ownership Structures and Concentration on Global Communication

Yuxi Wen, University of Nottingham Ningbo China
Mentor: Xiaoling Zhang, Angela Lewis

Goodpasture Syndrome: An Investigation of Disease Process, Diagnosis, Treatment, and Contribution to Intrapulmonary Hemorrhage

Julia Shield, Queensland University of Technology
Mentor: Mark B. Plenderleith

Complex Regional Pain Syndrome: A Scholarly Review

Joseph O’Brien, University of Rhode Island
Mentor: Guy W. Fried

Painting a Pretty Picture: The Role of Social Desirability in the Memory Self-Efficacy of Young and Older Adults

Keegan G. Sawin, Butler University
Mentor: Tara Lineweaver
VOLUME 7 REVIEWERS

Andrew Levy
Department of English

Mandy Hall
Department of Psychology

Joel Martin
Department of Psychology

Brian Day
Department of Psychology

Carol Reeves
Department of English

Brian Giesler
Department of Psychology

William Walsh
Department of English

Ania Spyra
Department of English

Clare Carrasco
Department of Music

India Johnson
Department of Psychology

Rocky Colavito
Department of English

Chris Bungard
Department of Philosophy, Religion, and Classics

Krista Cline
Department of Sociology and Criminology

Tonya Bergeson
Department of Communication Sciences & Disorders

Kenneth Colburn, Jr.
Department of Sociology and Criminology

Brenda Quincy
College of Pharmacy and Health Science

Margaretha Geertsema-Sligh
College of Communication

Xianming Han
Department of Physics and Astronomy

Shyam Sriram
Department of Political Science

William Rieber
Department of Economics, Business Law, and Finance

Amy Peak
College of Pharmacy and Health Science

VOLUME 7 BUTLER UNIVERSITY SPONSORS

College of Liberal Arts and Sciences
Office of the Provost
MISSION AND EDITORIAL POLICY

The mission of the Butler Journal of Undergraduate Research (BJUR) is to advance Butler’s commitment to undergraduate research by seeking to publish original, high quality empirical research undertaken by undergraduates from any university. The journal is national in scope, multidisciplinary, and seeks to publish the best empirical and scholarly research available from undergraduate students anywhere. Submission of original, scholarly research articles is open to undergraduates from any accredited college and university.

The publication of the Butler Journal of Undergraduate Research is designed to build upon and strengthen Butler’s commitment to quality undergraduate research by providing an outlet for the publication of outstanding undergraduate scholarship. BJUR recognizes and supports diverse disciplines, perspectives and methods across the social sciences, natural and life sciences, the arts and humanities. The Journal publishes papers that represent original research undertaken by a student or students during their undergraduate career.

BJUR is committed to intellectual integrity, rigorous standards of scholarship, and rational and civil discourse. It seeks to foster open and critical inquiry that privileges no particular disciplinary standpoint while operating within the limits of a standard of scientific, scholarly and empirical discourse committed to the pursuit of knowledge and truth. While every effort is made to publish issues that represent a fair balance of scholarship across the entire spectrum of undergraduate studies, the Journal’s first priority is always to publish the best undergraduate research available at any given time regardless of disciplinary representation.

POWERED BY

© 2021 Butler Journal of Undergraduate Research
digitalcommons.butler.edu/bjur
VIRTUE AND ANIMAL ETHICS: AN INTERDISCIPLINARY STUDY

ALISON K. OLIVER, UNIVERSITY OF SAN DIEGO
MENTOR: CHRISTOPHER CARTER

Abstract

Analysis of important components from the sciences, philosophy, and theology makes apparent that nonhuman animals require moral consideration beyond what they typically receive in Western culture. Humans are in the habit of inaccurately justifying harmful treatment of animals without any proper evaluation of ethical behavior; the current standards of morality regarding nonhuman species are insufficient. For instance, in the context of the latest scientific advancements in animal behavior and cognition, many aspects of the partition between human and animal intelligence have become ambiguous; however, the typical human mindset stubbornly maintains the notion that humans are ultimately superior, and the perceived lack of intellect of nonhuman animals (a premise that has yet to be empirically defined) is still employed as reason enough for humans to dominate these other species. Furthermore, Christian scripture emphasizes the importance of peace and consideration while simultaneously denouncing violence and viciousness; thus, a Christian theology premised on love, justice, and care for “the least of these” (Matthew 25:40) is in favor of an ethical reevaluation rather than the continuation of current animal maltreatment. Upon the examination of virtue across multiple significant fields, it is clear that a principle moral norm is ignored once extended beyond human interactions, that unwarranted cruelty is wrong. Nonetheless, the current standards of animal treatment are highly characteristic of unwarranted cruelty; humans have exploited the presence of nonhuman species to a point that is definitively unethical.

While we all, as a civilization, still have an incredible amount of work to do before we reach an ideal of human equality, many will not rest until their needs are met. However, there are those that are entirely unable to have a voice and oppose their oppressors. These beings share our world, our basic needs, and our unique abilities of perception and subjective experience yet do not share what we have defined as human rights. Animals, or more accurately, nonhuman animals, are not treated with the same moral consideration as are humans. While this reality is
thought of as fact rather than ethical dilemma, few have confronted animal ethics with the same severity as standard “ethics”—which involves moral behavior and interactions between humans and only humans. As Western civilization continues to address the marginalized and to move toward a goal of ethical treatment among humans, ignoring other species in this consideration should no longer be the customary practice for those who strive for virtue. The human use of nonhuman species has distorted necessity into the installation of exploitative industries that disregard those species’ status as living creatures and minimizes the gravity of unnecessary violence. The way our society has evolved requires new reflection on moral norms, which should include paying attention to the nonhuman members of our communities.

Many would argue that there is no place for theology or philosophy within the sciences, or vice versa, yet I consider these to agree in many aspects—and this is where truth lies. I will demonstrate how the fundamental traditions from religion and philosophical thought supplement new scientific findings in a complicated yet enlightening way, which I will apply to my analysis of morality as it pertains to nonhuman and human interactions.

The remainder of this paper will address how humans are in the habit of inaccurately justifying harmful treatment of animals without proper evaluation of ethical behavior, and how the current standards of morality regarding nonhuman species are insufficient. I will assess various beliefs and stances on morality and will discuss how and why these notions should be applied to all sentient creatures. Because human well-being does not rely on harmful, exploitative oppression of nonhuman animals, cruel treatment of nonhuman animals is unwarranted and must be modified. By extending the implications of virtue to all sentient beings rather than restricting them to interpersonal interactions, we can discover how an ethical relationship can be established between nonhuman and human animals. With this interdisciplinary analysis of virtuous thought, I will argue that nonhuman animals require moral consideration beyond what they typically receive, and I will explore how ethics can and should be extended to nonhuman animals.

**Understanding Morality**

What first must be asked is, Should ethics be applied to the treatment of species other than our own, and if so, how? Humans do have an instinctive tendency to first take care of themselves before anyone else, and although this may be natural, it is often taken to an extent that is unarguably immoral. As those in power face ethical dilemmas, they do not first ask, What would be the virtuous choice? but
rather, How does this benefit me? Rather than striving to love their neighbors or do the most good possible in any given situation, people desire outcomes that allow them to gain individual profits regardless of how the outcomes may affect others, especially when the others are not in direct sight.

This phenomenon is especially prominent in terms of animal ethics. Institutionalized animal violence prevails largely because it is out of sight and thus out of mind. For example, in his Animal Liberation, well-known Australian philosopher Peter Singer has compiled an extensive list of extremely destructive experiments performed on animals. Although many people may know that research is conducted using animals in the sciences, we are incredibly unaware of the extents to which researchers go to keep the public from noticing what occurs behind closed laboratory doors. Singer explains, “It is not surprising that the public still has not the remotest idea of the extent of animal experimentation. Research facilities are usually designed so that the public sees little of the live animals that go in, or the dead ones that come out. (A standard textbook on the use of animals in experimentation advises laboratories to install an incinerator, since the sight of dozens of bodies of dead animals left out as ordinary refuse ‘will certainly not enhance the esteem with which the research center or school is held by the public’).” Likewise, other institutions, such as the food and entertainment industries, expend effort to conceal the violence that is characteristic of the institutions—as well as the degree of that violence. The unwarranted cruelty that ensues is of a different magnitude than the cruelty that is already expected. For instance, in the book Animal Factories, Jim Mason and J. A. Keller have published images of what animals endure in the food industry. Although we know pigs, cattle, and chickens all die to be processed as food, does the public know that pregnant sows are confined until they give birth and are then immobilized until their piglets are weaned? That live chickens are carried upside-down by their feet on their way to be slaughtered? Countless appalling videos can be found with a simple Google search that shows the unabashed handling of animals before they are killed for food—for example, calves being dangerously flung into much-too-small trailers for transport to a different location.

As media becomes more available through multiple platforms, hiding such inhumane treatment is becoming more difficult; however, many refuse to acknowledge the cruelty that is so apparent—people would rather remain in the

---

1 Singer, Animal Liberation.
2 Singer, Animal Liberation.
3 Mason and Singer, Animal Factories.
dark than accept that the animal they are eating underwent horrific suffering from the moment it was born to the moment it was killed. It is unethical to consciously deny what we all can see, and additionally unethical for institutions to deliberately conceal their practices.

**Christianity and Human Responsibility**

A significant theme in the Bible is the request to be gentle and to avoid violence; copious verses warn of the evils that violence generates, in addition to verses that acclaim gentleness. Although these verses may be focused on interpersonal interactions, this theme is not restricted purely to humans. Christians are called to love and respect God’s creation—the entire earth—as we have a responsibility on this planet to do. Even though the Bible makes strong distinctions between the flesh of humankind and the flesh of other creatures, we are nonetheless asked to rule responsibly over the earth and its inhabitants. Importantly, this is where many fail to recognize the definitive discrepancy between having “dominion” (Genesis 1:28) and the exploitative domination of the earth and its nonhuman beings. In fact, when taken into context, God’s grant of human dominion over all else is actually a mandate for humans to care for and value all that God created, all that God declared as “good” (Genesis 1:25).

Importantly, many Christians refer to the numerous Bible passages that discuss meat eating in order to defend this practice: if God specifically expressed that animals were to be used for food, how could this be unethical? For instance, Genesis 9:3 conveys God’s declaration that “every moving thing that lives shall be food for you. And as I gave you the green plants, I give you everything.” From such passages, I acknowledge that God made man to be omnivorous and that certain species may have been created for the purpose of food. Many other factors to consider are constantly overlooked, however! First is the question of necessity. In biblical times, humankind obviously had significantly less access to basic necessities such as food and clothing, and the use of animals in these aspects thus greatly contributed to human flourishing, yet current technologies and capacities allow this use to be nonessential, so should such permission by God still be used to defend the killing of animals today? More notably, we must recognize the extent to which we have taken God’s consent to use animals for human benefit. Would God

---

appreciate the violent murder of various species in the name of fashion and vanity? Would God approve of the shocking viciousness of animal maltreatment in the food industry?

Christians have taken biblical passages to defend unethical behavior, manipulating the text to support certain their own agendas. For instance, the discussions of slaves and masters in the Bible was commonly construed as an approval from God that slavery is a righteous institution. Animosity toward those who diverge from Christian convention, such as homosexuals, individuals of other faiths or customs, or merely Christian people of color, has been and continues to be observed. Historically, White Christian males asserted themselves to be the undisputed authority over other races, women, nonhuman animals, and the earth—over all of God’s creation. Illustrated by European colonialism, enslavement of African Americans, and established sexist ideals, it is clear that those who considered themselves to be the most concerned with biblical thought in the past were also the most oppressive.

The rhetoric employed to describe Native and African Americans was often characteristic of dehumanization; nonwhite individuals were depicted and perceived as “savages” and “animals.” This sentiment exposes the juxtaposition that because other races were viewed as animalistic, White males were the unmistakable definition of “human”; thus, notions of humanity are built upon a foundation of oppression, discrimination, and fallacious interpretations of superiority. Syl Ko addresses the connection between animality and racism in Aphro-ism:

As authors of the racial framework, Western white men conceived of themselves as the representatives of humanity. They were the objects of morality and law, and, not coincidentally, the subjects that dictated how we should think about notions such as morality, law, and justice. Their notion of “the animal”—construed under their white supremacist framework as “subhuman,” “nonhuman,” or “inhuman”—is the conceptual vehicle for justified violence. ... Since racism requires this notion of animality, since racism and race-thinking would fail to make sense without animality, those of us interested in resisting or combatting racism need to take seriously why the status of “the animal” is what it is.5

Importantly, these White Christian males did not recognize how such oppression opposed the major teachings of their own religion; Jesus’s radically compassionate teachings and actions disregarded any apparent inequalities between

---

5 Ko and Ko, Aphro-ism.
the people with whom he interacted. As Jesus expressed, “For even the Son of Man did not come to be served, but to serve, and to give his life as a ransom for many” (Mark 10:45), and as Paul the Apostle later emphasized, “There is neither Jew nor Greek, there is neither slave nor free, there is no male and female, for you are all one in Christ Jesus” (Galatians 3:28). Jesus and his followers taught of the necessity of “love, joy, peace, patience, kindness, goodness, [and] faithfulness” (Galatians 5:22), yet the historically powerful individuals have chosen their power over the practice of these teachings. To recognize every person as equal in Christ would be a forfeit of status and power; therefore, artificial and unjustified hierarchies persevered, and still remain.

Nonetheless, the concept that is consistently emphasized throughout Christian tradition and scripture is love, which is directly referred to as the most important virtue to exercise. Despite this seemingly obvious element of Christianity, many have turned away from this virtuous way of thinking and behaving, and have used the Bible (among other resources) to assert certain claims that directly oppose the greatest commandments: “Love the Lord your God with all your heart and with all your soul and with all your mind. This is the first and greatest commandment. And the second is like it: Love your neighbor as yourself. All the Law and the Prophets hang on these two commandments” (Matthew 22:37–40). An emphasis should therefore be placed on love and gentleness rather than on engineering scripture to promote harmful notions of exploitation and objectification.

Furthermore, it is important to note that many of the prominent figures of the scientific revolution (e.g., Hooke, Boyle, Bacon, Faraday, Copernicus, Kepler, Galileo, Pascal, Newton, Mendel) and the contemporary philosophy period (e.g., Rousseau, Locke, Descartes, Kant) considered themselves to be Christian. Their findings and teachings remain incredibly influential, as much of modern science and philosophy is rooted in these White Christian males’ revelations. Consequently, certain views of moral responsibility, superiority, and humanity itself originated purely from the privileged members of society. In contrast, Jesus gave special attention to the poor and the marginalized, as he himself was one of the marginalized—a poor Palestinian Jew living under Roman occupation, with no political power beyond that of social influence. In many instances, he denounced the evils of oppression and viciousness, and was the exemplar of virtue: “The Spirit of the Lord is upon me, because he has anointed me to proclaim good news to the poor. He has sent me to proclaim liberty to the captives and recovering of sight to the blind, to set at liberty those who are oppressed” (Luke 4:18).
In his *The Spirit of Soul Food*, author, pastor, and theology professor Christopher Carter evaluates Jesus’s teachings in the context of the exploitation of human and nonhuman animals, and reveals the importance of considering all species, for those who identify as Christian as well as those interested in ethical behavior and equality between genders and ethnicities.⁶ He explains:

Our relationship with nonhuman animals should be influenced by Jesus’ teaching that those who are the peacemakers will be blessed (Matthew 5:9) and his refusal to use violence to prevent his arrest (Luke 22:49–51). In the last example, Jesus shows us that human beings should not harm other living beings. Harming another being or ending another living being’s life when it is unnecessary for our survival allows the logic that justifies an oppressive hierarchal relationship between one group over another to persist. In this way, the killing of nonhuman animals in general and specifically the killing of nonhuman animals for food perpetuate the logic of oppression that has justified the exploitation of non-white human beings, women, the GLBTQ community, and the environment. Thus, if we desire to eliminate oppressive hierarchal relationships, imitating Christ means that our plates and bowls should look drastically different than they may typically do.

As such, the defending of nonhuman animal exploitation on the basis of our “humanity,” or our “superior” hierarchical status promotes the same reasoning implicated in oppressive and exploitative human relationships—a major source of evil and viciousness in past and current Western societies.

Accordingly, although the Bible discusses killing nonhuman animals for the purpose of food or sacrifice, there is no biblical defense for the infliction of suffering and cruelty unnecessarily. As we are created in God’s image, humanity’s greater status than that of other creatures does not discount those creatures’ innate value and goodness, and we have a duty to respect all of God’s creation. The presence of evil in this world was brought about by human misjudgment, and animals are subject to these evils just as we are—even though they are generally helpless against human-inflicted violence. The Bible stresses the importance of peace and consideration while simultaneously denouncing violence and viciousness; thus, I contend that a Christian theology premised on love, justice, and care for “the least of these” (Matthew 25:40–45) is in favor of an ethical reevaluation rather than the continuation of animal maltreatment.

---

⁶ Carter, *Spirit of Soul Food.*
Scientific Considerations

*Anthropocentrism* is defined as the notion that humankind is the central or most important element of existence, particularly in comparison to other species or entities.7 This is a philosophical thought that often refers to an excessive elevation of human importance and is not typically incorporated into basic human rationale. Nevertheless, most, if not all, people unquestionably rank humanity as the ultimate superior being. This reasoning has been left unexamined, largely because it seems to be factual, yet in order to avoid assumptions, we must ask what criteria are used to determine human superiority. I observe that the essential measurement is manifested in human intelligence but also implicates our physical form, religious ideologies, emotional and insightful abilities, and many other distinguishing qualities that are thought to be restricted to humans. Even though copious research has been conducted to juxtapose humans and animals, the founding thought has remained that humans are smarter and more rational than, and therefore superior to, all other species. Before discussing the ethical concerns of assuming that intelligence corresponds to supremacy, I will first address comparative research that attempts to create a division between human and nonhuman animals.

The study of animal cognition, or the study of the mental capacities of nonhuman animals, is a significant field in the behavioral sciences. Interestingly, this study falls under a category of psychology called *comparative* psychology. This indicates that research of animal intelligence is fundamentally conducted through comparison—hierarchizing the mental abilities of nonhuman animals against those of human animals—which includes a significant problem found in the origin of comparative psychology. C. Lloyd Morgan was a 19th-century British ethologist and psychologist known for constructing an experimental approach for the study of animal cognition that prevails today: Morgan’s Canon. In his *An Introduction to Comparative Psychology*, Morgan states, “In no case is an animal activity to be interpreted in terms of higher psychological processes if it can be fairly interpreted in terms of processes which stand lower in the scale of psychological evolution and development.” 8 Basically, always assume there is a primitive explanation for an animal’s behavior; anthropomorphizing animal cognition will yield false conclusions. This basis of studying animal intelligence generates preceding biases that are left unaccounted for. If researchers begin with the expectation that higher intellectual processes are exhibited only by humans and

---

7 Boslaugh, “Anthropocentrism.”
8 Morgan, Comparative Psychology.
that their results must align with Morgan’s Canon, how can they conduct fair tests? Even though the technology currently used to test brain activity (in any species) has drastically advanced, we still lack a method that can identify what animals are truly thinking. We cannot empirically discern process (mental functioning) and performance (what is observed externally) in an experimental setting. To this end, our knowledge of animal intelligence, feelings, cognition, consciousness, and the like is lacking, rendering such defenses of ultimate human superiority irrelevant.

Furthermore, the concrete neurological and cognitive differences between humans and other animals are questionable. The brain-to-body size ratio is noticeably larger in humans than in other species, and neuroscientists have uncovered many other factors that contribute to the uniqueness of human intelligence and cognition. For example, humans have approximately 11.5 billion more cortical neurons than do most other mammals, meaning that regardless of brain size, the human brain exhibits the most neuronal density of any species. Furthermore, the myelination of nerve cells in humans is much thicker than that in other animals. Myelinated nerve cells allow for electric impulses to be sent quickly along axons in the brain; thus, the speed at which information travels through the brain is significantly faster for humans than for other species. Although numerous qualities demonstrated by the human brain set it apart from other species, such qualities are not entirely exclusive to humans. All primates, for instance, exhibit similar-diameter myelin sheaths, and whales and elephants have only about half a billion fewer cortical neurons than do humans, which is numerically close enough to discount any resulting critical cognitive differences. Furthermore, brain size has been found to be an inappropriate measure of overall intelligence, in that certain small-brained species, such as birds and rodents, display more cognitive complexity than do larger-brained species, such as horses and cows. Ultimately, qualities of the human brain that appear to suggest human uniqueness and intellectual superiority may be additionally found in other species or may be altogether obsolete.

Other intellectual capabilities that humans possess are often incorrectly assumed to be restricted to humans. For example, humans are not the only species capable of forming social relationships, exercising complex problem-solving skills, and finding and using external tools. New animal-behavior research has uncovered that some cognitive skills are more advanced in other species than in humans, including certain memory abilities, adaptability, and flexible learning strategies. Researchers have conducted many cognitive tests in which animals outperform their human counterparts, so why do we remain certain of our complete intellectual
supremacy over animals? As Ursula Dicke and Gerard Roth of *Scientific American* state in the article “Animal Intelligence and the Evolution of the Human Mind”:

> As far as we know, no dog can compose music, no dolphin can speak in rhymes, and no parrot can solve equations with two unknowns. Only humans can perform such intellectual feats, presumably because we are smarter than all other animal species—at least by our own definition of intelligence. … The lack of an obvious structural correlate to human intellect jibes with the idea that our intelligence may not be wholly unique: studies are revealing that chimps, among various other species, possess a diversity of humanlike social and cognitive skills.⁹

As our knowledge of neuroscience advances, many aspects of a seemingly definite line between human and animal intelligence continue to become blurred.

Importantly, intelligence continues to serve as the major measurement for one’s ranking within a dominance hierarchy, defined in animal behavior as “a form of animal social structure in which a linear or nearly linear ranking exists, with each animal dominant over those below it and submissive to those above it in the hierarchy.”¹⁰ This phenomenon is not restricted to nonhuman animal behavior but is also an innate factor of sociability. Philosopher Jean-Jacques Rousseau’s revolutionary work *Discourse on the Origin and Basis of Inequality Among Men* acknowledges that social living introduces artificial inequalities.¹¹ He addresses a form of self-love that is unnatural and hierarchical: *amour-propre*. This essentially refers to one’s desire to be “better” than others, the desire to be ranked higher in the established social order. As a result, inequalities are developed concerning whom can be used to what purpose and how valuable their contribution is in the wider community. Rousseau explains, “Here are all natural qualities set in action, every man’s rank and fate set, not only as to the amount of their goods and the power to help or to hurt, but also as to mind, beauty, strength or skill, as to merit or talents, and, since these are the only qualities that could attract consideration, one soon had to have or to affect them; for one’s own advantage one had to seem other than one in fact was.” Because of the inflation of comparative thinking that arises from sociability, Rousseau recognizes man in society as an unfree being that only finds meaning in this relational self-love, *amour-propre*.

---

⁹ Dicke and Roth, “Animal Intelligence.”
¹⁰ Editors of Encyclopedia Britannica, “Dominance Hierarchy.”
¹¹ Rousseau and Gourevitch, *Discourses*. 
We must acknowledge the consequences of the inclination to hierarchize on the basis of intelligence. Today, the theory of prominent philosopher Aristotle regarding “natural slaves,” or individuals born lacking certain intellectual qualities, is widely discredited. In Book I of his Politics, Aristotle argues, “For he is a slave by nature who is capable of belonging to another—which is also why he belongs to another—and who participates in reason only to the extent of perceiving it, but does not have it.” Furthermore, Thomas Aquinas held a similar view, as stated in his Summa Contra Gentiles: “For men of outstanding intelligence naturally take command, while those who are less intelligent but of more robust physique, seem intended by nature to act as servants.”

To be sure, it was not uncommon for past figures to defend the institution of slavery with the argument that “natural slaves” are revealed through certain intellectual weaknesses and that such weaknesses can be compensated for by their submission to other, smarter, people. This argument was not restricted to Aristotle’s and Aquinas’s generations and has been utilized by more recent figures. American psychologist and author Arthur Jensen, who taught at the University of California, Berkeley, and died in 2012, conducted numerous studies regarding racial differences in intelligence. In his book The g Factor: The Science of Mental Ability, Jensen discusses variability in the g factor, or the general intelligence factor, among White and African American individuals: “The relationship of the g factor to a number of biological variables and its relationship to the size of the white-black differences on various cognitive tests (i.e., Spearman’s hypothesis) suggests that the average white-black difference in g has a biological component. ... Racial populations differ in many genetic characteristics, some of which, such as brain size, have behavioral and psychometric correlates, particularly g.” Jensen published numerous works claiming that some races are inherently more intelligent than others, which have been determined to be malicious and incorrect. Indeed, the general academic population recognizes comparable oppressive works to be condemnable, yet intelligence (along with wealth and other artificial values that result from intelligence) still evidently determines the worth of both human and nonhuman animals alike.

Although we can clearly note the ethical violations present in such works and mentalities, using scientific studies pertaining to intelligence to advocate for the ranking of worth among species remains incredibly prevalent. Peter Singer

---

13 Aquinas, Contra Gentiles.
14 Jensen, The g Factor.
articulates this correlation in his *Animal Liberation*: “‘Speciesism,’ by analogy with racism, must also be condemned. … If possessing a higher degree of intelligence does not entitle one human to use another for his or her own ends, how can it entitle humans to exploit nonhumans for the same purpose?” Regardless of humanity’s accepted higher degree of intelligence, intellect does not correspond to freedom or amount of moral consideration due. Furthermore, in reference to scientific fact, we know that animals have the abilities to be scared, to feel pain, and to protect their kin. In other words, humans are aware that animals are capable of feeling and thus of suffering.

It is important to note Jeremy Bentham’s acknowledgment of such reality:

The French have already discovered that the blackness of the skin is no reason why a human being should be abandoned without redress to the caprice of a tormentor. It may one day come to be recognized that the number of the legs, the villosity of the skin, or the termination of the os sacrum are reasons equally insufficient for abandoning a sensitive being to the same fate. What else is it that should trace the insuperable line? Is it the faculty of reason, or perhaps the faculty of discourse? But a full-grown horse or dog is beyond comparison a more rational, as well as a more conversable animal, than an infant of a day or a week or even a month, old. But suppose they were otherwise, what would it avail? The question is not, Can they reason? nor Can they talk? but, Can they suffer?”

The ethical weight of a being’s capacity for suffering is much greater than the weight of a being’s level of intelligence, yet we still employ nonhuman animals’ lack of intelligence (a premise that has yet to be empirically defined) as reason enough to harmfully exploit other species.

**Human Morality**

This paper serves as a confrontation of the negligible moral norms applied to nonhuman animals. While there is no complete, explicit definition of morality, the most agreed-upon theories from philosophers throughout history have emphasized that vicious, harmful, unnecessarily cruel acts are unethical. Such acts can be measured by the degree or amount of suffering they produce, and, as articulated by C. S. Lewis in his anti-vivisection argument, “whenever pain is inflicted it requires justification.”

---

15 Bentham and Mill, *The Utilitarians*.
16 Lewis and Walmsley, C.S. Lewis Essay Collection.
cutting of or operation on a living animal usually for physiological or pathological investigation” or, broadly, “animal experimentation especially if considered to cause distress to the subject.” Accordingly, there is a difference between quickly killing livestock for the purpose of food—the death can happen in an instant and the suffering is quickly over—and the inhumane treatment of animals from the moment they are born to the moment they are relieved of the harsh, agonizing lives they have been subject to by the hand of “moral” agents.

In addition to intelligence, morality itself has been referred to as yet another element of humanity that separates us from other animals. Presumably, humans are the only beings who are aware of morality and are moral agents—defined as people capable of distinguishing right from wrong, or morally responsible or blameworthy, for behaving immorally. Immanuel Kant developed Kantian ethics, in which moral agents have a duty to act in accordance with moral law. Specifically, he writes, “An action done from duty has its moral worth, not in the purpose to be attained by it, but in the maxim according with which it is decided upon; it depends therefore, not on the realization of the object of action, but solely on the principle of volition in accordance with which, irrespective of all objects of the faculty of desire, the action has been performed.” Furthermore, Kant contends that moral agents must be rational, autonomous beings: “The will is conceived as a faculty of determining oneself to action in accordance with the conception of certain laws. And such a faculty can be found only in rational beings.” Notwithstanding our total uncertainty regarding whether nonhuman animals (or even human animals, for that matter) have free will, virtue ethics predominately concludes that humans have a certain moral responsibility to choose to do what is virtuous and to refuse the adverse, immoral choice. As Aristotle conveys in The Nicomachean Ethics, “Virtue lies in our power, and similarly so does vice; because where it is in our power to act, it is also in our power not to act.” To this end, do we, as moral agents, choose to behave ethically in our treatment of animals? Likewise, do we exercise our autonomous abilities and knowledge of morality to refuse unethical treatment of animals? As Peter Singer identifies:

We rarely stop to consider that the animal who kills with the least reason to do so is the human animal. We think of lions and wolves as savage because

---

17 Merriam-Webster Dictionary.
18 Rosen, “Responsibility and Moral Ignorance.”
19 Guyer, Kant’s Groundwork.
20 Kant and Abbott, Fundamental Principles.
21 Welldon, Nicomachean Ethics.
they kill; but they must kill, or starve. Humans kill other animals for sport, to satisfy their curiosity, to beautify their bodies, and to please their palates. Human beings also kill members of their own species for greed or power. Moreover, human beings are not content with mere killing. Throughout history they have shown a tendency to torment and torture both their fellow human beings and their fellow animals before putting them to death. No other animal shows much interest in doing this.22

Sadly, the moral implications of human treatment of nonhuman animals are consistently avoided; many would rather ignore these truths than consider ethical treatment of animals.

In terms of human morality, there is a common psychological disconnect in which we perceive our overall character to be good, regardless of our individual actions. Few people evaluate their sins or unethical behavior and decide, “I often behave immorally, so I must be a bad person.” I believe this disconnect to be a consequence of guilt, which is defined in psychology as “a self-conscious emotion characterized by a painful appraisal of having done (or thought) something that is wrong and often by a readiness to take action designed to undo or mitigate this wrong”23; however, likely because animals cannot outwardly express their pain and suffering in the same way humans can, the dimension of guilt that encompasses wanting to “undo or mitigate this wrong” is not experienced in the majority of humans. Interestingly, it is important to consider whether, if animals acquired the ability to speak, we would still treat them as we do. Is the fact that animals cannot directly exclaim, “I am in pain!” the determining factor that supports the continuation of human-inflicted violence? Although verbal language is a critical feature of human uniqueness, we know that animals experience pain as we do; in fact, much of what we know regarding the somatosensory responses to painful stimulation was obtained from (painful) research conducted on animals! Because guilt is an uncomfortable feeling, and because animals cannot demand their own liberation as humans can, their unprovoked suffering therefore remains acceptable.

Likewise, when animal ethics is addressed, opposition is a typical response. For instance, Peter Singer’s Animal Liberation, a significant resource in the deliberation of animal ethics, received widespread objection after its publication. Animal ethics, in itself, is rarely considered as a valid discipline by those in power, at the top of the socially constructed dominance hierarchy, those historically

22 Singer, Animal Liberation.
23 VandenBos, Dictionary of Psychology.
oppressive individuals. Arguments against the moral consideration of animals are manifested in animals’ assumed irrationality, lack of moral perspective, and lack of free will. For example, philosophy professor Tibor Machan contends in his work *Putting Humans First: Why We Are Nature’s Favorite*:

So far, there is no clear evidence that any species other than human beings reside at the top of the natural hierarchy. When human beings emerge in the natural world, so does the capacity to think and exercise self-initiative. That is why Aristotle calls us ‘the rational animal’. … Animals, furthermore, have no central, crucial need of thinking, whereas human beings cannot begin to survive without thinking. And unlike animals, human beings cannot count on instincts to guide them automatically. For those lower animals that do exhibit some rudimentary cognitive capacity, it is very much a side issue, elicited usually by human beings in highly circumscribed and unusual circumstances (such as laboratories).

As such, critics of animal ethics habitually establish their reasoning in humanity’s superiority: because we exhibit rationality, morality, and autonomy, we are to be ethically considered as distinct from nonhuman animals.

Conversely, the assertion that human beings are the only beings that exhibit rationality is not factual, as we do not have evidence that confirms nonhuman animals to be definitively irrational. Moreover, by characterizing the human species as rational, we dismiss the entirely *irrational* behaviors and attitudes that humans exhibit. In an interview with the American Psychological Association’s Kirsten Weir, psychologist and behavioral economist Dan Ariely discussed human irrationality and dishonesty, explaining, “Classical economics assumes human beings behave rationally. But if you believe everyone is rational, and you look at humanity and see how much misery there is in the world, then you have to conclude that this is the best we can do ... the world isn’t like this as an outcome of the decisions of eight billion rational people. It’s like this as the outcome of eight billion irrational people.” Arguably, humans are the only species that knowingly and unnecessarily harm themselves and others without true reason.

Furthermore, in acknowledgment of the significant vices of the human species, we must note the driving forces of many of our actions: greed, gluttony, arrogance, and apathy. In *Environmental Virtue Ethics*, Philip Cafaro speaks of these vices and of how humanity’s submission to these vices opposes reason and

24 Machan, *Putting Humans First.*
25 Weir, “*Totally Irrational.*”
actually harms the entire planet. In reference to traditional philosophical thought, Cafaro writes,

The tradition sees vice as contradicting and eventually undermining reason, hence destroying our ability to understand our proper place in the world and act morally. Aristotle expresses this in his distinction between incontinence (the tendency to pursue pleasure even when we know it is wrong to do so) and the full-blown vice of intemperance (where the continued pursuit of illicit pleasure has so clouded our judgement that we no longer recognize right from wrong). The vices are habits of thought and action. Left unchecked, they tend to cloud reason, the voice of both conscience and prudence.26

Regarding humanity’s continual destruction of the planet, he later exposes our refusal to accept that in harming the planet, we are likewise harming ourselves: “We falsely assume that we can keep separate harms to nature and harms to humanity, harms to others and harms to ourselves. We do not see that environmental vices do not just harm nature; they harm us and the people around us.” Driven by greed and selfishness, those in power often succumb to vices and irrational behaviors that directly cause universal damage: harm to our planet, to nonhuman animals, and to humanity. In claiming that humans are the only rational, moral creatures, we also must admit that we are irrational and immoral, as our knowledge of virtue requires our immoral actions to be the choice to act immorally rather than virtuously. The aspiration for improvement within our species must be accompanied by a conscious decision to choose virtue over vice; otherwise, our immoral presence on the earth will continue to cause destruction.

Virtuous Treatment of Nonhuman Animals

Again, I contend that sentient beings do require moral consideration beyond what they are currently given. No number of discrepancies that permit the human species to govern the rest of the world’s inhabitants can excuse the institutionalized abuse of nonhuman—but still thinking, feeling, valuable—beings. In fact, these very discrepancies bestow on us the moral responsibility to consider improvements, to strive for virtue in the treatment of all beings.

Indeed, certain consequences occur when virtuous thought is applied to the treatment of nonhuman animals. Practices that have become traditions are threatened, and an overall human mentality is put into question. Nonetheless, the

26 Sandler and Cafaro, Environmental Virtue Ethics.
progressions of Western culture have historically yielded beneficial outcomes, and an acceptance of ethical treatment of animals should be viewed as another triumph over oppressive attitudes. Similar to the abolitionist movement or the feminist movement, animal liberation shares an end of challenging oppression, objectification, and exploitation. Moreover, we must acknowledge what is actually at stake in the moral consideration of other beings. Are the majority unwilling to sacrifice their ability to eat animal meat? Will scientific research be hindered by the moral consideration of animals? Does animal ethics minimize the value of humanity? No, the application of moral norms to nonhuman species does not require the use of nonhuman species to be entirely eliminated and does not imply that they are equal to humans. Instead, in terms of nonhuman and human animal relations, we should be asking, How can we reshape our current standards of animal treatment virtuously? When and how is the exploitation of animals justified? What constitutes ethical treatment? Rather than being viewed as a threat to human flourishing, it should be perceived as another way in which we each can exercise virtue.

Virtue ethics encompasses a purpose of cultivating virtuous habits in order to identify how one can grow and become a better person. Ethicists emphasize that consistency is an essential component of morality; one’s ethical reasoning must be reliable in all of one’s decision-making. When one alters one’s ethics case by case, one is not practicing virtue but instead allowing oneself to behave in any way that is of direct benefit to oneself. As emphasized by Kant’s categorical imperative “act only according to that maxim whereby you can at the same time will that it should become a universal law,”27 sinful acts, even those that appear to be harmless, are nonetheless sinful. By this reasoning, purchasing animal meat is an action that supports the institutionalized cruel treatment of animals, and while it may be an indirect method of support that appears to be harmless, consumers of animal meat are the direct cause of animal slaughter. In practice, discern where and how the “meat” you buy lived; was it factory-farmed or truly pasture-raised? Are your eggs actually free-range, or do they merely adhere to the weak standards set by the USDA? If vegetarianism or veganism is not a practical lifestyle for monetary, dietary, or other personal restrictions, spending a few extra dollars and minutes of research to practice more-ethical meat consumption can be a great contribution toward fostering virtuous treatment of nonhuman animals. The cascade of events that follows consumer habits works fast, and even individual practices slowly but surely will alter the destructive principles of the Western food industry. In

27 Guyer, Kant’s Groundwork.
Conclusion

In consideration of the recent breakthroughs in such areas as modern science, technology, and communications, the Western mindset commonly reasons that the current status of humankind is superior to that of past generations. There is an unrecognized yet prevalent bias toward modern times that increases alongside such advancements, especially within the younger generations of people. Many traditions of the past are becoming outdated, and younger individuals are frequently drawn to breaking the old status quo in favor of, hopefully, progressive outlooks and acceptance.

I recognize the importance of an ethical reflection in terms of our relationship with other species. Do we want to be the generation of people that puts an end to the unnecessary human-inflicted animal violence? Or do we want this singular aspect of modern life to remain unchanged and unevaluated from a moral perspective?

I suggest that we use virtue ethics to discern appropriate treatment of animals because certain institutionalized evils are completely unnecessary yet are customary because the vast majority of people do not directly experience this violence. In biblical times, people had a different relationship with the animals they killed for food, clothing, and sacrifice. They raised their own flocks and killed them with their own hands, and this was done under God’s instructions (“you may slaughter animals from the herds and flocks the Lord has given you, as I have commanded you” [Deuteronomy 12:21]), in the kosher way that involves using a knife to sever the trachea and esophagus—the quickest and most painless method, in which the animal does not endure much suffering at the time of its death.28 Now, however, animals intended to be meat suffer their entire lives and are not known to their consumers as living beings but are always and exclusively considered meat. Even though many would argue that they disagree with the institutionalized mistreatment of animals, they are under the illusion that change is not within their power. People therefore continue to purchase what is at their fingertips—an already dead animal that they cannot do anything to save. Of course, true change will not happen instantaneously, but the importance of determination for a greater purpose

28 Kesselman, Guide to the Law.
is essential for humanity’s moral growth. Martin Luther King Jr. acknowledged this in a speech at New York University in 1961: “Human progress is neither automatic nor inevitable. ... Every step toward the goal of justice requires sacrifice, suffering, and struggle; the tireless exertions and passionate concern of dedicated individuals.” Clearly, his cause for justice and equality among human individuals requires much more sacrifice, suffering, and struggle than a new consideration of ethical animal treatment will require.

Humans’ demonstrable superiority in our capacity to affect the conditions in our world does not imply a superiority in the intrinsic value of human lives. That we exhibit certain uniqueness in our intellectual capabilities does not grant us clearance to discount other lives as insignificant. The infliction of suffering on the basis of superiority is not an ethically sound notion but is a submission to oppressive ideals that have historically plagued our civilization. Western societies do not rely on the exploitation of animals to flourish, and thus, the unjustified violence that ensues is entirely immoral. We are the cause of animals’ suffering, but to what end? What benefits do we achieve that outweigh the degree of suffering that animals are subject to?

Because the current intentions of animal killing are fixed in profit rather than necessity, however, and because the majority of humans have disconnected themselves from the consequences of this violence, practices characteristic of unwarranted cruelty to nonhuman animals continues and the immoral practices are not widely opposed. Those who have directly resisted the unethical treatment of animals are not given sufficient attention, because most people would rather avoid this reality than admit that major customs, and their own behaviors, are wrong. Biblical defenses of animal killing have become significantly obsolete in industrialized countries, yet the central principles of Christian theology—love, justice, compassion, gentleness—are as necessary to embrace as ever. It is essential to make an effort in cultivating habits that produce virtue rather than evil in this time when each action affects so many others around us, both human and nonhuman. We must seek to end the alienation of animals in our culture, in order to resist the vices that such alienation continues to generate.

The maltreatment of animals perpetuates the same logic that was, and still is, used to justify the mistreatment of human beings—which is unmistakable in the factory farming industry. Significantly, the employees of these establishments are typically people of color and are typically treated as less than human. Charlie

29 “Realizing a Dream.”
LeDuff reports on these inhumane occurrences in his *New York Times* article “At a Slaughterhouse, Some Things Never Die.” LeDuff conveys the terrible working conditions that the immigrant, marginalized, and desperate workers undergo; as one worker is quoted as saying, “This job’s for an ass. They treat you like an animal.” The factory farming industry outwardly seeks out those in need; it prevails in its violence by pursuing those who feel stuck in marginalized positions. LeDuff explains that in the slaughterhouse, “the turnover is 100 percent. Five thousand quit and five thousand are hired every year. You hear people say, They don’t kill pigs in the plant, they kill people. So desperate is the company for workers, its recruiters comb the streets of New York’s immigrant communities, personnel staff members say, and word of mouth has reached Mexico and beyond. The company even procures criminals. Several at the morning orientation were inmates on work release in green uniforms, bused in from the county prison.” Clearly, this institution as a whole not only propagates violence toward animals but also engages in the dehumanization and inhumane treatment of human beings. In contemplating our civilization’s capacity for virtue, we must acknowledge that this major industry is an immoral industry.

Furthermore, the routine acts of violence in the agriculture industry not only include immoral treatment of the animals but are also psychologically damaging to the associated employees. James McWilliams, history professor and author of *Just Food: Where Locavores Get It Wrong and How We Can Truly Eat Responsibly,* writes about this phenomenon in his article “The Dangerous Psychology of Factory Farming.” He evaluates the current practice of killing animals for food in comparison to how this practice was conducted in the past: “Before 1850, when most animal husbandry happened on a relatively small scale, farmers viewed their animals as animals. That is, they saw them as sentient beings with unique needs that, left unaddressed, would result in an inferior product.” The post-1850 mindset that emphasizes profitability above all else in the animal agriculture system dismisses the acknowledgment of the animals as animals, viewing them instead as lucrative objects. Of this, McWilliams reports,

> Beginning with plants, and then moving to animals, they became less concerned with individual idiosyncrasies and more concerned with collective evaluations of productivity. The chain of production expanded, and, as it did, farmers came to speak in terms of nutrient input, breeding schedules, confinement space, and disease management. By the 1870s,

---

30 LeDuff, “At a Slaughterhouse.”
31 McWilliams, “Dangerous Psychology.”
farmers were regularly referring to their animals not as animals but, literally, as machines being built in factories. … [Factory farming’s] impersonal, highly rationalized structure is designed to protect those involved from the emotional consequences of killing.  

Because of the objectification of animals in the agriculture industry, farmers do not have to face the reality of animal slaughter but can train their minds to recognize living beings as objects, machines, and revenues. This notion alone reveals the extent to which humans avoid ethical treatment of animals. That factory farmers knowingly have to objectify and demoralize the animals they slaughter in order to preserve their own mental well-being exposes that this industry promotes the cultivation of evil rather than virtue, disregarding both the factory farmers themselves and the animals they work with.

In accordance with theological and philosophical thought, the ethical treatment of animals must entail practicing consistently distinguished virtues. In particular, Plato and Aristotle emphasized prudence, courage, temperance, and justice as the four cardinal virtues. Rousseau identified pity, or compassion, as the essence of human morality. Jeremy Bentham taught of the weighing of pains against pleasures, and Kant explained how to exercise goodwill in his categorical imperative. Among these past philosophers, as among many other virtue ethicists, the concept of morality is defined in multiple ways, yet none of these methods of exercising virtue is sufficiently applied to the treatment of animals. Furthermore, Christian ethics maintains that Jesus is the embodiment of righteousness and is the ultimate role model for humanity, and that love, joy, peace, patience, kindness, goodness, faithfulness, gentleness, and self-control are fundamental characteristics. As Philippians 4:8–9 states, “Whatever is true, whatever is honorable, whatever is just, whatever is pure, whatever is lovely, whatever is commendable, if there is any excellence, if there is anything worthy of praise, think about these things.” It is not difficult to see that our consideration of nonhuman animals does not align with any of these virtues or methods of practicing virtue. We must therefore reflect upon our moral obligations to nonhuman beings and discontinue the institutionalized, unjustified animal violence that hinders our society’s capacity for virtue.

Virtuous behavior toward nonhuman animals would entail a complete reconsideration and elimination of the unwarranted institutionalized cruelty and violence involved in the exploitation of other species. In refusing the oppression of animals, we are refusing this major vice that has historically harmed the human

---

32 McWilliams, “Dangerous Psychology.”
members of our communities, just as it continues to harm the nonhuman members. Those who care about exercising virtue or imitating Christ in their lives should confront the immorality that accompanies the exploitative oppression of animals and should subsequently consider changing their diets to reflect their priorities. Because virtue is an action, we must be active in our practice and recognition of virtue—active in our moral consideration of both human and nonhuman beings.
Bibliography


Abstract

In this paper, I argue that John Milton’s writings demonstrate a subtle preoccupation on the part of their author with the inherent limitations presented by various forms and conventions. In addressing and challenging the restrictive nature of poetic conventions, the human form, ontological conceptions, perceived gender binaries, and poetic verse itself, I find that Milton’s writings illustrate an undercurrent of rebellion against the imposed limitations and conventions of the mediums they inhabit, as well as a fascination with the idea of form free from confinement. The lines and construction of the majority of Milton’s English sonnets and his timeless lapsarian epic, Paradise Lost, as well as other works such as Areopagitica, “Lycidas,” and “On the Morning of Christ’s Nativity,” bear significant support for a rebellion against form and convention. The implication is that Milton exercised some degree of metaliterary thought amid the composition of his works and their unprecedented defiance of standard and convention.

Before there were such well-acknowledged norm-benders as the genre-busting megastar musician Prince, the revolutionary American gender theorist Judith Butler, or the diversity-minded rap-opera theater-maker Lin-Manuel Miranda, there was John Milton. The seventeenth-century poet, polemic, civil servant, and intellectual of sustained international notoriety nursed a preoccupation with the inherent limitations that various forms and conventions present. He challenged the restrictive nature of poetic conventions, ontological conceptions, perceived gender binaries, and poetic verse, and even the human form itself. Milton’s writings illustrate an undercurrent of rebellion against the imposed limitations and conventions of the mediums they might inhabit, as well as a fascination with the idea of confinement-free form. And yet at the same time, he seemed to purposely seek out constraints, from the strict sonnet form to the highly stylized epic and pastoral genres. The majority of his English sonnets and his postlapsarian epic Paradise Lost—as well as other works such as Areopagitica, “Lycidas,” and “On the Morning of Christ’s Nativity,” although less blatantly so—
bear significant support in their lines and construction for an attraction to constraints as well as a simultaneous rebellion against typical form and convention. Milton’s simultaneous seeking of and resistance to typical forms indicate a subtle but distinct fascination on his part with norms and how to purposefully push their boundaries.

Milton’s portfolio of sonnets, twenty-four poems strong, first consisted of poems that respectfully and artfully explored Petrarchan/Tuscan sonnetary conventions. As a young man, he wrote “a fragmentary group of Italian poems” between 1629 and 1632 that intended to “explor[e], in the correct Tuscan dialect, the origins of the love sonnet” (Kerrigan et al. 137). Until 1632, the vast majority (if not entirety) of known sonnets, Tuscan or otherwise, were concerned with matters of love. That year, Milton broke poetic convention with his seventh sonnet, a self-reflective poem written in English. As Kerrigan et al. write, “Nothing […] could have anticipated Milton’s effort [in this sonnet] to gather religious strength in contemplating the end of an apparently unpromising youth” (137). Such a topic had never before been breached in sonnet form.

As Milton himself stated within the sonnet, “It shall be still in the strictest measure even” (10), and indeed, he played within the foundations of the Petrarchan sonnet with Sonnet 7 in his thematic structure and Italian language. The first quatrain introduces conflict, which—prior to Milton’s sonnets—was always romantically grounded. In this seventh Italian sonnet, though, Milton writes to a friend of his progressing age and seemingly stagnant development as a writer: “How soon hath Time, the subtle thief of youth, / Stol’n on his wings my three and twentieth year! / My hasting days fly on with full career, / But my late spring no bud or blossom shew’th” (1–4). He continues by developing this conflict in the second quatrain, then beginning the traditional sonnety sestet with a volta that introduces a pronounced change in tone—in this case, a reaffirmation of self, that he will continue to march toward whatever goal “Time” and “the Will of Heaven” lead him (12). Milton’s specific writing, while meeting the structural qualifications for a Petrarchian sonnet, could also be read as three quatrains and a couplet, which evokes the measure of Shakespearean sonnets. In Sonnet 7, Milton chronicles his woe and delay morphing into steadfastness and confidence within a poetic container previously exclusive to tales of love and loss, thus renegotiating the boundaries of sonnets and the stories within their stanzas.

The vast majority of Milton’s other sonnets further challenged poetic convention. His tenth, thirteenth through seventeenth, twentieth, and twenty-first sonnets celebrate well-known individuals and his relationships with them, while his
eighth and eighteenth sonnets provide commentary on wartime skirmishes; his eleventh and twelfth sonnets defend his *Tetrachordon* divorce tracts, and his final English sonnet, “On the New Forces of Conscience,” was written in opposition to the Presbyterians. Milton defied convention with these poems in that he appropriated the sonnet as a means of political discourse best suited to his own compositional style as a self-identified poet. “Far from standing at odds with one another, poetic creation and political engagement [thus] turn out to be intimately connected. The artistry of the sonnets keeps pace with […] Milton’s specifically political understanding of events, parties, and leaders as well as his sense of their significance to himself and his nation as political entities” (Mueller 477). Thus, I argue not that Milton challenged the structural form of sonnet poetry but that, in harnessing his dexterity as a poet to supplement his political prose with sonnets, he transformed the perceived boundaries of sonnets.

Beyond politically repurposing their poetic form, Milton also used sonnets to express explicit feelings of frustration with other forms’ limitations. Sonnets 19 and 22 both testify to Milton’s frustrations with the limitations of form, in that their lines depict Milton’s doubts and anxieties concerning his own human existence, limited by visual impairment and mortality. In both sonnets, the forty-three-year-old poet contemplates whether his developed blindness might make impossible his perceived divine responsibility to produce a literary work “unattempted yet in prose or rhyme” (1.16)—a subject that, like those touched on by the majority of his English sonnets, had never before been addressed in this particular medium. We know, however, that despite his blindness, Milton ultimately did produce such a work in *Paradise Lost*, a literary masterpiece that illustrates Milton’s form obsession perhaps more than any other piece he conceived.

*Paradise Lost* is a mutiny against form in myriad ways, one of which is its construction—that is, the medium in which John Milton produced it. In his 1672 *Artis Logicae*, Milton defines matter as “the cause from which a thing is [*causa ex qua res es*]” and form as “the cause through which a thing is what it is [*causa per quam res est id quod es*]” (*Artis* 866–67; Lieb 207). This work’s ontological exploration of form provides a logical basis for the great deal of thought that Milton seems to have put into choosing an appropriately limitless form for his intended masterpiece. Milton’s Trinity College manuscript from the early 1640s contains a series of outlines and ideas for such a work, four of which are outlines for *Paradise Lost* as a tragic drama (Kerrigan et al. 252). It seems, however, that Milton decided that a theatrical medium could not realistically accommodate Adam and Eve’s “first naked glory” (4.1115) and switched forms, opting instead for a work of “English
heroic verse without rhyme, as that of Homer in Greek, and of Virgil in Latin,” as he stated in his postpublication argument, “The Printer to the Reader.” Milton explicitly targets the limitations of form—specifically, poetic form—in this originally unintended foreword of sorts, having written and included it with his epic as a response to public outcry concerning “why the poem rhymes not”:

[R]hyme being no necessary adjunct or true ornament of poem or good verse, in longer works especially, but the invention of a barbarous age, to set off wretched matter and lame meter [...] [t]his neglect of rhyme so little is to be taken for a defect, though it may seem so perhaps to vulgar readers, that it rather is to be esteemed an example set, the first in English, of ancient liberty recovered to heroic poem from the troublesome and modern bondage of rhyming. (Paradise 291)

In his Between Worlds: The Rhetorical Universe of Paradise Lost, Will Pallister argues that the particular blank verse form Milton chose for his lapsarian epic enabled Milton to weave near-countless political, religious, philosophical, historical, mythological, literary, justice-related, and gender-related elements and references into the work’s “encyclopedic scope” while still maintaining such elements’ subordinacy to the work’s “overarching poetic agenda” and “aesthetic achievement” (241). Renaissance critical theory, too, offers support for the notion of the epic as a compendious storytelling medium. At frequent points within Paradise Lost, Milton references the similarly blind poet Homer, a Grecian whose epics were widely considered an original source for philosophy, mathematics, history, geography, military, art, and numerous other disciplines in their time (Lewalski 5). The continuous union of all these fields via the relatively unrestrictive nature of Paradise Lost’s rhetorical form aligns with Milton’s own monist ontology, a philosophy entailing that everything in the universe is constructed of the same matter—“one first matter all” (5.472)—but “endued in various forms, various degrees / Of substance” (Paradise Lost 5.473–74; Lieb 224). Milton’s monism underlies much of the anti-form sentiments expressed within the actual line and lyric of Paradise Lost.

Paradise Lost’s narrative content exhibits many challenges against form. Narrative character descriptions serve as one means by which Milton achieves the undertone of formlessness pervading his epic. For instance, the narrator begins Paradise Lost by calling on its muse, to which the narrator ascribes abilities characteristic of both genders: “Dove-like sat’st brooding on the vast abyss,” behaving as a female dove would behave (1.21), “[a]nd mad’st it pregnant,” which is an ability typically attributed only to males (1.22). The narrator’s description of
the fallen angels later in Book 1 further blurs the binary form that Milton’s contemporaries assigned to gender, simultaneously offering a restrictive interpretation of the human form through its verbiage: “For spirits when they please / Can either sex assume, or both” (1.423–24). The concrete nature of shape also lands under Milton’s microscope, as the narrator explains that fallen angels’ otherworldly essences are “[n]ot tied or manacled with joint or limb, / Not founded on the brittle strength of bones, / […] but in what shape they choose / Dilated or condensed, bright or obscure, can execute their airy purposes” (1.426–30). This shapeshifting frequently resurfaces throughout Paradise Lost with regard to the fallen angels. They shrink as they enter Pandemonium, as “[t]hey who seemed / In bigness to surpass Earth’s giant sons / Now less than the smallest dwarves” (1.777–79), and Satan demonstrates their ability to take on earthly forms when he infiltrates Eden within “his borrowed visage” of the serpent (3.154–56). In fact, Milton’s narrator goes so far as to defy the very concept of visualizable concrete form itself with the description of Death. Introduced as a shapeless, indistinguishably limbed being of equal parts shadow and substance (2.666–79), Death defies the confines of physical form altogether to such a degree that we as readers struggle to mentally envision him (Trubowitz 393).

On the other end of the heavenly spectrum, too, Milton ingrains formlessness into his epic: in its address at the introduction of Book 3, the narrator uses ephemeral but self-contradictory language to describe God and the Holy Trinity. The narrator praises two halves of the Holy Trinity simultaneously—“Hail holy light, offspring of heav’n first-born” (3.1)—in referring to the Son, but the narrator continues, “Since God is light, / And never but in unapproachéd light / Dwelt from eternity, dwelt then in thee” (3.3–5), a description that evokes the same sense of depth as did that of Death. In this sense, Milton uses his narrator to establish an underlying defiance of form from the very beginnings of Paradise Lost.

Milton continues to develop the form-defying, gender-fluid characteristics conveyed within his descriptions of his muse (1.21–22) and the fallen angels (1.423–28) in the dialogue he invents between Adam and the archangel Raphael. In their lengthy conversation, Raphael provides insights into the formless, fluid nature of Milton’s pre-lapsarian universe. After unabashedly gushing about the logistics of heavenly sex—“and obstacle find none / Of membrane, joint, or limb, exclusive bars” (8.622–26)—the archangel directly touches on Milton’s monist philosophy in response to one of Adam’s questions concerning the basis for all things’ existence: “one first matter all” (5.432). He explains to Adam that everything in the entire universe is composed of the same matter and that everything exists on a
“scale of nature” (5.509) up which humans can ascend toward angelic heights by following God’s commandments: “Time may come when [hu]man[s] / With angels may participate, and find / No inconvenient diet, nor too light fare: / And from these corporal nutriments perhaps / [Their] bodies may at last turn all to spirit” (5.493–97). Milton thus fundamentally undermines the notion of concrete form altogether in *Paradise Lost*, as pre-lapsarian humans could attain angel status through obedience.

The notion of concrete form is later critiqued by none other than Satan himself, upon his transformation into a serpent. Satan bemoans the slimy form he chose to inhabit, having once been of the highest angelic form yet now “constrained / Into a beast” (4.164–65). In this manner, Milton reframes earthly form as something of a cage rather than a vessel. Toward the end of his epic, Milton also reiterates Satan’s frustrations with the confines that his serpentine form imposed upon him in establishing that God’s punishment for Satan and the fallen angels’ misdeeds would be permanent imprisonment in snake form:

> [H]e hears / On all sides, from innumerable tongues / A dismal universal hiss, the sound / Of public scorn; he wondered, but not long / Had leisure, wond’ring at himself now more; / His visage drawn he felt to sharp and spare, / His arms clung to his ribs, his legs entwining / Each other, till supplanted down he fell / A monstrous serpent on his belly prone[.] (5.504–14)

In short, *Paradise Lost* appears ripe with elements both formless, form-defying, and form-critiquing. Many of its primary characters—Satan, the Holy Trinity, Death, all the angels, and all the fallen angels—either transcend or explicitly condemn form. The ontological philosophy within which *Paradise Lost* operates undermines the concrete nature of form. The very form in which Milton wrote it was seemingly chosen in repudiation of literary form’s confines and the restrictions that a dramatic or poetic form would otherwise impose. All in all, *Paradise Lost* seems to mutiny against form in both narrative and construction.

Although Milton’s sonnets and epic poem can be said to most prominently convey his personal rebellion against the expectations of form, his other works express similar attractions to and resistance against their forms’ confines. *Areopagitica*, Milton’s 1644 argument for unlicensed printing in England, could be interpreted as a polemic argument rallying against the unjust restrictions inherent to the regulatory confines—in other words, the form—with which Parliament decided that all writing should be printed and distributed. He argues that restricting what the population can and cannot read not only hinders those who have written
pieces that Parliament deems unfit for print but also stifles the intellectual growth and moral fortitude of those who would otherwise have read such purportedly “evil” works (Areopagitica 939). Playing off of the same heretical concept underlying the Book of Genesis’s tale of the Fall of Humanity, upon which Paradise Lost would eventually be based, Milton puts forth the notion that people cannot determine what is good without first discovering what is evil: “It was from out the rind of one apple tasted that the knowledge of good and evil as two twins cleaving together leapt forward into the world” (Areopagitica 939). Thus, the man who would hypocritically become a print licensor himself under Cromwell’s reign asserts that licensed printing removes all opportunity for readers to learn for themselves the difference between good and evil works and seems to be a flawed regulatory form at its core.

Milton’s “Lycidas” and “On the Morning of Christ’s Nativity,” rather than offering explicit critiques on one form or another, both defy the conventions of their respective forms. “Lycidas,” “one of the most famous, most powerful, and most studied poems in English literature,” is an elegy written in the wake of the untimely death of Edward King, a contemporary of Milton’s at Christ’s College of Cambridge (Kerrigan et al. 99). Milton elegizes King with a “pastoral elegy,” the first of its kind: he uses the character Lycidas, a relatively common name in classical pastoral (Kerrigan et al. 100) in place of King, depicting how shepherds, the sea god Neptune, and even Nature herself mourn Lycidas’s premature death. Using gentle, pastoral imagery to elegize the dead is a Miltonian invention, and such departure from convention did not come without criticism. Samuel Johnson argued in his review of “Lycidas” that “[w]here there is leisure for fiction, there is little grief.” Regardless of how Johnson or others might have received it, however, Milton certainly broke from convention in writing Edward King’s elegy and challenged the expectations of elegiac form in the process. Similarly, “On the Morning of Christ’s Nativity” depicts a refreshingly original take on the traditional Christian nativity poem. Kerrigan, Rumrich, and Fallon write that other Nativity odes dwell on the traditional Christmas-card aspects of the Christ child’s birth, whereas in lieu of such conventional themes, Milton depicts “a sense of the meaning of the Nativity within the full scope of Christian history” as well as within “the various cultures, religions, and artistic traditions found in the world at its advent” (18). Much like his elegy of Edward King, Milton’s Nativity poem therefore departs from the form that convention would expect such a piece to take.

With the understanding that Milton was both drawn to and compelled to resist constraints, I close by considering the wider implications of such a claim. The
connection between Milton’s definitions and exploration of form and matter in his *Artis Logicae* and his chosen blank-verse epic medium for *Paradise Lost* implies metaliterary strategy in Milton’s other works. In other words, Milton seemed to choose mediums that would inherently restrict his forms and then, within those restrictions, to push back. His sonnets also support such a concept; he was the first to appropriate the sonnet as a vehicle of political discourse. These plausible realities bring up the question: Why? Why consciously choose highly constrained forms, such as the sonnet, the epic, and even the iconic first male and female, Adam and Eve, only to then resist them? Contemporary neuroscience and social science research suggest the reason. In 2019, Oguz, Tarakci, and van Knippenberg reviewed almost 150 empirical studies and found not only that people thrive amid constraints but also that constraints make people more, not less, creative. Somehow, Milton already knew that.

Milton’s preoccupation with both form and form-play, if you will, bears out contemporary research on limitations. Milton’s *Paradise Lost*, *Areopagitica*, “Lycidas,” “On the Morning of Christ’s Nativity,” and English sonnets all suggest that the famed poet, polemic, civil servant, and intellectual found paradise in first consciously choosing and then defying form’s confining bars. John Milton’s version of paradise, it seems, was a paradise uncaged.
Works Cited


MUSICAL HYBRIDIZATION AND POLITICAL CONTRADICTION: THE SUCCESS OF ARTHUR HONEgger’S ANTIGONE IN VICHY FRANCE

EMMA K. SCHUBART, UNIVERSITY OF NORTH CAROLINA, CHAPEL HILL
MENTOR: SHARON JAMES

Abstract

Arthur Honegger’s modernist opera *Antigone* appeared at the Paris Opéra in 1943, sixteen years after its unremarkable premiere in Brussels. The sudden Parisian success of the opera was extraordinary; the work was enthusiastically received by the French public, the Vichy collaborationist authorities, and the occupying Nazi officials. The improbable wartime triumph of *Antigone* can be explained by a unique confluence of compositional, political, and cultural realities. Honegger’s compositional hybridization of French and German musical traditions, as well as his opportunistic commercial motivations as a Swiss composer working in German-occupied France, certainly aided the success of the opera. Additionally, the contradictions that characterized the fascist Vichy state and the rise of modernism in France also abetted the opera’s success.

Arthur Honegger’s opera *Antigone* premiered in 1927 at the Théâtre Royal de la Monnaie in Brussels. This musically innovative opera, set to a libretto by French writer Jean Cocteau, featured sets by Pablo Picasso and costumes by Coco Chanel. Initial reviews of the opera were generally positive, but the work was not enthusiastically received until sixteen years later at its French premiere at the Paris Opéra in 1943 (even though it had originally been rejected by the Paris Opéra for being too modern).\(^1\) Once on the Parisian stage, the opera was a sensation; when especially moved by the opera, audiences would interrupt with bursts of applause. In addition to receiving acclaim from the French public, Honegger’s *Antigone* was surprisingly approved by both Vichy officials and the occupying German authorities. How is it that this opera, which is now largely forgotten, could be enthusiastically endorsed by the Parisian public, the collaborationist French government, and Nazi occupiers, particularly given that the opera fearlessly addresses the conflict of morality and legality (as did the original Greek tragedy

\(^1\) For the French premiere, Jean Cocteau designed the sets and costumes.
penned by Sophocles)? The answer to this question has extraordinary implications vis-à-vis an understanding of culture within Vichy France. This paper will analyze how the political climate in Paris in 1943 and Honegger’s own willingness to navigate that climate enabled the success of a modernist opera that was based on a politically charged Greek tragedy.

Sophocles produced his tragedy *Antigone* sometime in or around 441 BCE. The play begins right after the deaths of brothers Eteocles and Polynices, who died on opposite sides of the civil war fighting for the throne of Thebes. Creon, King of Thebes, decreed that Eteocles would receive an honored burial while Polynices’s body would be left to rot on the battlefield without being sanctified by holy rites. Such a refusal to allow burial was the mark of extreme disgrace and dishonor. The play begins with Antigone deciding to disobey her king (who is also her uncle) by burying and sanctifying Polynices’s body. After she accomplishes this, she is summoned before Creon and is questioned. She fearlessly admits to administering the funeral rituals to Polynices, defends the morality of her actions, and denounces the immorality of Creon’s law. Creon then condemns her to being buried alive in a cave. The play ends with Antigone’s suicide, the suicide of her betrothed, Haemon, and the suicide of Creon’s wife, Eurydice. This ancient Greek tragedy is simultaneously accessible and distant: it grapples with the ageless conflict between morality and legality, yet the text is obviously situated in the context of Ancient Greece. As Oudemans explains, “The text oscillates between the familiar and the alien because the *Antigone* is part of a cosmology … which differs radically from our own.” Nonetheless,

the cosmology of the *Antigone* is familiar to us, not merely because this tragedy touches emotional chords in the modern European mind, but primarily because it tries to cope with cosmological problems with which we are confronted as well, although its solutions and ours are mutually exclusive. In this sense, the *Antigone* is a thorn in the side of modern European cosmology; small wonder that a range of interpretative efforts have been made either to remove the thorn or to turn it into something beneficial. The “familiarity” of the ancient Greek tragedy was palpable when it appeared on the Parisian stage in 1943. Evidently, it struck a chord with its factious European audience. Instead of hindering such a musically unconventional opera from taking

---

3 Oudemans and Lardinois, 1.
hold, the complicated “cosmology”—the political and cultural realities of Vichy France and, more broadly, of Europe during World War II—actually seem to have facilitated its accessibility and, in turn, its success.

To understand the Parisian success of Antigone in 1943, one must first examine the political climate in which the opera premiered. The Vichy regime was the French puppet government that lasted from the German invasion of France in 1940 until 1944. On June 22, 1940, occupied France, led by new head of state Marshal Pétain, signed an armistice with Nazi Germany that effectively cut France in half. According to the Hitler–Pétain pact, in exchange for complete cooperation, Pétain would be free to control the “free zone” of the country and German forces would remain in the occupied north. Even though Pétain claimed to be the protector of the French public, his regime was a fascist one with its own ultraconservative agenda: divorce was prohibited, members of the French Resistance were arrested, phone calls were monitored, and the press was censored. During and immediately after German occupation, the Vichy regime depicted itself as the lesser of two evils. To use the words of Prime Minister Pierre Laval, it was thanks to the Vichy government that France managed to “éviter le pire.”4 In his statement before the High Court of Justice in 1945, Pétain defended his government, claiming his regime had stood between the French public and the enemy: “I used my power as a shield to protect the French people. Every day, a dagger at my throat, I struggled against the enemy’s demands. History will tell all that I spared you, though my adversaries think only of reproaching me for the inevitable.”5

Despite Pétain’s claims, Vichy certainly was no shield between the French public and its Nazi occupiers. As Paxton notes, “In fact, France became the largest single supplier to Germany of foreign male labor in all occupied Europe in 1943, east or west.”6 Indeed, the Vichy regime did not spell safety for the nearly 80,000 Jewish refugees who were shipped from France to death camps. It was not until decades after the end of World War II that the Vichy government’s role in the Holocaust fully came to light. As Lorraine Boissoneault explains,

The misconception that the Vichy Regime was the lesser of two evils endured only for the first few decades after the war. Since then, as more

4 Paxton, Vichy France, 358.
5 Paxton, 358.
6 Paxton, 366.
archival material has been made public, historians have gradually come to see the collaborators as willing participants in the Holocaust.\(^7\)

Hannah Arendt points to particularly “hair-raising” proof of the Vichy government’s part in the Holocaust. In the summer of 1942, 4,000 Jewish children who had been rounded up with their parents in Paris were left at the concentration camp at Drancy after their parents had been shipped to Auschwitz. After spending ten days at the concentration camp, these children, too, were shipped to Auschwitz. It was “[Prime Minister] Laval himself [who] had proposed that children under sixteen be included in the deportations; this meant that the whole gruesome episode was … the outcome of an agreement between France and Germany, negotiated at the highest level.”\(^8\) During the summer and fall of 1942, twenty-seven thousand stateless Jews—eighteen thousand from Paris and nine thousand from Vichy France—were deported to Auschwitz.”\(^9\)

Key to the Nazi implementation of a collaboration d'etat with Vichy France was the divided and disorganized political landscape in France. Stanley Hoffmann outlines the political “incoherence” that defined Vichy France and identifies those members of French society who condemned the republic as an oppressor that “had both corrupted the French polity … and weakened French society,” the socially discontented, anti-republican conservatives, and the fascists who sought either “French regeneration or revitalization” via totalitarian leadership or insurrection.\(^10\) This was the splintered political landscape in France in 1940 upon which Hitler capitalized. “What made Hitler's apparent ‘moderation’ ”—that is, the exploitation of France without implementing total Nazi occupation—“of June 1940 so Machiavellian was its very ability to maximize France's inner tensions.”\(^11\)

While the political fragmentation in France lent itself to exploitation by its Nazi occupiers, it also translated to a disintegration of any distinct French cultural identity. As Jane Fulcher explains, “What constituted French values during Vichy was far from clear. The contradictory nature of the government’s goals led to inconsistency and confusion in its cultural policies.”\(^12\) This cultural confusion was systematically exploited by German programs that were implemented to undermine French culture. For example, German propaganda sought to undercut the prestige

\(^7\) Boissoneault, “Puppet Government.”
\(^8\) Arendt, Eichmann in Jerusalem, 164.
\(^9\) Arendt, 164.
\(^10\) Hoffmann, “Collaborationism in France.”
\(^11\) Hoffmann.
\(^12\) Fulcher, “French Identity,” 651.
of French music by emphasizing the similarities between French and German music, insinuating that French music did not belong to France. Vichy officials actively participated in these propaganda efforts. There were concerts throughout France, sometimes with lectures, specifically organized to emphasize the superiority of German music. This cultural ambiguity in French society, complicated by massive political disorder, actually seems to have been the perfect backdrop for the composer Arthur Honegger, as he never defined himself or his music as distinctly French or German and instead managed to straddle the traditions of both cultures.

Arthur Honegger was born in France in 1892 to German-speaking Swiss parents. As World War I was brewing, he obtained Swiss citizenship, which he assumed would be his safest option (though he would live in France for most of his life). Honegger was therefore a man without a straightforward sense of national identity, living in Europe during a time in which the political rhetoric was centered upon nationalism. As Fulcher describes, “this was a period saturated with nationalist rhetoric concerning art and the natural ‘soil,’ which created ambiguities within Honegger that would long remain.”

Honegger’s personal ethnic ambiguities were mirrored in his compositional education. He studied at the Paris Conservatoire from 1911 to 1918 with Charles-Marie Widor and Vincent d’Indy. As Honegger himself put it, he arrived in Paris already “nourished on the [German] classics and romantics, enamored of Strauss and Reger,” but he quickly fell in love with the French tradition when he found that in Paris, “it was not that [German] school, but the [French] Debussyites in full bloom.”

Honegger’s affinity for both French and German musical traditions would fortuitously lend itself to a successful career in German-occupied France.

The German–French eclecticism that defined Honegger’s ethnicity and education came to characterize Honegger’s music in Antigone. This was also evidenced in the composer’s controversial political behavior throughout World War II. Honegger was a member of Les Six, a group of six notable French composers who were given this title in 1920 by French music critic Henri Collet in his article entitled “The Russian Five, the French Six, and M. Erik Satie.” The five other members of Les Six were Darius Milhaud, Francis Poulenc, Georges Auric, Louis Durey, and Germaine Tailleferre. Les Six spearheaded musical modernism in France. Musical modernism refers to the new culture of music that emerged after

13 Fulcher, 653.
14 Fulcher, Composer as Intellectual, 184.
15 Honegger, I Am a Composer, 91.
World War I that embraced atonality (without necessarily dismissing tonality) as well as innovative rhythmic, harmonic, and melodic concepts. Edward Campbell explains the modernist musical aesthetic as one that is based upon “the conviction that music is not a static phenomenon defined by timeless truths and classical principles, but rather something which is intrinsically historical and developmental,” and even though “the belief in musical progress or in the principle of innovation is not new or unique to modernism, such values are particularly important within modernist aesthetic stances.”

Les Six actively used their music as an instrument of revolt against social, political, and cultural norms. Throughout the 1920s and 1930s, Les Six became more and more politically engaged. This meant—with the notable exception of Honegger—celebrating traditional French values and condemning German fascism. Unlike the other members of Les Six, Honegger was not as concerned with the disruption of cultural and political institutions. He also differed from his contemporaries in that he was never anti-German. While the other five members of Les Six were heavily involved in the anti-German Resistance movement throughout the war, Honegger’s career was benefitting not only from the support of French civilians but also from German and Vichy endorsement. In 1941, he was invited to join the French Resistance group the Front National des Musiciens after he wrote a column defending French music for the arts journal Comoedia. As Harry Halbreich writes, “Honegger’s popularity had never been as great as during those years when [his compositions] Jeanne d’Arc au Bûcher, La Danse des Morts and the Symphony for Strings united audiences in packed concert halls in a single spirit of enthusiasm and hope.” Later that same year, however, Honegger was invited to Vienna to participate in the international Mozart festival, which was organized by the German Ministry of Propaganda under Joseph Goebbels. After this trip, Honegger was granted exit visas by German officials to give concerts of his piece Pacific 231, which Nazis actually considered “degenerate” music, throughout Europe. In 1942, he attended a German embassy reception for Heinz Drewes, who was the head of Goebbels’s music department in Berlin. That same year, French orchestras began reviving Pacific 231, and in June and July, German officials allowed French musicians to celebrate Honegger’s birthday with weeklong festivals. In fact, Honegger was the only living composer to receive such an honor in occupied France.

16 Campbell, Boulez, Music and Philosophy, 37.
17 Halbreich, Arthur Honegger, 164.
18 Sprout, “Unlocking the Mystery of Honegger.”
By 1943, in light of his connections to Nazi officials, Honegger was dismissed from the Resistance; however, as Leslie Sprout notes, “the timing of these events [his connections with Nazi authorities] suggests that Honegger was motivated not by political sympathy for the occupiers but by professional opportunities.”19 After the war, Honegger sought to defend his dealings with German authorities when he wrote, “Going to the enemy’s camp does not automatically mean that one supports his cause”20; nonetheless, records of Honegger’s wartime compositions certainly do not exonerate him. His *Chant de liberation* was performed in 1944 in one of the first concerts in Paris after the Liberation. At the time, French author Maurice Brilliant wrote of the piece, “For our joy and honour, Honegger was a composer of the Resistance.”21 The score was lost until 2010. The melody of the piece comes from the music for the film *Joan of Arc* that Honegger composed in 1942. The film was a piece of Vichy propaganda meant to depict Pétain as a virtuous protector of the French people against the enemy of Great Britain. Honegger also composed an oratorio on *Jean d’Arc au Bucher* (Joan of Arc at the Stake), which the Vichy regime toured throughout southern France. Honegger later claimed that he used the melody from *Joan of Arc* in a piece that he composed as a Resistance song as early as April 1942, however. This composition celebrated the Allied powers and incorporated quotations from “It’s a Long Way to Tipperary,” “The Star-Spangled Banner,” and “La Marseillaise.” If the date of this composition had been correct, it would prove that Honegger’s allegiance lay with the Resistance even while he was outwardly in contact with Nazi authorities—but when *Chant de liberation* was uncovered in 2010, it bore a date of April 1944, two years later than Honegger had claimed. “Instead of placing Honegger as a Resistance composer,” Daisy Fancourt explains, the 1944 composition date “suggests that he feared he would be accused of collaborating, so [he] wrote a Resistance song and lied about its date to provide evidence that might save him during a trial.”22

Tellingly, the next year, Honegger produced a song entitled “Hymne de la délivrance,” composed in the style of Resistance songs, which he used in the score of a film about Resistance fighters, *Un Ami Viendra ce Soir*, later that year. Honegger’s flurry of anti-German Resistance compositions, all produced when Allied victory was becoming likely, points to the opportunism of the composer. Throughout World War II, Honegger’s behavior with the Resistance, the Vichy

---

19 Sprout.
20 Quoted in Sprout.
21 Quoted in Fancourt, “Les Six.”
22 Fancourt.
collaborationist government, and Nazi authorities was indicative of a composer whose motivations were not ideological as much as commercial. His ability to toe the line between French and German politics was mirrored in his music, and this ultimately facilitated approval by a wide audience.

Honegger revered the music of Beethoven, Wagner, and Bach but was also strongly influenced by Debussy, Ravel, and d’Indy. He not only managed to exist in both musical contexts but was also able to highlight the musical idioms of both worlds in his music. He conveyed the modern French style within the framework of the German tradition, becoming something of a musical hybrid of a composer. Honegger embodies what Jay Winter describes as the emotional response of the post–World War I generation to move forward with “a refashioned set of ideas and images derived from a range of older traditions”; instead of completely rejecting tradition, this generation sought to “recast” it, as a way of “walk[ing] backwards into the future, struggling to understand the chaotic history of this century.”

Honegger’s music certainly did walk backwards into the future; his compositions embraced tradition while striding boldly into the new post–World War I era of music modernism.

This duality of traditional and modern may be best seen in Honegger’s opera Antigone. French composer Olivier Messiaen often heard Honegger announce, “If something should last after my death, it ought to be Antigone.” Although the opera is largely unknown today, it has nonetheless been singled out by prominent composers as a masterful, pivotal work. Antigone was considered by French composer and conductor Pierre Boulez to be a pioneering opera, charting new territory for the next generation of composers. The innovation within Honegger’s Antigone lies in his realization of Cocteau’s libretto. Through his libretto, Cocteau sought to revitalize Antigone in a post-World War I context. After seeing a performance of the tragedy by Sophocles at the Comédie Française, which Cocteau described as “incredibly boring,” he decided to update the play with a condensed three-act libretto. In 1922, Cocteau explained that he meant to “put a new dress on the old Greek tragedy, adapting it to the rhythm of our own language.” He accomplished this by using French colloquialisms (as did Sophocles, though in Greek) and by making the libretto compressed and synoptic. Cocteau compared his libretto to a “pen drawing after a painting by an old master” and “an aerial

\[\text{\footnotesize\textsuperscript{23}}\] Winter, Sites of Memory, 221–222.
\[\text{\footnotesize\textsuperscript{24}}\] Quoted in Shapiro, Les Six, 182.
\[\text{\footnotesize\textsuperscript{25}}\] Quoted in Fulcher, “French Identity,” 658.
\[\text{\footnotesize\textsuperscript{26}}\] Quoted in Fulcher, 658.
photograph of the Acropolis.” As his analogies suggest, Cocteau’s libretto was rough and ungarnished. This presented an exciting challenge to Honegger, who was inspired by Cocteau’s bare-bones approach.

Honegger himself explained that, to complement Cocteau’s innovative writing, he would have to be inventive as he set the libretto to music: “[As] … I was composing the music for Antigone, on a violent and even brutal text, I sometimes said to myself: ‘If I combine words and music for this text in the customary fashion, it will lose its character, its power.’ ” Besides the challenge posed by the austerity of the libretto, Honegger was also attracted to its fast pace.

He believed that slowness was the great failing of lyric theater and felt that twentieth-century audiences were conditioned by technological advancements in the media for rapidly unfolding plots. As he explained in his 1929 address to the Rice Institute:

The principal reasons for the lack of success of the great part of the work by the lyric theatre seem to me to be the following: the slowness of the action, exaggerated by the chorus and the symphony and the impossibility of understanding the text. In this day we are accustomed to speed. The motion picture has given us the taste for the swift succession of tableaux and the public no longer has any patience.

As such, Honegger composed Antigone with the goal of complementing the condensed libretto with music that was just as swift and succinct.

Specifically in the case of Antigone, as he composed the vocal lines, Honegger sought to avoid the patterns of stress and intonation that are inherent in the French language. To achieve maximum clarity, he often displaced the natural syllables and accents of the language. He was guided not by the French language but by the theatrical context of the words he was setting to music. In so doing, he managed to strike a balance between the Italian bel canto tradition and the Debussian tradition of recitatives. Honegger disliked bel canto—a style of singing that prioritizes lyricism above the realities of prosody—because the singer can hold the vowel for as long as he or she wishes, which obviously interferes with the audience’s comprehension of the text. As Honegger put it, it is thanks to bel canto

---

27 Quoted in Fulcher, 658.
28 Quoted in Spratt, Music of Arthur Honegger, 131.
29 Quoted in Shapiro, Les Six, 183.
30 The tradition of continuous recitatives is attributed to French composer Claude Debussy because he employed this technique throughout his opera Pelléas et Mélisande.
that “in the lyric theatre one can never understand the singers. Now, ninety-nine times out of a hundred, it is not the singers’ fault, but the fault of the composers.”31 Clearly, Honegger meant to set himself apart from such misguided composers. He also rejected the operatic tradition of recitatives—a compositional technique of employing continuous holds in the music as the singer “sings” in the rhythm of ordinary speech on the same note—which Honegger described as “monotonous repetition” and “impassive syllabism.”32 Instead, he created a compositional hybrid that combined bel canto lyricism and the realistic prosody of recitatives by emphasizing the consonants of the words. He believed consonants facilitate language, and with this in mind, he accentuated the consonants in the vocal lines in Antigone.

Years after he composed Antigone, Honegger explained his compositional process: “I sought for the right stress, especially on the attacking consonants. … What is important in the word is not the vowel, but the consonant: it really plays the role of a locomotive, dragging the whole word behind it … the consonants project the word into the hall.”33 To illustrate this consonant-focused compositional technique, Honegger used the following example from Antigone:

At one instance, Créon violently interrupts the choir, crying out: ‘Assez de sottises, vieillesse!’ The conventional prosody suggests the following stress: ‘Assez de sottises, vieillesse!’ Try to project this phrase in anger with this rhythm: the aggressive effect is immediately blunted. Respecting the dramatic situation and Créon’s anger, I made it: ‘Assez de sottises, vieillesse!’ leaning on the roots of the words.34

In keeping with Cocteau’s spirit of revitalization, Honegger breathed new energy into this Greek tragedy. He broke with operatic convention in order to allow the text to reach the audience with maximum effect. He musically shed what he felt was distracting and superfluous about the French language so a reenergized twentieth-century Antigone could pierce the hearts of her audience. The enamored audiences’ eruptions of applause throughout the opera proved that Honegger’s efforts to modernize the opera were successful.

In addition to the Parisian public, Vichy and German critics were equally enthusiastic about Antigone. Music critic Arthur Hoérée praised the opera in the

31 Quoted in Spratt, Music of Arthur Honegger, 131.
32 Quoted in Spratt, 131.
33 Quoted in Spratt, 131.
34 Quoted in Spratt, 131.
Vichy-approved journal *L’Information musicale*. Tellingly, in his glowing review of the opera, Hoérée claimed that Creon, not his defiant niece Antigone, deserves empathy. Similarly, Nazi-approved forces also commended *Antigone*. German composer Werner Egk, who was a favorite composer of Joseph Goebbels, praised Honegger’s eclectic compositional approach.

Enthusiastic reception of Honegger’s modernist opera sheds an interesting light on the goals of the Nazi state with regard to art censorship. Traditionally, historical analyses of art sanctioned by the Nazi regime indicate that the Nazi state was primarily interested in promoting the musical compositions of ethnically German composers—Bruckner, Bach, Strauss, Beethoven, Wagner, etc. It is critically important to remember, however, that the Nazi regime also sought to establish itself as unmistakably modern and therefore encouraged modernist art. As Michael Kater highlights, “On close examination we can find in the Third Reich elements we would not expect in the dictionary definition of a totalitarian regime,” such as “avant-garde attempts at modernism that may not have conformed with the modernist conceptions of the early republic but were nonetheless novel and were in fact officially welcomed and even subsidized as such.” This is reflected in a speech given by Joseph Goebbels in Munich in 1936: “The National Socialist Weltanschauung is the most modern thing in the world today, and the National Socialist state is the most modern state. There are thousands of motifs for a modern art in the spirit of this weltanschauung.”

Apparently, Honegger’s *Antigone*, an opera derived from the ancient Greeks and served up in the middle of World War II in France, fit perfectly with the Nazi objective of cultural modernism.

All art is subject to varying, even conflicting, interpretations. It is an extraordinary piece of art that can withstand the test of contradictory interpretations arising from a fraught political climate and a highly contentious cultural zeitgeist. Politically speaking, *Antigone* was able to endure through myriad interpretations thanks to its composer’s ethnic ambiguity, commercial opportunism, and compositional language of musical hybridization. In addition, it must be underscored that this opera satisfied the cultural needs of the occupied French public, the Vichy government, and the Nazi forces, which helped to ensure its relevancy. The opera’s music, however, seems to be the primary reason for

---

35 Beethoven proved an interesting challenge for Nazi officials who wanted to claim his music as their own but had to considerably revise his biography in order to do so. See Schubart’s “Tainted or Transcendent” for an explanation of the lengths Nazi officials went to in order to revise Beethoven’s questionable (i.e., not Aryan) heritage.


37 Quoted in Kater, 177.
Antigone’s wide cultural appeal. The innovative musical realization of the text, the atonality, and the innovative rhythmic figures combined to form a piece of modernist music that concomitantly satisfied the Nazi regime’s agenda of appearing avant-garde, pleased the collaborationist Vichy government, and gave the culturally sidelined French public a particularly timely story.

In its day, the opera was hugely popular and was predicted to go down in history as a masterpiece. This prediction was incorrect, as today, Honegger’s Antigone has been largely forgotten. Regardless of its relative obscurity, however, this opera charted new, modernist musical territory and paved the way for the next generation of composers to expand the limits of musical expression. Both musically and culturally, Honegger’s Antigone was made possible by an extraordinary confluence of political circumstances. An innovative French librettist collaborated with a cultural hybrid of a composer at precisely the moment in history when an occupied French nation, its collaborationist French government, and Nazi occupiers would all be supportive of the rebirth of a story that fearlessly and incongruously questions the rule of law. The enthusiastic Parisian reception of Honegger’s Antigone reveals the chaos of a culturally fragmented and politically hypocritical Vichy France. At the same time, the Parisian response to the opera reveals a common desire among Honegger’s audiences: the desire to see the traditional reclothed in the modern. The very admittance of this opera onto the Parisian stage reveals the political, social, and cultural conflicts that existed within the fascist Vichy regime and sheds light on the development of French modernism at a moment in history when French culture was quite literally under siege.
Bibliography


THE STRATIFICATION OF JUSTICE: EVALUATING THE RELATIONSHIP BETWEEN GENDER, RACE, AND CRIME

CATHERINE BAIN, BUTLER UNIVERSITY
MENTOR: KRISTA CLINE

Abstract

Today, 2.3 million Americans are incarcerated; 1,564,000 of those people are not White. This illustrates a clear racial disparity within the U.S. justice system in terms of which individuals are sentenced to prison. Although previous research, examining mostly state justice systems, has found that racial bias leads to longer sentences and has noted differences in sentencing length for male and female defendants, a paucity of research examines the impact of race and gender on sentencing lengths within the federal justice system. This study seeks to fill that gap. Data from the years 2000 through 2016 were obtained via the U.S. Sentencing Commission’s Monitoring of Federal Criminal Cases and were analyzed using multiple linear regression featuring control variables of education level, age, socioeconomic status, and criminal history. The data set included 1,011,988 defendants, of whom 87% were male, 71% were White, and 78% had a criminal history. Results found that, on average and across crimes, White individuals received shorter sentences than did people of color, and women received shorter sentences than did men. When comparing ethnicity, on average and across crimes, White men received shorter sentences than did men of color, and White women received longer sentences than did women of color. These findings suggest that a judge may base the length of sentence not only on the facts of the case but also on the defendant’s race or gender. These results alert federal judges to possible biases of which they must be aware, and they highlight the need for a policy to control for demographic factors in determining sentences.

Keywords: criminology, race, gender, sentencing length, justice, justice system

Race and, subsequently, racism seem to be written into the DNA of America. Whether the topic is mass incarceration, police brutality, or political policies such as stop-and-frisk, race is always a factor. This conversation does not
stop as we cross over into the field of academia. Scholars argue that racial bias was a main motivator behind the war on drugs as well as behind laws such as stop-and-frisk and the three-strikes policy (Alexander, 2010). In addition, we know that the label of “criminal” is more likely to be assigned to a person of color than to a White or White-passing individual, and links have been made between the label of criminality and rates of recidivism (Chiricos et al., 2007). The field of research focusing on race as it pertains to criminality and criminal justice is not new. In fact, this type of research has a long and complicated history, ever more complicated by the many types of relationships that can be examined in any given court case. This study seeks to examine these relationships and to build upon them. Specifically, this study aims at examining the relationships between race and sentencing length, gender and sentencing length, and the intersection of race and gender with sentencing length. To examine these relationships, however, we must first examine previous findings about these relationships.

Literature Review

Race

Race has a long and complicated history with the U.S. justice system. The 13th Amendment, often credited with ending slavery in America, did not fully do so. In fact, it declares the enslavement of those incarcerated to be legal, a fact taken advantage of quite frequently during Jim Crow (Alexander, 2010). When looking at the influence of race within our justice system today, we see the literature convening on three major focuses of study: the race of the defendant, the race of the victim as it pertains to the race of the defendant, and the race of the victim.

The race of the defendant has been thought to influence various things within the justice system, including the likelihood of conviction (Gibson, 1978; Miller & Hewitt, 1978) and even the likelihood of sentencing with the death penalty (Radelet, 1981). In terms of conviction rates, research has shown that juries are more likely to recommend longer sentences for cases in which the defendant was of a race dissimilar to the racial makeup of the jury (Miller & Hewitt, 1978) and has shown geographical patterns of racial bias (Gibson, 1978). For example, Gibson (1978) found three major patterns of conviction throughout the United States—pro-Black, anti-Black, and nondiscriminatory—and that these patterns seem to correlate with geographical location: for example, anti-Black patterns of conviction tended to occur in the more southern area of the United States.
In many cases, such as the geographical connections seen in Gibson’s (1978) work, culture can play a part in influencing our opinions of others. For example, we typically are drawn to those individuals who look like us and avoid those who do not. This fear of “the outsider,” commonly referred to as xenophobia, is believed to be an evolutionary trait (Cook et al., 2018), for we think that those who look like us will not harm us but those who look different may cause us harm. This pattern of thinking led to the second major relationship of focus in this study: race of the victim as it pertains to the race of the defendant. Previous research has shown that cases with defendants of color and White victims are the most likely to end with the defendant being sentenced to jail (Radelet, 1981) but that cases with White defendants and victims of color did not see the same likelihood. This indicates that not only the dissimilarity of race of victim versus race of defendant but also which individual is of which race is important to sentencing outcome (Radelet, 1981).

These findings therefore lead to the last relationship studied: the race of the victim. Would the race of the victim still matter when the race of the defendant was controlled for? Pierce et al. (2017) found evidence that cases with White victims were more likely to result in death sentences than cases with victims of color even after controlling for the racial makeup of the jury and the race of the defendant.

**Gender**

When it comes to the role that gender plays within our justice system, we know four main things. First, research shows that men and women tend to commit different types of crimes (Butcher et al., 2017; Lauritsen et al., 2009). Specifically, men are more likely to commit violent crimes than are women (Butcher et al., 2017), although that difference in likelihood seems to be decreasing (Lauritsen et al., 2009).

Second, men, in general, are more likely to be sent to prison than are women (Butcher et al., 2017). This difference is seen even if the type of crime committed is controlled for, meaning that a man who committed first-degree murder would be more likely to go to prison than a woman who committed first-degree murder (Butcher et al., 2017). We also know that men are more likely than women to commit crimes in general (Steffensmeier & Allan, 1996).

Third, we tend to associate men and women with violence differently. To better understand this difference, we must first understand the social phenomena that are gender roles. Dainty, kind, shy, generous, sensitive, and quiet are all
adjectives typically associated with women; that is how we expect women to be. *Strong, aggressive, loud, unbothered, greedy, insensitive*; these are the words we use to describe men. We have a belief that, to be considered “manly,” one must fit this mold and that to be a valuable member of society as a man, one must be manly. This is taught over and over through sports, video games, movies, and other forms of media. We teach men to be aggressive, and we all too often equate aggression with violence (Baugher & Gazmararian, 2015; Connell, 1996; Cornwall, 1997; Santana et al., 2006). It should therefore be no surprise that we associate men and women differently with crime.

Finally, we see that the gender of the victim is relevant. Pierce et al. (2017) found that cases with female victims were far more likely to result in death sentences than were cases in which the victims were male. Additionally, in the past, it was seen that females were more likely to have violent acts committed against them, although more recently, no statistically significant difference across gender has been found in terms of who is being victimized (Morgan & Truman, 2017).

**Other Extralegal Factors**

An extralegal factor is a factor pertaining to the case that is outside the scope of the law, such as race, gender, socioeconomic status, and the like. Research has shown that those who are of a lower socioeconomic status receive inadequate defense as a result of relying on federally funded defense attorneys who are often overworked and are therefore unable to designate adequate time to a defendant’s case (Gould & Leon, 2017).

Another extralegal factor that may contribute to sentencing is that of stereotypes. Welch (2007) argues that there has been continual discrimination of people of color throughout America’s history via our use of stereotyping. The study links this idea with the caricatures of people of color portrayed during the blackface era as well as with current policies such as stop-and-frisk that allow for racial profiling. Welch argues that by imposing these policies, we are reinforcing the false idea that people of color are somehow criminal in their biology. To support this theory, the study cites the disproportionate prison population and how it skews public perceptions of crime to be unrealistically correlated with race, and the impact this perception has on the way law enforcement officers and public officials do their jobs. In addition, Welch’s research reaffirms the theory that a problematic relationship exists between false public perception and the way that perception influences policy making, as was previously found in the work of Baldus (2004).
One other important piece of background information to know is how crimes are processed in the United States. When an individual commits a crime in this country, the crime can be processed in either one of two systems: the state justice system or the federal justice system. Each of these systems processes different types of crimes. The state system deals with things such as murder, arson, robbery, rape, theft, and burglary. The federal system deals with fewer classes of crimes because the crimes must involve a national or federal interest. This is not to say that the federal system does not deal with things such as murder or rape; in some cases, they do. Provisions for such cases come from the federal criminal code (18 U.S.C., 1970). This covers both violent and nonviolent crimes. In addition to these crimes, any crime committed on federal property, such as national parks or federal courthouses, is deemed a federal crime.

The current, quantitative, study fills two gaps in the literature by exploring the impact of race and gender on the specific length of sentencing and by looking specifically in our federal system. In addition, given previous knowledge on the impact of socioeconomic status (Brown & Males, 2011) and age (Farrington, 1986) on criminality, this study controls for these variables. Given the previously established racial bias seen in our justice system in terms of what defendants are sentenced to prison, as well as our knowledge about the relationship between gender and criminality, I expected three things: (1) longer sentences are given to individuals of color than to White individuals, (2) women are given shorter sentences than men, and (3) with regard to the interaction between race and gender, men of color are given the longest sentences and White women are given the shortest sentences.

Given the importance of our justice system in America, a study looking at possible further bias has the potential to enhance our understanding of the complex relationship that America has with both race and gender. In addition, this research has the potential to influence future policy implications working against this type of bias. Finally, by demonstrating who is most vulnerable to bias in our justice system, this study could help to identify those who are most likely to benefit from policies working against it.

Methodology

Research Design

This study involved a quantitative research design that utilized a secondary data set. This design was chosen to eliminate bias caused by obtaining the data and
because of restrictions in feasibility in collecting original data. Overall, doing a secondary data analysis provided the most reliable use of measurement. The universe of data contains all cases received by the United States Sentencing Commission (USSC) with sentencing dates between October 1, 2000, and September 30, 2016.

Participants

Participants included 1,011,988 individuals (29% White) aged 16–97 (M = 35.42, SD = 10.94) from a variety of educational backgrounds. The sample included a greater proportion of men (87%) than women. Additionally, more cases with a defendant with a criminal history (78%) were obtained than for a defendant without a criminal history. To ensure that age and criminal history did not influence the results of the primary analyses, it was included as a covariate. In addition, highest level of education was used as a covariate to control for socioeconomic status because the two variables are known to be positively correlated.

Measures

Data were obtained from the USSC’s Monitoring of Federal Criminal Cases data sets for the years 2000 through 2016. The USSC obtained these data through a repeated cross-sectional design. This data set has a universe containing all cases received by the USSC that had sentencing dates between October 1, 2000, and September 30, 2016. Cases were reported to the USSC only if they were sentenced according to the Sentencing Reform Act (SRA) of 1984. All sentences recorded in this data set were deemed to be constitutional. The SRA of 1984 simply put forth a range of sentencing lengths that are allowed based on a given crime. It is important to note that all the data used in this study were from after the act was put in place. The original study obtained this data from judgments of conviction, guideline worksheets, statements of reason, the Federal Probation Sentence and Supervision Information System, plea agreements, and presentence reports.

Dependent Variable

The dependent variable in this study is sentencing length. Sentencing length was operationalized as the number of months the defendant was sentenced to at the time of original sentencing.
Independent Variables

The independent variables of interest are race and gender. This study operationalized gender as male and female, for this is how gender is defined in the federal justice system. Race was operationalized in terms of White individuals and people of color, resulting from previous literature’s findings that policies determined to be discriminatory toward one minority group, such as stop-and-frisk, are discriminatory toward most, if not all, other minority groups (Gelman et al., 2012).

Control Variables

The current study included three control variables: age, criminal history, and educational status. Age was defined by the number of years the defendant had been alive at the time of sentencing. Criminal history was coded 0 for no criminal history and 1 for criminal history. Educational status included four categories originally coded as 1 for did not finish high school, 2 for high school diploma/GED, 3 for pursued higher education, and 4 for earned a trade school degree. Because educational status was the only control variable coded categorically, educational status was recoded into dummy variables upon analysis, making the value for each either 0 or 1.

Data Analysis

Data from a total of 1,011,988 defendants sentenced between September 1, 2000, and September 30, 2016, were analyzed using SPSS. Descriptive statistics were run on both the dependent variables and independent variables as well as on the control variables. Data analysis was then conducted using multiple linear regression, looking for a significant $p$ value of less than .05. For the data set containing all 1,011,988 defendants, four models of regression were run.

The first model included the control variables of age, education level, and criminal history as well as an independent variable of the year the sentencing took place, on the dependent variable, sentence length. The second model included these variables as well as the independent variable of race on the dependent variable. The third model included the control variables and the independent variables of gender and year of sentencing on the dependent variable. The fourth model included the control variables and the independent variables of year of sentencing, as well as the intersection of gender and race on the dependent variable. The models were run in this order to see the base effects of the control variables when dealing only with the
year of sentencing first, and then looking at the influence that the independent variable of interest had on the prior results. After the overall analysis was complete, an additional two models of analysis were run examining the variables of race while controlling for gender. In the original analysis, it was seen that gender had a larger effect on length of sentencing than did race, so an additional model was needed to compare women of color to White women and men of color to White men most accurately.

Models were run on individual years. No significant differences were seen between years, however, so in terms of analysis, the overall data set was the major point of focus.

Results

Descriptive statistics are provided in Tables 1–4. Four models were run on the overall data set. The first model contained the control variables and the year of sentencing, and the remaining three models contained the independent variables of race and gender, and the interaction of race and gender, respectively.

Relationships Among Key Variables

Standard correlation analysis was run in IBM SPSS Statistics to explore the relationships between all variables of interest within the study.

Unsurprisingly, defendants with a criminal history were given longer sentences than those without \( (r = .129, p < .01) \). In accordance with prior literature, younger defendants were given longer sentences \( (r = -.014, p < .01) \). Finally, educational status had a negative relationship with length of sentence if the defendant had either not finished high school, \( (r = -.036, p < .01) \), pursued a level of education past high school, \( (r = -.041, p < .01) \), or obtained some other form of degree, such as one from a trade school \( (r = -.010, p < .01) \), yet a positive relationship was found between graduation from high school and length of sentence \( (r = .079, p < .01) \). Consistent with my first hypothesis, White defendants were given lesser sentences than were defendants of color \( (r = -.016, p < .01) \). Consistent with my second hypothesis, female defendants received lesser sentences than did male defendants, \( (r = -.130, p < .01) \). Inconsistent with my third hypothesis, a negative relationship was seen between race and gender intersection and length of sentence \( (r = -.071, p < .01) \).
Mediating Effects of Race, Gender, and Interaction

To test the strength of influence of each independent variable on sentence length, I ran a series of regression analyses to test the potential mediating effects of race, gender, and race-gender interaction (Tables 3 and 4). For the variable of education level, pursuit of higher education served as the comparison variable in the analysis.

All control variables were found to be significant \( (p < .01) \) in all four models except Model 2, in which age and the educational level of other degree were not found to be significant at this level \( (other\ degree\ was\ significant\ at\ the\ .05\ level,\ p = .013) \), as shown in Tables 3 and 4. Year was also found to be a significant predictor of length of sentence. The first model found a significance with a small effect size \( (B = -0.464, SE = 0.017, p < .001) \), as did the second, third, and fourth models \( (Model\ 2: B = -0.494, SE = 0.017, p < .001; Model\ 3: B = -0.447, SE = 0.016, p < .001; Model\ 4: B = -0.474, SE = 0.017, p < .001) \).

The independent variable of race was calculated only through Model 2, where it was shown to be significant \( (B = -3.868, SE = 0.185, p < .001) \). The independent variable of gender was calculated only in Model 3, where it was shown to be significant \( (B = -26.876, SE = 0.231, p < .001) \). The independent variable of the interaction between race and gender was calculated only in Model 4, where it was shown to be significant \( (B = -23.762, SE = 0.356, p < .001) \). Although all models remained significant, the effect sizes in Models 1 \( (R^2 = .023) \), 2 \( (R^2 = .023) \), 3 \( (R^2 = .036) \), and 4 \( (R^2 = .027) \) were small.

After the overall analysis was complete, an additional two models of analysis were run to examine the variables of race while controlling for gender. The first additional model examined only females and showed race to be significant \( (B = 2.879, SE = 0.387, p < .001) \). The second additional model examined only males and also showed race to be significant \( (B = -3.920, SE = 0.206, p < .001) \). The effect size was small in both of the additional models \( (R^2 = .009 \ for\ the\ first,\ R^2 = .023 \ for\ the\ second) \).

Discussion

This study strongly supports the previously documented reports of both racial and gender bias in the justice system in the United States, but it also augments prior research through an exploration of the impact of race and gender on the specific length of sentence. The negative relationship between a defendant’s race
and length of sentencing seen in this study is consistent with prior findings that suggest a racial bias in the justice system (Gibson, 1978). More importantly, the current study expands on prior literature that has documented instances of bias within the justice system by illustrating relationships between race, gender, and sentencing length. Consistent with my first hypothesis, defendants of color were seen to be given longer sentences than were White defendants. Additionally, in support of my second hypothesis, it was found that female defendants were given shorter sentences than male defendants. When looking directly at the interaction of gender and race, a negative relationship was seen, although this may be due to the higher strength of the effect of gender as compared to race on sentencing length.

Despite the established relationships that race and gender have with sentencing length, results of this study do not support all of the anticipated mediating effects among these three variables. I hypothesized that racial influence would be the same across genders. Specifically, I anticipated that both women and men of color would be given longer sentences than were White men and women, but this was not supported. In examining race with the control of gender, analysis showed that White women were given longer sentences than were women of color, and White men were given shorter sentences than were men of color. Although race did not have the same effect on both genders, it did significantly affect sentencing length in both cases. This suggests that an outside variable may affect this given sentencing length. Previous research indicates that a difference exists in types of crimes committed by men as compared to women (Lauritsen et al., 2009), and this difference could be contributing to the difference in effect of race across genders. Another reason this difference is seen may be the sample of cases used in this study. The sample contained a much higher percentage of men (87%) than women, which may have produced skewed results.

Interestingly, age did not always prove to be significant. This finding is contrary to the results of Farrington (1986), which illustrate a strong relationship between age and likelihood to commit crimes. This finding is also surprising in its irregularity, for significance not only varied across years but also across models within a single year. The irregularity and the opposing results suggest that although age may be a significant predictor of likelihood to commit crimes, it is not as reliable in its ability to predict length of sentence.

Another result to note is that of the significance of year. This is interesting because it illustrates that the influence of racial and gender bias when looking at sentencing length is not static across years. Overall, looking at the cumulative data from 2000 through 2016, year was seen to have a negative relationship with
sentencing length, which indicates that we may be seeing a decrease in the amount of influence that race and gender have on sentencing length in recent years as compared to the early 2000s.

Despite this study’s promising findings about race, gender, and sentencing length, its results should be interpreted within the context of its limitations. First, the cases used in this study were reported to the USSC, and although the cases span the United States, they may be skewed to one geographic location over another. Most defendants are individuals of color with some criminal history. Given these demographic characteristics, these defendants may not be representative of the average defendant in the federal system. For example, defendants with criminal histories are likely given longer sentences than those without (Bushway & Piehl, 2007); thus, the majority of sentences within this data set may be longer than they would be on average in the federal system as a whole.

Additionally, the data set chosen to complete this survey did not contain a variable pertaining to type of crime committed. According to U.S. federal law, crimes are given different mandatory minimums and, as a result, different average sentences. As previously mentioned, men and women tend to commit different crimes; the inability to control for type of crime committed may therefore have influenced the results.

Despite these limitations, two of my three original hypotheses were supported: longer sentences for individuals of color as compared to those for White individuals, and longer sentences for male defendants than for female defendants. Although my third hypothesis was not entirely supported, study findings did support the hypothesis that men of color are given longer sentences than are White men. This conflicting result indicates that another variable may be at play, influencing sentencing length outside of the independent variables tested. Furthermore, the current study significantly contributes to the existing body of research by being one of the first to look directly at the influence of the interaction of race and gender on sentencing length, as well as by being one of the first to look specifically at the federal justice system. As such, this study provides valuable information that begins to elucidate the complex interplay among race, gender, and length of sentence, not only providing a foundation for new avenues of research but also identifying new information that politicians can utilize to promote policy change and reduce bias within the federal court system.
References


U.S. Const. amend. XIII.


Table 1. Means and Standard Deviations for Study Variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>White (1 = White, 0 = person of color)</td>
<td>0.29 (.451)</td>
</tr>
<tr>
<td>Female (1 = female, 0 = male)</td>
<td>0.13 (.341)</td>
</tr>
<tr>
<td>Race and gender interaction</td>
<td>0.0510 (.22008)</td>
</tr>
<tr>
<td>Age</td>
<td>35.42 (10.941)</td>
</tr>
<tr>
<td>Length of sentence (months)</td>
<td>50.6850 (79.14561)</td>
</tr>
<tr>
<td>Criminal history (1 = yes, 0 = no)</td>
<td>0.78 (.411)</td>
</tr>
<tr>
<td>Did not graduate high school</td>
<td>0.4806 (.49962)</td>
</tr>
<tr>
<td>Graduated high school</td>
<td>0.2939 (.45553)</td>
</tr>
<tr>
<td>Pursued higher education</td>
<td>0.1973 (.39796)</td>
</tr>
<tr>
<td>Attained other degree</td>
<td>0.0282 (.16560)</td>
</tr>
</tbody>
</table>
Table 2. Correlations of Race, Gender, and Control Variables

<table>
<thead>
<tr>
<th></th>
<th>White</th>
<th>Female</th>
<th>Race and gender interaction</th>
<th>Age</th>
<th>Length of sentence</th>
<th>Criminal history</th>
<th>Did not complete high school</th>
<th>Graduated high school</th>
<th>Pursued higher education</th>
<th>Attained other degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>.082**</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race and gender interaction</td>
<td>.367**</td>
<td>.588**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>.228**</td>
<td>.025**</td>
<td>.055**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length of sentence</td>
<td>-.016**</td>
<td>-.130**</td>
<td>-.071**</td>
<td>-.014**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Criminal history</td>
<td>-.073**</td>
<td>-.163**</td>
<td>-.084**</td>
<td>-.060**</td>
<td>.129**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Education Level**

<table>
<thead>
<tr>
<th></th>
<th>Did not complete high school</th>
<th>Graduated high school</th>
<th>Pursued higher education</th>
<th>Attained other degree</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-.302**</td>
<td>-.109**</td>
<td>-.119**</td>
<td>-.170**</td>
</tr>
<tr>
<td>Did not complete high school</td>
<td>.126**</td>
<td>.025**</td>
<td>.046**</td>
<td>0.001</td>
</tr>
<tr>
<td>Graduated high school</td>
<td>.209**</td>
<td>.087**</td>
<td>.079**</td>
<td>.187**</td>
</tr>
<tr>
<td>Pursued higher education</td>
<td>.064**</td>
<td>.053**</td>
<td>.043**</td>
<td>.060**</td>
</tr>
</tbody>
</table>
Table 3. Estimated Effects of Race, Gender, and Control Variables on Sentencing Lengths.

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>R-squared</strong></td>
<td>.023</td>
<td>.023</td>
<td>.036</td>
<td>.027</td>
</tr>
<tr>
<td><strong>Year</strong></td>
<td>–0.464</td>
<td>–0.494</td>
<td>–0.447</td>
<td>–0.474</td>
</tr>
<tr>
<td></td>
<td>0.017</td>
<td>0.017</td>
<td>0.016</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td>–0.040</td>
<td>–0.011</td>
<td>–0.043</td>
<td>–0.025</td>
</tr>
<tr>
<td></td>
<td>0.007</td>
<td>0.007</td>
<td>0.007</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>0.157</td>
<td>0.000</td>
<td>0.001</td>
</tr>
<tr>
<td><strong>Criminal history</strong></td>
<td>25.030</td>
<td>24.964</td>
<td>21.702</td>
<td>24.188</td>
</tr>
<tr>
<td></td>
<td>0.194</td>
<td>0.194</td>
<td>0.195</td>
<td>0.194</td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Did not graduate high school</td>
<td>–1.467</td>
<td>–2.557</td>
<td>–3.503</td>
<td>–2.696</td>
</tr>
<tr>
<td></td>
<td>0.216</td>
<td>0.222</td>
<td>0.215</td>
<td>0.216</td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Graduated high school</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11.720</td>
<td>11.458</td>
<td>11.031</td>
<td>11.469</td>
</tr>
<tr>
<td></td>
<td>0.231</td>
<td>0.231</td>
<td>0.229</td>
<td>0.230</td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Attained other degree</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.278</td>
<td>1.234</td>
<td>2.649</td>
<td>1.818</td>
</tr>
<tr>
<td></td>
<td>0.495</td>
<td>0.495</td>
<td>0.492</td>
<td>0.494</td>
</tr>
<tr>
<td></td>
<td>0.010</td>
<td>0.013</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>White</strong></td>
<td></td>
<td></td>
<td>–3.868</td>
<td>0.185</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>Female</strong></td>
<td></td>
<td></td>
<td>–</td>
<td>0.231</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>26.876</td>
<td></td>
</tr>
<tr>
<td><strong>Race and gender interaction</strong></td>
<td></td>
<td></td>
<td>0.356</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>23.762</td>
</tr>
</tbody>
</table>


Table 4. Estimated Effects of Race when Gender Is Controlled.

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th></th>
<th></th>
<th>Model 2</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>R-squared</td>
<td>.009</td>
<td>.023</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>SE</td>
<td>Sig.</td>
<td>B</td>
<td>SE</td>
<td>Sig.</td>
</tr>
<tr>
<td>Year</td>
<td>0.197</td>
<td>0.039</td>
<td>0.000</td>
<td>–0.577</td>
<td>0.018</td>
<td>0.000</td>
</tr>
<tr>
<td>Age</td>
<td>0.015</td>
<td>0.017</td>
<td>0.000</td>
<td>–0.027</td>
<td>0.008</td>
<td>0.001</td>
</tr>
<tr>
<td>Criminal history</td>
<td>12.061</td>
<td>0.385</td>
<td>0.000</td>
<td>23.774</td>
<td>0.221</td>
<td>0.000</td>
</tr>
<tr>
<td>White</td>
<td>2.879</td>
<td>0.390</td>
<td>0.000</td>
<td>–3.920</td>
<td>0.206</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Did not graduate high school</td>
<td>3.729</td>
<td>0.478</td>
<td>0.000</td>
<td>–5.471</td>
<td>0.247</td>
<td>0.000</td>
</tr>
<tr>
<td>Graduated high school</td>
<td>3.083</td>
<td>0.475</td>
<td>0.000</td>
<td>11.943</td>
<td>0.258</td>
<td>0.000</td>
</tr>
<tr>
<td>Attained other degree</td>
<td>–0.081</td>
<td>0.889</td>
<td>0.928</td>
<td>3.181</td>
<td>0.572</td>
<td>0.000</td>
</tr>
</tbody>
</table>
“OK BOOMER!”: HOW A PEJORATIVE CAN PROMOTE SOCIETAL CHANGE THROUGH HUMOR, IMPOLITENESS, AND CARNIVALESQUE

GWENDOLYN M. GOODHEAD, NORTHWEST UNIVERSITY
MENTOR: JACOB WITT

Abstract

The “OK boomer” pejorative has become very popular on social media platforms because of its ability to encode both impoliteness and humor. The encoded impoliteness and humor work together in both ironically polite and mockingly impolite ways by utilizing satire and parody to subvert and ridicule the target of the joke. This encoding creates a unique opportunity that allows those from younger generations to deride people from older generations because the two groups adhere to differing ideologies. The pejorative is also valuable for younger generations because when it is employed against older generations or even one’s own peer group, it promotes carnivalesque degradation and renewal in order to bring societal change. Although “OK boomer” may seem like a simple insult, the phrase actually follows the pragmatic rules of impoliteness and humor, prompting change through carnivalesque decay and renewal in the users it is directed at on Twitter.

In the interconnected sphere of social media platforms, differing ideologies abound within a rigid social hierarchy. There are many levels to this hierarchy; the lowest levels encompass young adults and teenagers who have little inherent power over their communities, and the higher levels contain older adults and seniors who tend to hold more sway over how their societies are shaped. When debating contrasting ideologies with people from different levels of the social hierarchy in person, one needs to attempt to find common ground while moderating one’s rhetoric in order to have a polite dialogue. When arguing on social media platforms with those who have different ideological views, however, users tend to neglect these polite formalities.

Instead of using a genuine sense of politeness, users attempt to appear civil in their online discussions by incorporating polite-sounding speech into their arguments in order to be heard by those on different levels of the hierarchy. Stephen Pihlaja and Andreas Musolff (2017) posit that the arguments between users of
differing hierarchical standings exist as a means to develop ideological communities among like-minded users and to ridicule those who do not hold these same values. Instead of upholding politeness, users on social media platforms act in certain ways and use distinct language to signal their ideologies while also mocking those who believe differently. Social media platforms such as Twitter (2021) have created guidelines to prevent users from employing hateful language like that of racism, sexism, or ageism. They enforce these policies with “permanent account suspension” to help manufacture a polite environment among people with dissimilar ideologies and social standings (para. 23). Despite these guidelines, users continue to develop new impolite words and phrases as a way to find community with some and to ridicule others in a subversive and humorous way.

One such phrase, “OK boomer,” has been used on various social media platforms such as Twitter to mock people from older generations, especially those from the baby boomer age group. Tweets using this pejorative display the versatility of the phrase and how it disrupts a sense of politeness between older and younger people on the social media platform by mocking anything from Thanksgiving guests’ opinions (Champion, 2019) to then-presidential nominee Joe Biden’s campaign bus (Holland, 2019). The pejorative is immensely popular with younger generations because it offers the chance to subtly deride others in a way that seems polite at first but actually conceals an impolite connotation within itself. “OK boomer” is also attractive because of the humor that is encoded in its underlying meaning. The phrase employs satire and parody in order to degrade the target of the joke in a subversive and humorous manner, giving rise to its continued popularity because of its power to make the teller of the joke seem clever in a subtle way.

The “OK boomer” phenomenon, which sparked substantial discussion in the press, has been mostly ignored by academics. Scholars have not yet begun researching or writing about why the “OK boomer” pejorative has developed into such a significant insult in online forums. Some researchers have attempted to understand the pragmatics of impoliteness, with linguists such as Geoffrey Leech (2014) and Jonathan Culpeper and Dániel Kádár (2010) exploring the various forms that impoliteness can assume in a conversation and the historical and cultural forms of impoliteness, respectively. Several scholars, such as Helen Davies and Sarah Ilott (2018), as well as Susan Stein (2000), have conducted research on humor and its relationship to Mikhail Bakhtin’s (1965/1984) theory of carnivalesque. These scholars write about the nature of parody and satire, with Davies and Ilott putting a particular emphasis on the satire found in modern popular culture. Although
researchers have focused on the pragmatics of impoliteness, the nature of parody and satire, and the theory of carnivalesque, a gap exists in the study of how these elements have converged to develop phrases such as “OK boomer.”

Throughout this thesis, I endeavor to illustrate how the “OK boomer” pejorative is more than a simple online insult or a passing phase. “OK boomer” is linguistically fascinating because of its compliance with the pragmatics of impoliteness and its use of subversively satirical and parodic humor in this impoliteness, as well as its proficiency at bringing change through carnivalesque humor when it is targeted at others on social media platforms. Although the “OK boomer” pejorative may seem like a simple insult, the phrase actually follows the pragmatic rules of impoliteness and humor, prompting change through carnivalesque decay and renewal in the users it is directed at on Twitter.

Impoliteness Theory

Ironic Politeness and Mock Impoliteness

Geoffrey Leech (2014), linguist and author of The Pragmatics of Politeness, posits that several kinds of impoliteness exist, but he reveals two in particular that delve into the humor surrounding impoliteness. One of these forms of impoliteness allows a speaker to say something seemingly polite on the surface but have it actually mean something entirely impolite if one looks at it more closely (p. 233). Leech explains that this form of ironic politeness is meant to be an unstable form of a polite utterance that can be interpreted only as a direct attack against the listener (pp. 233–234). He uses the example phrase “A fine friend you are!” to show that this ironically polite sentence can only be taken as a verbal attack toward the person it is aimed at because of the emphasis on specific words—“you,” in this case (p. 234). This phenomenon of listeners understanding the implicit meanings of ironically polite sayings is referred to as pragmatalization (p. 234).

The other form of humorous impoliteness is mock impoliteness, which occurs when a speaker says something inherently rude to a peer and means for that peer to interpret it as something polite (Leech, 2014, p. 238). For example, a tweet made by Denz™ (2021) mockingly references a post he made several years ago about his desire to be nice to women at the seeming expense of his dignity. Ridgeway (2021) replied to this tweet with “Was? LOL,” displaying a sense of impoliteness at first by questioning Denz’s assertion that he has changed and then following it with the acronym for “laughing out loud” to show that she was only teasing. Mock impoliteness can be used by speakers to create a sense of belonging
and camaraderie among those who are close to them “in terms of both vertical and horizontal distance” (Leech, 2014, p. 239). The distance that Leech refers to is the gauge with which people determine their relationship standing with others and how close they are to these people. If one feels as though one is close to someone both vertically and horizontally, one feels as though one is on the same level of the social hierarchy and within the same peer groups as the other person, allowing for a sense of camaraderie between the two people. This form of humorous impoliteness is very effective in situations where everyone gathered feels comfortable mocking each other and trading friendly banter; a sense of familiarity with those present is required to achieve this humor (p. 240). According to Leech, this type of impoliteness cannot be used by strangers, as the inherent politeness would be lost on the listener, given the lack of camaraderie (p. 239).

**Sarcasm and Parody**

Delving further into how these two forms of humorous impoliteness operate, two different forms of humor emerge to work in tandem with these two impolite forms of speech. Ironic politeness employs satire in an effort to attack the object of the joke in a way that will make the speaker and other listeners laugh. Susan Stein (2000) explains that the goal of satire is to show that the speaker is better than the person being satirized in both morality and intellect, while also showing that satire is “more serious than playful” in its humor when compared to parody’s unique jollity (para. 6). This sense of inherent superiority is evident in ironic politeness when a speaker means to ridicule the target of a joke subtly while showing that the speaker is being moralistic in their apparent politeness. It also allows the speaker to reveal their higher intelligence if the person who was ridiculed did not understand that they were being satirized. Because of this fact, satire cannot be separated from ironic politeness during its utilization. Moreover, to be taken as a nicety by the other person, it requires a sense of seriousness in its delivery.

Just as ironic politeness makes use of satire in its humor, mock impoliteness utilizes parody to create laughter and camaraderie. Stein (2000) discusses the idea that parody is more concerned with imitating objects and people in humorous ways, as well as with being more playful in its usage, than is satire. As mock impoliteness is used to create a feeling of equality among peers, parody works well in achieving this feeling because it makes light of seemingly important issues such as politics or societal norms by imitating them in a funny way. Parody creates a space within peer groups that allows members to exchange playful banter and to liken each other to things or people that would be considered humorous in that space.
Leech (2014, pp. 239–240) explains mock impoliteness, and specifically playful banter, by using an example of a man who went to an unfamiliar bar and attempted to get a drink. This man tried being polite with the bartender but managed to get a drink only when one of the regular patrons stood up for him by calling the bartender a “dozy bastard” in a friendly and endearing way in order to get the bartender’s attention. In this instance, the bartender was most likely not acting as a “dozy bastard” and the regular patron called him this name only because he knew the bartender well enough that such a comparison would be funny and in no way insulting to the bartender. Mock impoliteness and parody are used together to develop a relaxed, fun relationship between friends.

Although a strict dichotomy may be struck between ironic politeness and mock impoliteness, the jokes within each category can be incredibly diverse. Deborah Tannen (2005) writes about this idea after recording her family’s sense of humor at a Thanksgiving meal (p. 164). She finds varying levels of seriousness in joking, with some of her relatives using explicit emphasis in their humor in order to make the joke obvious and others favoring a more deadpan, serious tone in their jests (p. 175). Jokes in either the ironic politeness or mock impoliteness category can take on a more serious or more playful tone depending on the speaker’s form of humor.

Within Leech’s (2014) discussion on ironic politeness, he uses a few examples of humor that employ different tones. These can vary from emphatic, obvious humor to biting rebukes masked in polite tones. He provides one example of a woman who was late to work being greeted by her boss with a very emphatic “Good afternoon, Sue,” as well as an example of a biting “Sorry I asked!” apology (p. 235). Speakers use a vast range of tones within the two separate categories of impolite humor, allowing for incredibly diverse joking within these spheres of ironic politeness and mock impoliteness.

**Impoliteness Theory and “OK Boomer”**

With this context in mind, applying Leech’s (2014) theory to the “OK boomer” pejorative reveals several key nuances. This paper examines posts pulled from significant dates on Twitter, namely when the phrase “OK boomer” began to appear regularly on the platform (January 1–February 1, 2019), when the phrase peaked in popularity (November 1–December 1, 2019), and as its usage declined (January 1–February 1, 2020). All of these tweets were pulled from threads that featured the phrase “OK boomer,” although they represent only a small selection of a vast number of tweets using this pejorative.
Approximately three-quarters of the “OK boomer” tweets pulled can be labeled ironically polite, and the remaining tweets can be considered mockingly impolite. A few tweets did not fit clearly into either category and were left out of my analysis. In addition to these outliers, the tweets found within the two categories can employ a wide range of tones, from good-natured fun to rather scathing. I have endeavored to place each pulled tweet into the category that best describes its implicit meaning, but this designation is inherently subjective and must be taken into consideration while reading this essay.

Ironically Polite Tweets

Examples abound from the ironically polite category that detail how the “OK boomer” pejorative can be deftly used satirically to ridicule the object of the attack. A Twitter user by the name of Ny (2019) used the pejorative to mock another user who claimed that “Google should not be your only check” when fact-checking journalists (Bund, 2019). The implied meaning in this instance is that this other user was out of touch with the way that people use Google to find out whether a news outlet is spreading factual evidence and that the user should stop attempting to reprimand others for being innovative in how they get their news information. Another example of this use was displayed when the musician Jacob Sartorius (2019) tweeted some general ideas that he typically heard from adults. He tweeted, “I LOVE SCHOOL EDUCATION IS THE MOST IMPORTANT THING YOU WONT [sic] BE ANYTHING WITHOUT SCHOOL,” and replied to his own comment with the phrase “OK boomer” to summarize that these general sentiments about education are old-fashioned and not relevant to modern youth.

Both of these examples fit well into the ironically polite category because of their inherent satirical tone in using “OK boomer” to deride their targets on Twitter. These tweets using the pejorative are meant to sound either neutral or even agreeable on the surface, but their position as ironically polite utterances leaves no room for them to be understood as anything other than biting humor at the expense of their targets. As an added bonus for those who use this pejorative on Twitter, it “boosts the face of the ironist while attacking the face of the target” because the one who uses the phrase lampoons the target while showing their own cleverness in the situation (Leech, 2014, p. 235).
Mockingly Impolite Tweets

Along with the examples of tweets in the ironically polite category, some tweets showcase mockingly impolite parodic humor. Many of these tweets use the pejorative self-deprecatingly. One such example was written by Eliza Ronalds-Hannon (2019), who shared a story about her mom “referring to herself affectionately—and in [third-person]—as ‘Boomer.’” Ronalds-Hannon’s tweet explains that her mom realized that she might be doing things in a slightly old-fashioned way and was lightheartedly mocking herself by calling herself a boomer. Another example employing the “OK boomer” pejorative in an even more parodic way is a tweet by Juiblex (2019) that makes a series of observations, such as “Me pointing at an old person: ok boomer” and “Me pointing at a flower: ok bloomer.” This user was making a mockery out of the pejorative itself by using it to refer to an older individual and then rhyming it with other words such as “bloomer” and “groomer” to make it lose its satirical power.

These two examples fit well into the mockingly impolite category in that they use parody implicitly to connect with others and make them laugh. The mother who calls herself a boomer most likely knows that the word has taken on a pejorative meaning and so is making this word inclusive and a parody of its new usage by lightheartedly calling herself a boomer (Ronalds-Hannon, 2019). In the same manner, Juiblex (2019) used “OK boomer” in a parodic way by rhyming the term “boomer” with rhyming words, causing the term to lose the power it retains by infusing it with lighthearted humor. Both of these users were having a good-natured laugh with other Twitter users by mocking themselves and the “OK boomer” pejorative to connect with their peers who also enjoyed this type of humor.

As previously mentioned, on Twitter, the usage of “OK boomer” in an ironically polite way is significantly higher than its usage as a mockingly impolite form of humor. Although the pejorative can be used to create camaraderie between peers through parody and lighthearted fun, its main purpose is to satirize a particular target in a covert way. The pejorative is versatile, but most of the posts that feature the phrase are focused on getting a quick laugh at the expense of the target.

The Ideology Behind “OK Boomer”

The Encoded Meaning of “OK Boomer”

From these two categories, a distinct topic trend is emerging in the usage of the “OK boomer” pejorative. Within the mockingly impolite category, the topics
stayed fairly similar throughout the research period. These topics include popular culture, parody, self-deprecation, and friendly banter. In the ironically polite category, the topics shifted a bit more. At the beginning and middle periods of the pejorative’s lifespan, the topics dealt with politics, technology, and provoking an anger response in other people. The current usages of the pejorative still include the ones previously mentioned but also deal with popular culture and daily life. The distinct topics of the “OK boomer” pejorative found in the ironically polite and mockingly impolite categories are starting to meld together into a focus on the “boomer” as a person who is out of touch with both popular culture and the daily life experienced by younger generations. The current usages of the pejorative still include the ones previously mentioned but also deal with popular culture and daily life. The distinct topics of the “OK boomer” pejorative found in the ironically polite and mockingly impolite categories are starting to meld together into a focus on the “boomer” as a person who is out of touch with both popular culture and the daily life experienced by younger generations. This is different from the pejorative’s initial usage as referring only to someone who talks about politics or does not understand technology. As these two categories begin to share a definition of a boomer as someone who is out of touch with current trends in areas that are important to younger generations, both the ironically polite and mockingly impolite usages of “OK boomer” highlight how much the younger and older generations have diverged.

Differing Ideologies

Just like a time capsule, the “OK boomer” pejorative displays the frustrations that people from younger generations have experienced for many years in interactions with people of older generations. The amount of scorn directed at older people for their believed indifference toward the issues that younger people are fighting against is funneled into this pejorative phrase, which is then hurled satirically at anyone who does not agree with the common views of younger generations. Taylor Lorenz, a journalist at The New York Times, discusses this idea in a 2019 article that brings up the pejorative from the perspective of those who were born into Generation Z, a generation that currently includes both young adults and teenagers. Lorenz writes that many people from Generation Z believe that older people are unconcerned with climate change and were brought up on old-fashioned ideas with which younger generations today disagree. This divergence of beliefs and opinions about the world has led to the creation of a pejorative that can dismiss the ideas of older people humorously by simply reminding them that they are older and therefore out of touch with the reality of the world that younger generations are inheriting.

The ideological divergence that has occurred between older and younger people highlights a hierarchical structure in which the older generations hold much of the power over societal conduct. To illustrate this ideological divergence, Lorenz
(2019) writes about several younger people’s opinions about how older people disparage the beliefs of Generation Z. One of the people whom Lorenz interviewed shared her viewpoint that older people are out of touch with the issues that younger people deal with and that “[e]verybody in Gen Z is affected by the choices of the boomers, that they made and are still making. … Those choices are hurting us and our future. Everyone in my generation can relate to that experience and we’re all really frustrated by it” (para. 8). From the perspective of younger people, the hierarchical order is such that older generations are able to shape the future of the next generations, and one of the only ways that younger people can have a say in their future is by using the “OK boomer” pejorative to call out the mindsets and behaviors of older people.

The Theory of Carnivalesque and “OK Boomer”

Carnivalesque Equality

The hierarchical structure that favors older generations over younger ones reflects the hierarchy that was in place during the days of medieval carnivals. According to literary theorist Mikhail Bakhtin (1965/1984), the medieval carnival allowed a sort of “temporary liberation … from the established order; it marked the suspension of all hierarchical rank, privileges, norms, and prohibitions” (p. 10). In the medieval carnival, all people were considered equal, in direct opposition to traditional feasts, which exemplified hierarchical inequality (p. 10). During the carnival, people from different levels of the social hierarchy could interact with each other in ways that were normally not permitted, allowing them to temporarily do away with polite societal practices and to exchange the normal practices for good-natured humor among equals (p. 10). The carnival was therefore a place that produced openness and equality for all.

Degradation and Renewal

The ideas and common practices of the medieval carnival are developed and distilled into a theory that Bakhtin (1965/1984) terms “carnivalesque” (p. 15). This theory posits that two important stages need to occur in order for someone to not only experience equality with others but also bring about change in a society (p. 10). The first stage is degradation, which occurs when a speaker parodies or satirizes their target to the extent that what was once sacred and untouchable is now vulgar and material (pp. 20–21). Degradation causes everyone to be situated on the lowest level of the hierarchy together. The second stage happens after a target has
been thrown “down to the reproductive lower stratum, the zone in which conception and new birth take[s] place” (p. 21). The target is thus degraded to the point where a rebirth through humility is possible. This stage allows renewal to begin in the target and “bring[s] forth something more and better” in the target’s life (p. 21). Degradation and renewal work in tandem to radically improve societies and those living in them.

The two stages of carnivalesque humor create a space for social correction in a less threatening way by masking the reprimand in a joke made at the target’s expense. Degradation can be used by anyone on any level of the social hierarchy to subvert and mock those above them, and the renewal that follows can develop a sense of humility in the target. Bakhtin (1965/1984) posits that these two stages also affect the creators of the jokes, as well as the audience members who simply listen in on the joking (p. 12). The equality and interconnectedness that everyone experiences when taking part in this degrading and renewing humor work to bring reformation to the societal practices of as well as the people who live in these communities. The humor of carnivalesque points societies toward utopian living by asking them to embrace humility and their own humanity (p. 12).

The stages of degradation and renewal are inextricably linked with humor, relying on laughter to bring about change in a society. As mentioned previously, ironic politeness employs satire to ridicule a target covertly, whereas mocking impoliteness uses parody to evoke laughter. In the case of societal change through degradation and renewal, ironic politeness requires a very biting piece of satire that inspires humility in a target who is higher in the social hierarchy, whereas mocking impoliteness requires that both the speaker and target be from the same peer group and share a jovial camaraderie in order to create change. Both satire and parody can be used to change a society, but they require vastly different tones and intensities to create the change sought by the speaker.

**Ironic Politeness and Carnivalesque in “OK Boomer” Tweets**

The “OK boomer” pejorative functions on Twitter as an attempt to bring about societal change humorously. Some users on Twitter employ the pejorative in ironically polite tweets with the hope that people will try to think and behave in a fashion similar to that of younger generations. One user tweeted that they were seated near an older gentleman who let his phone ring while in a movie theater, prompting the user to ask the man to turn off his phone (Ho Toy Noodle, 2019). The man told the user, “[I] don’t even know how to work it,” to which the user replied that the man should have left the phone behind if he did not know how to
silence it. The user then followed this remark with “OK BOOMER?” to signify confusion about the fact that the older man did not know how to operate his phone, as well as to imply that the man needed to learn more about current technology. Another user tweeted that she spent a substantial amount of time painting a piece of intricate artwork but her mom noticed only the background colors of the painting (King Max⁷⚡♛, 2020). This user followed her mother’s comment by saying, “ok boomer,” suggesting that she thought her mom did not understand her modern artwork and needed to develop a sense of what is considered tasteful in art today. Both of these users employed ironically polite usages of the “OK boomer” pejorative to satirize their targets and to try to change their targets’ behavior or thinking on issues important to them.

These two tweets follow the stages of degradation and renewal found in carnivalesque humor. The first tweet, from a user who sat next to an older gentleman, uses satire against the older man—on a higher level of the social hierarchy—in order to degrade him by making him seem incredibly incompetent at operating a phone. Because this gentleman is implied to be older than the user, the implication is that he has more power and a higher position in the hierarchy, which causes the user to strive for equality with him by degrading him to the level that the user resides on. The tweet also involves renewal, in that the older gentleman is humbled on Twitter and given a chance to increase his technological skills. In the second tweet offered as an example, the artist’s mother is also both degraded and renewed on Twitter. The mother experiences degradation by having her comment ridiculed in the tweet, but she also finds renewal when her opinion about art is humbled and she is implicitly asked to develop a more modern sense of artistic taste. The degradation and renewal represent an attempt by the daughter to humble her mother in order to have a better connection with her mother, who is in a higher position on the social hierarchy.

Mock Impoliteness and Carnivalesque in “OK Boomer” Tweets

Twitter users have similarly utilized “OK boomer” in a mockingly impolite fashion to bring about change through friendly humor. A user named Nauticalist (2019) employs the pejorative in a reply to a user who typed the wrong word when referring to a collection of emoticons on their phone (Joe, 2019). The pejorative is used in a humorous way for the purpose of lightly correcting the other user on word choice in a nonthreatening manner. The reasoning for using the pejorative here is that the second user should have known the correct word because it is related to a frequently used form of technology. Moreover, Nauticalist is simply trying to make
Another Twitter thread features the “OK boomer” pejorative in a conversation between a couple of friends. One user tweets that he does not “play games on [F]ridays,” to which his friend Goldy responds with the “OK boomer” pejorative (Goldy, 2020; TUCKER, 2020). This dialogue seems harsh at first, but Goldy is just lightly criticizing a friend for not playing on a Friday. The usage of “OK boomer” implies that Goldy’s friend is acting like an older person in not wanting to play games on a Friday in the way people their age normally do, and Goldy is trying to get the friend to change by pointing out this detail.

These tweets exemplify both degradation and renewal in their usage of the “OK boomer” pejorative. When Nauticalist (2019) lightly corrects their friend, this degrades the friend by humorously pointing out the incorrect usage of the word “reel” and showing that this incorrect usage makes the user seem like they are acting above their social position in the hierarchy. The errant user experiences renewal when replying to Nauticalist by saying that they “like to switch things up now and again,” implying through humor that they know they have made a mistake in word usage and are going to try to not make the mistake again, in order to remain relevant and relatable to their friend (Joe, 2019). When Goldy (2020) mocks his friend for not playing games with him on a Friday, he is degrading his friend by showing that his friend is acting outside of the friend’s position within the social hierarchy by not accepting Goldy’s invitation to play games. Renewal for the friend occurs because the friend has to consider changing their usual Friday activities to include playing games with Goldy in an effort to maintain their position in this peer group.

Carnivalesque and the Modern Satire of “OK Boomer”

Although the “OK boomer” pejorative employs satire in an ironically polite fashion in order to degrade and renew others on Twitter, Bakhtin’s (1965/1984) views on modern satire must also be addressed. Bakhtin writes that carnivalesque is an inclusive laughter that draws all people together into both degradation and renewal (pp. 11–12). The modern satirist fails to draw others together by trying to position themselves as above the target of their joking while also failing to include everyone in their humorous appeal (p. 12). The nature of online formats such as Twitter necessitates that all tweets, including those involving the “OK boomer” pejorative, are public, and tweets are posted in a way that allows everyone to see both the pejorative tweet and the thread on which it has been posted. This ability to
look at tweets using the “OK boomer” phrase makes the satire in these tweets open for all to laugh at so all can join in the degradation and renewal of carnivalesque humor. A case may be made that the satire of these tweets can still prop the writer up as wittier and above the target of the joke but can also be just as effective at pointing out hierarchical injustice. Helen Davies and Sarah Ilott (2018) posit that satire is capable of “offer[ing] a critique of entrenched values about … social power inequalities” (p. 11). Even though a satirist can use the “OK boomer” pejorative to place themselves above the target of their joking, they can also use the pejorative to subvert the power of social hierarchies by illustrating this inequality humorously. Bakhtin may not have believed that modern satire could bring carnivalesque degradation and renewal, but I want to suggest that the “OK boomer” pejorative uses satire in order to subvert social hierarchies and prompt carnivalesque change in others.

**Conclusion**

The “OK boomer” pejorative is incredibly good at employing the pragmatics of impoliteness, humor, and carnivalesque to establish change in those who are on higher levels of the social hierarchy and in those who reside on the same level in online social media platforms. Because this pejorative is still new, more research needs to be conducted in an effort to understand if another pejorative like this one will find its way into popular online forums; if it does, researchers will need to discover whether it becomes popular for the same generational reasons. Some of this research could include understanding how younger generations became so divided from older generations that they feel the need to create a pejorative in an effort to subvert and change the behavior of other people. Another aspect of this research could focus on the humor that Bakhtin (1965/1984) addresses in his book and could apply that humor to areas online that display subversive tendencies, such as Internet memes. All of these questions could be raised in order to better understand how humor, impoliteness, and carnivalesque come together to create the “OK boomer” pejorative.

The many intricacies of “OK boomer” show the malleability and adaptiveness of language. Jokes that at one time needed inflection and tone of voice in order to be comprehensible can now be understood simply through implied meaning in the particular choice of words that one utilizes. The “OK boomer” pejorative is understood by younger generations as either an ironically polite remark toward people who disagree with their viewpoints or a mockingly impolite way to build camaraderie with their peers. Younger generations also implicitly
know that this pejorative is meant to bring change through degradation and renewal. It would seem that younger generations have developed a means to detect, comprehend, and find humorous the implicit messages encoded in text written on social media platforms. The greatest question that remains from this research is whether older generations will adapt to comprehend the encoded messages found in phrases such as “OK boomer” and their viewpoints change based on their understanding of these messages.


References


Culpeper, J., & Kádár, D. Z. (Eds.). (2010). *Historical (im)politeness*. Peter Lang. https://books.google.com/books?id=H1bQARN3bk8C&printsec=frontcover&dq=pragmatics+of+politeness+most+notable+linguists&hl=en&newbks=1&newbks_redir=0&sa=X&ved=2ahUKEwix2JDLoaIVLoZ4KHV0NDcQ6AEwA3oECAAAQAg#v=onepage&q=pragmatics%20of%20politeness%20most%20notable%20linguists&f=false


Denz™ [@TheDopest_D]. (2021, January 10). Why was I such a simp [Face with tears of joy emoji x3] [Image attached]. [Tweet]. Twitter. https://twitter.com/TheDopest_D/status/1348440599158870016


Ho Toy Noodle [@hotoynoodle]. (2019, November 30). sat next to one in the movies the other day. his phone rang. it was in his shirt pocket. it [Tweet]. Twitter.
https://twitter.com/hotoynoodle/status/1200782785037905924?s=20

Joe, D. [@Dino4Congress]. (2019, January 8). It was the first dopey face I found while scrolling through my reel. [Tweet]. Twitter.
https://twitter.com/LoGravityTweets/status/1082848257141690368?s=20

https://twitter.com/LordJuiblex/status/1200847411893227520?s=20

King Max⚡♛ [@peachijeon]. (2020, January 31). Me takes hours on this:
Mom: I like the background Me: ok boomer. [Image attached] [Tweet]. Twitter.
https://twitter.com/peachijeon/status/1223349173094944771?s=20

https://www.academia.edu/33129448/The-Pragmatics-of-Politeness


https://twitter.com/LoGravityTweets/status/1082848257141690368?s=20

https://twitter.com/TyranTheRed/status/1088683682729209861?s=20

https://eds.b.ebscohost.com/eds/ebookviewer/ebook/bmxlYmtfXzE1OTUzNDBfX0FO0?sid=c6e29acb-f47c-4964-afca-07ea2bb5a032@pdc- vessmogr02&vid=1&format=EB
https://twitter.com/WinterRidgeway/status/1348442792255651840

https://twitter.com/ElizaHannon/status/1200794168878125056?s=20

https://twitter.com/jacobsartorius/status/1200853086744522753?s=20

https://search.proquest.com/docview/219711590?accountid=28772

https://books.google.com/books?id=0ZdhWYV_Re0C&printsec=frontcover&dq=inauthor:%22Deborah+Tannen%22&hl=en&newbks=1&newbks_redir=0&sa=X&ved=2ahUKEwj35J7b6u_nAhVSLs0KHboUAR4Q6AEwCXoECAgQAg#v=onepage&q&f=false

TUCKER [@JERICHO]. (2020, January 31). I don't play games on fridays, but tomorrow Im down. [Tweet]. Twitter.
https://twitter.com/GoldGloveTV/status/1223318255420370944?s=20

Appendix A

This appendix provides the research that I have compiled regarding the “OK boomer” pejorative on Twitter. The hyperlinks listed have all been pulled from Twitter, using Twitter’s search feature to look for the phrase “OK boomer” at specific points in the lifespan of the pejorative. I separated these tweets into both ironically polite and mockingly impolite categories and divided them further into the specific date ranges from which they were searched.

As there are far more tweets to choose from than is possible with simple searching and the timeframe available to me for this work, more research could be done in this area to discover the most accurate percentage of ironically polite and mockingly impolite tweets. This research can be a good start to a more in-depth assessment of how the “OK boomer” pejorative is utilized on Twitter.

Ironically Polite Tweets

January 1–February 1, 2019

• https://twitter.com/boomerwache/status/1089841989569236992?s=20
• https://twitter.com/TyranTheRed/status/1088683682729209861?s=20
• https://twitter.com/VegetasLeftNut/status/1085905553962950656?s=20
• https://twitter.com/Zhir_Vengerskye/status/1085430553665429505?s=20
• https://twitter.com/Dalmas_i_vinden/status/1087229805844676608?s=20
• https://twitter.com/MrsFifn/status/1083400402232578053?s=20
• https://twitter.com/yudovico/status/1082579592748118016?s=20
• https://twitter.com/slmyers11/status/1082137206578241536?s=20
• https://twitter.com/katejellybean/status/1081650921019338752?s=20
• https://twitter.com/Lea_Avi/status/1080880271367593984?s=20

November 1–December 1, 2019

• https://twitter.com/YerboutiSheik/status/1200926969015283712?s=20
• https://twitter.com/nathanhartswick/status/1200910283776176128?s=20
January 1–February 1, 2020

- https://twitter.com/fcarucci/status/1223390849528000512?s=20
- https://twitter.com/DJStinky/status/1223333974686171136?s=20
- https://twitter.com/Alykkat/status/1223329325698535424?s=20
- https://twitter.com/_ChemaMendiola/status/1223384361992622080?s=20
- https://twitter.com/YeetProduction/status/122342863842906112?s=20
- https://twitter.com/carleebullock7/status/1223356733076066305?s=20
Mockingly Impolite Tweets

January 1–February 1, 2019

• https://twitter.com/LoGravityTweets/status/1082848257141690368?s=20

November 1–December 1, 2019

• https://twitter.com/monteiro/status/1200915302315225088?s=20
• https://twitter.com/ikoneedy/status/1200924687318028289?s=20
• https://twitter.com/Rx_Pixel/status/1200910397185748992?s=20
• https://twitter.com/TroyDreyfus/status/1200871062285770754?s=20
• https://twitter.com/LordJuiblex/status/1200847411893227520?s=20
• https://twitter.com/weareyourfek/status/1200832235210866688?s=20
• https://twitter.com/ElizaHannon/status/1200794168878125056?s=20
• https://twitter.com/Rachel_Sennott/status/1200781833467772932?s=20

January 1–February 1, 2020

• https://twitter.com/GoldGloveTV/status/1223318255420370944?s=20
• https://twitter.com/jasonkersey/status/1223296235261644800?s=20
MULIER EST URBIS CONFUSIO?—THE FUNCTION OF RHEA SILVIA, TARPEIA, AND HORATIA AS EXEMPLA IN LIVY’S THE EARLY HISTORY OF ROME

BHADRAJEE S. HEWAGE, PRINCETON UNIVERSITY
MENTOR: ANDREW FELDHERR

Abstract

In The Early History of Rome, Livy focuses his narrative upon the moral, emphasizing it through the employment of several devices of literary art. Given that the Roman women Rhea Silvia, Tarpeia, and Horatia all form part of the colorful cycle of stories that make up the early growth and development of Rome, this paper explores how these women function as exempla and how Livy frames them in the broader context of his masterpiece. Livy uses these female exempla not in a factual record but rather in a masterpiece of moral instruction, with the women serving as exempla of a flexible or open kind.

The study of history is the best medicine for a sick mind; for in history you have a record of the infinite variety of human experience plainly set out for all to see; and in that record you can find for yourself and your country both examples and warnings; fine things to take as models, base things, rotten through and through, to avoid.

—Livy, The Early History of Rome, Preface

Livy’s famous preface clearly proclaims that history is a valuable moral device. The specific wording of the preface clearly connects history and historiography as partners in moral instruction, yet Livy’s use of both does not necessarily mean that his use of either results in clear-cut moral lessons for the reader to simply internalize. “History” was full of “records,” or exempla, for moral agents of Livy’s time. Through understanding of the events of history, proof “of the infinite variety of human experience” can be found. According to scholar Tara Welch, Livy “draws in his audience not only as impressionable moral agents but also as readers and learners” (2015, p. 137). Classicist Jane Chaplin adds, “At any point, Livy may intend his various internal and external audiences to respond to the same exemplum in different ways” (2000, p. 4).
Livy’s exempla evidently have multiple audiences in mind, but in using them as moral devices, Livy seeks to establish an important relationship between those within the text and those outside it. The audience therefore appears to occupy a central role in the interpretation of these exempla, given the myriad ways to respond and to take moral lessons from them. Indeed, if “history” and “records” are the source materials that form the content of Livy’s chef d’oeuvre, the use of these source materials as teaching and advising tools to “find for yourself and your country” and “to take as models” reveals the means by which the constituent material comes to the reader. This means the historiography that refers to the meeting point between the material and the learner. It is not, therefore, the past through which Livy delivers help to the audience as “impressionable moral agents” but rather the interpretative “study” of that past, yet what is the role accorded to women such as Rhea Silvia, Tarpeia, and Horatia in this interpretative exemplum scheme?

Welch (2015) writes that the three women exist in a gap while at the same time being that which constitutes the gap while also acting as the means by which that gap is bridged. She explains, “Rhea Silvia is the point of connection between the Alban kings and Rome’s rulers; Horatia is both Roman and Alban; and Tarpeia exists in the moment between complete Sabine hostility and Sabine cooperation.” She suggests that through the actions of Rome’s early men, Livy highlights to his reader the centripetal forces at play, driving the development of his pluralistic city. Indeed, Roman men are at the center of continually driving Livy’s narrative forward. The positions of these founding-figure women and the colorful cycle of their stories, however, indicate to Welch the centrifugal forces that are also at play in Rome’s foundational stories. She writes that in positioning these women as capable of moving away from Roman norms, Livy highlights how Roman women are able to counteract the city’s patriarchal centripetal forces (2015, p. 163). This paper therefore explores how these women function as exempla and how Livy frames them in the broader context of his masterpiece. This paper will analyze these female exempla not as a factual record but rather, as Livy himself intended in his preface, as a masterpiece of moral discourse. Indeed, these women all serve as interesting exempla of a flexible or open kind and highlight the tension between the centripetal and centrifugal forces in operation during Rome’s founding.

Rhea Silvia features early on in Livy’s work, given her primary contribution as the mother of the fraternal twins Romulus and Remus; however, it is possible to understand her as an innocent victim of Amulius’s lust for his brother’s throne. Tom Stevenson notes, “The common view … is that she was raped by the war god
Mars and gave birth to twins. ... For her ‘crime,’ she was shackled, thrown into prison and not heard from again” (2011, p. 176). Livy naturally does not dwell much on developing her character, yet Welch indicates how Livy does explicitly describe the tyrannical behaviour of Amulius, who confines Rhea Silvia as a Vestal Virgin for having annihilated his brother’s male stock (2015, p. 162).Appearances can lead to deception, and what seems like a prestigious honor bestowed on Rhea Silvia is in fact a tactic to stem the family line. As with the other women featured in Livy’s exempla, Rhea Silvia is thus also victim to the extreme control that Roman menfolk seek to exert on the city’s women.

Livy does not definitely relate that Rhea Silvia was violated by Mars, however, writing instead, “Mars, she declared, was their father—perhaps she believed it, perhaps she was merely hoping by the pretence to palliate her guilt” (Livy 2002, p. 34). The absence of an equivocal statement is important, as while the violator’s identity is kept ambiguous, the innocence and moral purity of this exemplum are called into question. Indeed, the reader cannot but question the truth surrounding not only Rhea Silvia but also her sons and Amulius himself. Did she undermine her vows as a Vestal Virgin? Were Amulius’s actions toward her for undermining her vows in themselves indicative of those of a tyrant? Who, exactly, was the boys’ father? Although the danger posed by women in incorporating external bloodlines is readily present here, the record remains unresolved and is certainly complicated and intricate rather than simple and straightforward. What is clear, however, is that Rhea Silvia exists in a gap and serves as a point of connection between the Albans and the Romans.

Livy’s usage of exempla also motivates the reader to delve more deeply into each particular record and to understand more the general process of deciphering them. Regarding Tarpeia, another Vestal Virgin and daughter of the Roman commander Spurius Tarpeius, Stevenson describes how common accounts of her story reveal her as “a traitor who receives just punishment” (2011, p. 178). A simple reading of Livy’s narrative uncovers how Tarpeia betrayed Rome to the leader of the Sabine army, Titus Tatius, and consequently was murdered as punishment. Livy offers three different explanations to justify the incentive behind Tarpeia’s perfidy, however, thus again leaving the reader questioning the true interpretation that can be obtained from this exemplum. Livy writes first, “When she had gone outside the walls to fetch water for a sacrifice, [Tarpeia] was bribed by Tatius … to admit a party of his soldiers into the fortress” (2002, p. 44). The blame is thus first posited on Tatius and the Sabines to show that Tarpeia’s treachery came not from greed but from a scrupulous bribe. It is almost as if her betrayal of Rome was not her direct
fault but rather the consequence of Sabine bribery. Although Tarpeia did indeed “admit a party of his soldiers into the fortress,” her motivation came from the force of attraction rather than from inherent greed or selfishness.

Yet Livy conveys another account of Rome’s betrayal by remarking that Rome’s exposure came not from bribery but from greed itself: “There is also a story that this girl had demanded as the price of her services ‘What they had on their shield-arms’ ” (2002, p. 44). Here, Livy posits that Tarpeia’s avarice was the key factor in her ultimate death. Such was her greed and rapacity that she, in essence, sold out her city, the disgust at and aversion to which prompted Rome’s very enemies to heap their shields upon her, resulting in her death.

Finally, Livy’s third version of events explaining Tarpeia’s perfidy fuses the two previous accounts: “Some say that after bargaining for what they ‘had on their left arms’ she did actually demand their shields” (2002, p. 44). Similarly to his previous account, Livy again blames Tarpeia here, but unlike in the prior version, Tarpeia in this one directly demands the shields, rather than the gold, of the Sabine soldiers. In this scenario, Tarpeia even appears to be acting in Rome’s best interests. As Stevenson keenly notes, she can be interpreted as a “national heroine, who attempted to disarm the Sabines by trickery” (2011, p. 179).

Although these three versions feature the same result—Tarpeia’s eventual death at the hand of the Sabines—the true meaning behind Tarpeia’s exemplum is thus difficult to discern. Livy’s presentation of Tarpeia’s tale clearly leaves much material for the reader to analyze. What is beyond doubt is that as a direct result of Tarpeia’s actions, Rome was betrayed and Tarpeia received a fatal punishment. What is clear, however, is that Livy’s description of alternatives frames the overall narrative of the story. His refusal to come down and make a judgment on one side or the other opens Tarpeia’s exemplum to the interpretation of the reader. Indeed, after relating the three various points of view, Livy moves seamlessly to advance his general narrative, immediately continuing, “The Sabines were now in possession of the citadel” (2002, p. 44). Welch concludes, “In Tarpeia’s case, Livy places himself in the maze-walker’s position, confronted with the forks and paths of alternative traditions amongst which he may choose” (2015, p. 145), yet according to Stevenson, “[t]he implications of the various traditions once more frustrate resolution” (2011, p. 179). Given so many options for the original narrative of Tarpeia’s betrayal beyond the mere fact that she opened Rome to the Sabines, doubt exists. The meaning of an exemplum is thus not fixed or set but evolves through time and with context. The point behind each of Livy’s exempla therefore appears to be more about the process itself than the end product. Livy’s skill lies in
crafting the development of Tarpeia’s exemplum rather than in providing a clear resolution.

The record of Horatia, which gives a far more comfortable outcome than that of Tarpeia, offers a particularly strong point of comparison. As Welch explains, “In contrast to Tarpeia’s story, in which no one is heroic, Horatia’s story is full of commendable behaviors” (2015, p. 136). Horatia’s exemplum is conventionally described in relation to family and state disloyalty. Indeed, her life coincided with the hostilities between Rome and Alba Longa, which featured a prominent battle between her brothers, the Horatii, and three Albans, the Curiatii. When the sole surviving Horatii returned victorious to Rome but saw his sister mourning the death of a Curiatii to whom she had been betrothed, he immediately killed her, exclaiming, according to Livy, “What is Rome to such as you, or your brothers, living or dead? So perish all Roman women who mourn for an enemy!” (2002, p. 61). As Stevenson notes, “If the story were to end at this point, a superficial reading might generate sympathy for Horatius’s point of view. ... The story continues, and the events which follow make it clear that the simple reading of Horatia as a traitor to her family and to Rome is inadequate” (2011, p. 182). In fact, Livy proceeds to describe a deep difference of opinion among those who were present at Horatia’s killing.

Indeed, Horatia’s story also does not have a clear-cut resolution. As scholar Joseph Solodow outlines, “The people initially had been ambivalent about Horatius upon his slaying of his sister … and at the end they are still of two minds” (1979, p. 257). The debate surrounding whether Horatius’s heroism on behalf of Rome outweighed the murder of his sister ultimately resulted in the acquittal of Horatius. As Livy notes, “Though he was guilty in law, popular admiration of his quality obtained his acquittal” (2002, p. 62). Horatia’s record seems to indemnify everyone, with Horatia garnering sympathy from the crowd because “there were none who did not feel the horror of this deed” (Livy, 2002, p.61) and the law seemingly vindicating Horatius.

This is in contrast to Tarpeia’s tale, which flatters no protagonist. Tarpeia’s commonly viewed status as a renegade thus also becomes unsure. Her exemplum appears to hint at an underlying instability of sacrifice, in which the Roman community sympathizes with both the sacrificant and victim. Horatius the sacrificant, having defended the honor of his family and state, becomes willing to kill his own sister for dishonoring both, with the sacrifice seemingly expiating the wrong in the eyes of both their father and early Roman law. Horatia the victim, despite dishonoring her own clan and state by mourning the enemy, has also lost
her betrothed and her hopes for the future. The conditions thus appear for the reader to sympathize with both Horatius and Horatia, despite the gravity of both of their actions. As Soldow notes, “Even at the end … we, like the people of Rome, cannot be sure how to judge” (1979, p. 257).

From the characters of Rhea Silvia, Tarpeia, and Horatia, it is therefore clear that Livy gives a definite degree of attention to records involving women within his general foundational narrative. Rhea Silvia serves to connect Rome’s rulers with Alban kings. Tarpeia exists during the moment between Sabine belligerence and Sabine collaboration, and Horatia is in essence both Roman and Alban. Through their function as exempla, the women appear to highlight the tension between the centripetal and centrifugal forces at play in Rome’s founding story. Thus, as Welch concludes, the function of these women highlights how “women may exert themselves toward or away from Rome, and may be pulled toward or away from Rome” (2015, p. 163). When Livy (2002) says in his preface that the “study of history is the best medicine for a sick mind; for in history you have a record of the infinite variety of human experience plainly set out for all to see,” it is not that he has lost creative control of his narrative amongst the presence of conflicting accounts and sources; rather, Livy wants his readers to use these women to reconsider their own multiple stances, which brings an added nuance to his moralistic discourse.
References


PRIVILEGE & VOTING FOR PREDATORS: EXPLORING THE RELATIONSHIP BETWEEN PRIVILEGE AND VOTING FOR SEX OFFENDERS

REILLY SIMMONS, BUTLER UNIVERSITY
MENTOR: GREGORY SHUFELDT

Abstract

This paper was inspired by my confusion that women voted for sexual predators in several elections in 2016, despite the fact that women are usually the victims of sexual misconduct. My research question was, Does privilege affect political party crossover when a candidate is accused of sexual misconduct? I hypothesized that a woman’s likelihood of crossing political party lines when a candidate is accused of sexual misconduct decreases as her privilege increases, and that the opposite would therefore also be true: as a woman’s privilege decreases, her likelihood of crossing political party lines increases when a candidate is accused of sexual misconduct.

I used a split-sample survey experiment, then ran four bivariate tests. The first bivariate test measured the impact of partisanship on political party crossover when the candidate is accused of sexual misconduct. The second bivariate test tested the impact of privilege, controlling for partisanship. The third bivariate test was done differently for men and women. The last bivariate test measured whether privilege—controlling for partisanship—had an impact on the dependent variable separately among men and women.

I found my hypothesis to be wrong. A woman’s privilege does not significantly influence whether she is willing to support a candidate accused of sexual misconduct. Partisanship is the pivotal force in determining a woman’s voting habits when the candidate has been accused of sexual misconduct. I also found that a male’s increasing privilege increases his likelihood of crossing political party lines when a candidate has been accused of sexual misconduct.

After witnessing several high-profile elections, such as the 2016 U.S. presidential election, and the Alabama senate election, in which candidates accused of sexual misconduct either won or came incredibly close to winning the election, one should question why. Donald Trump ultimately became president of the United
States after more than 15 women came forward to say he sexually assaulted them (Pearson et al., 2020). In Alabama, Roy Moore lost his U.S. Senate race but still received 48.4% of the vote (Bloch et al., 2017). He was undeterred after his 2017 loss and ran again in 2020. These are two examples of higher-profile cases. It should not be overlooked or understated that candidates and government officials at any level have been accused of sexual misconduct.

One might assume that women, who are most often the victims of sexual misconduct, would change political parties to prevent the reward of a political office to a sex offender, yet that is not the case, as voters elected Donald Trump and Roy Moore only narrowly lost. This paper asks why women specifically are still willing to vote for candidates accused of sexual misconduct. After evaluating the literature on partisanship and elections with sexual misconduct accusations, previous studies have yet to examine whether the several levels of privilege a woman possesses might explain the results of such elections. This paper seeks to answer the research question of whether privilege affects political party crossover when a candidate is accused of sexual misconduct.

**Literature Review**

Certain aspects of a person’s identity are powerful predictors of their voting patterns. Many factors predict why people vote the way they do. Previous literature shows how a person’s race, sexual orientation, and socioeconomic status are all indicators of how that person may vote (Black & Stone, 2005; Crenshaw, 1997; Dunlap et al., 2007). A person’s privilege, and how privilege is defined, can determine voting habits or patterns (McIntosh, 2007). Partisanship is another powerful predictor of a person’s vote. This paper explores the connection between a woman’s privilege and her likelihood to cross over political parties when a candidate is involved in a sexual-misconduct scandal. This connection is explored because of the gap in the literature regarding voting for candidates accused of sexual misconduct and the types of people voting for these candidates. Here, the focus is specifically on women because sexual misconduct affects women at a greater rate than it does men (World Health Organization, 2002). This study questions whether privilege affects political party crossover in the instance of a woman’s party candidate being accused of sexual misconduct. The hypothesis is that with an increasing amount of privilege, a woman is less likely to cross political party lines when a candidate is accused of sexual misconduct.

Partisanship plays a powerful determining role in voting habits (Bafumi & Shapiro, 2009; Campbell et al., 1960; McCarty et al., 2016; Miller, 1991).
Partisanship has returned in a form that is both more ideological and more issue-based along liberal-conservative lines than it has been in more than 30 years. Voters are more influenced by partisan factors because these factors are more long-term compared to election-specific factors, which are seen as short-term. Voters have a certain attachment to their parties, and these loyalties to parties are what create the basic division in voters. The strength and direction of party identification are facts of central importance in accounting for attitude and behavior. Most Americans hold this sense of attachment to one party or another (Campbell et al., 1960). This attachment to one’s party has held constant over different time periods, regions, genders, and races. Miller (1991) explains that year in and year out, women have been no more likely than men to cast party votes or to defect and cross party lines to vote for a president. Beginning in the mid-1970s, politics became much more divisive, with more Democrats staking out consistently liberal positions and more Republicans supporting exclusively conservative ones. This trend made it increasingly difficult for moderates to win their parties’ contests (McCarty et al., 2016).

Additionally, society tends to view various types of scandals differently. Overall, people are more forgiving of sexual-misconduct scandals than of financial scandals, such as tax-evasion scandals, with the passage of time (Doherty et al., 2011), perhaps because the former depress evaluations not only of the politician’s personal attributes but also of his or her professional judgment. Another explanation for this discrepancy may be that although the incidents behind moral scandals could imaginably happen “in the heat of the moment,” financial scandals usually involve forethought and/or continued criminal behavior. Involvement in a financial scandal may thus be seen as a stronger signal of a politician’s fundamental character than may involvement in a sex scandal. For a political figure, reputation is a decisive asset. Inappropriate behavior, such as involvement in a scandal, may leave the candidate vulnerable to commentary from other political candidates, yet candidates involved in scandals have still been elected to office. One factor that may be particularly influential in determining the extent to which past involvement in a scandal can be used against a candidate is whether it occurred recently or long ago. In the research, however, the distinction of “moral” scandal is not explicitly made clear. Berinsky and colleagues (2011) found that White voters penalize Black candidates more than White candidates when there are instances of sexual misconduct. If the scandal cue plays out in a racialized way, White voters who hold negative attitudes toward Blacks should respond more strongly to that scandal. In other words, a racialized cue should work more strongly among people who are already receptive to it. In their experiment, Berinsky et al. found that Barack Obama
suffered a higher penalty than John Edwards in overall evaluations. A negative story involving rumors of a sexual-infidelity scandal hurt Obama more than it hurt Edwards both directly and immediately, on his overall favorability rating, as well as indirectly and potentially, through perceptions of his liberal ideology. The results are not limited to a particular partisan or ideological group. In these ways, Black candidates do seem to suffer a racial disadvantage with White voters.

There are different levels of privilege and ways to measure the amount of privilege someone has. One way of measuring privilege uses a person’s socioeconomic status (Dunlap et al., 2007). Economic privilege can be defined as having been reared in socioeconomic conditions that are financially stable and secure. In their article on social class bias and implications for training and practice, Liu et al. (2007) focus on White middle-class privilege. They forgo all other possible intersections related to race, making this a limited argument. In society, it is seen as normal to want to move upward with regard to socioeconomic status. When a person does not express desire for upward social mobility or cannot move up, this is viewed as deviant. Middle-class members frequently expect to get what they want because they are favored in economic institutions. Individuals from lower socioeconomic statuses believe that they are not as likely to have support from economic institutions and that they are more likely to face discrimination. The middle class is favored because it is the largest social class in society. The systems are set up to benefit middle-class members at a greater rate than lower-class members. McIntosh (2007) explains how socioeconomic privilege embodies itself through housing and the ability to move if needed. McIntosh then goes further to show that socioeconomic privilege includes the ability to choose where to receive an education. Other aspects include affording access to legal and medical help. Because the vast majority of societies are capitalistic, this allows the rich to exploit the poor underclass.

A second way of measuring privilege is via race (Crenshaw, 1997; Ghitza & Gelman, 2013; Jackson, 1999; Lucal, 1996). Crenshaw (1997) discusses how it is uncommon for White people to plainly mention their whiteness in political and academic discourse. Whiteness categorizes people into social areas that are easily distinguishable. A majority of White people view themselves as normal nonracists and are therefore often silent in times of struggle for people of color, in order to protect their own racial privilege. This silence is a type of language that many politicians take advantage of. By protecting their White privilege, they reinforce the idea of White superiority. This causes distrust and resentment between White people and people of color, and the underlying resentment builds into racial
tensions that drive political preferences. Ghitza and Gelman (2013) argue that the usual political discourse is that different demographic groups vote as homogenous blocs when, in fact, they show notable heterogeneity. The likelihood of one of the richest African Americans voting for a Democratic candidate was higher than that of one of the poorest African Americans voting for a Democratic candidate in the 2004 and 2008 elections. The turnout swing of voters is largely driven by African Americans and other young minorities. Younger White voters did not increase their turnout. Ghitza and Gelman’s study is limited because it focused on the elections of only 2004 and 2008 and provided only inferences about a small quantity of demographics.

Jackson (1999) explains that when White people deny the existence of White privilege, they must be conscious that it exists. White people are systemically advantaged in society because our society was built around White people. White people do not have to argue their identities like people of color might, and they are in a position of comfort, knowing that they are usually the majority in social and political situations. These social and political situations lead to White people being in a position of power over people of color. White participants in Jackson’s study felt empowered to keep a feeling of superiority. Lucal (1996) describes how race is thought to specifically apply to people of color, which leads White people to see it as something that does not affect them. Racial inequality discourse is spoken in ways that do not affect White society. Thus, White society can look at racial discrimination with disengagement. Whites have opportunities for not realizing how race works and not remembering how it works, whereas people of color are not afforded that luxury. White privilege gives White people the option of hearing or not hearing people of color. They can choose when they do and do not listen to different voices.

A third way of determining privilege is the sexual orientation of a person (Black & Stone, 2005; Blumenfeld, 1992; Bohan 1996). Black and Stone (2005) find that sexual-orientation privilege is based on heterosexuality being viewed as the most common expression of sexual orientation in society. Any orientation that strays from this expression is looked down upon, deemed inferior and wrong. Heterosexual people have a sense of superiority and an exaggerated belief in their self-worth. They also may have misperceptions about the world around them. Heterosexual people may believe that because they are heterosexual, they deserve special power and entitlements. For example, because they are heterosexual, the perceived norm, they may feel that their values are the correct values, or the only values that should be accepted, simply because they represent this norm. This
distortion of reality is painful for people to accept because they demand an examination of the consequences and demand accountability. Being accountable for privilege means that the privileged are ready to relinquish benefits and entitlements to which they have become habituated and that they recognize their aspect in the plausible oppression of others. Blumenfeld (1992) finds that any sexual orientation other than heterosexuality is thought of as able to be changed or that a person chooses to deviate from the norm. Even though homosexuality has been present throughout history, society still views sexual orientation as something that can change and that needs to be changed. Those who stray from the existing normative behaviors of heterosexual people are sometimes thought of as defectors of their sex. There are preconceived notions of what someone’s sexual orientation needs to conform to or what someone’s sexual orientation should avoid. When heterosexual people feel uncomfortable in an environment, they lash out at the minority sexual orientation.

Bohan (1996) explains how society rewards heterosexual people with benefits such as socially supported marriage and dating, therefore granting heterosexual people security within their sexuality. Heterosexual people do not have to suffer from violence or discrimination from institutions of any kind. They are also guaranteed a level of self-acceptance, in that society favors and approves of their sexual orientation. The system favors heterosexuality and does not question the actions of heterosexual people, whereas any other sexual orientation is questioned and the actions of people who are not heterosexual are under constant scrutiny because they are not approved of in totality by society. The dominance of Christianity in Europe and the United States feeds the narrative of heterosexuality being the only correct and natural sexual orientation; Christianity deemed homosexuality as a sin against God and an offense against the state.

The literature provides information about how partisanship heavily influences the way people vote, how the context of privilege affects the way people vote, and how a candidate’s involvement in a scandal and the race of that candidate affects how people vote. This paper focuses on the relationship between a woman’s privilege and her likelihood of crossing over political party lines when a candidate is accused of sexual misconduct, because this is one gap found in the literature. The more privilege a woman has, the less likely she is to cross political party lines when a candidate is accused of sexual misconduct.

This gap is important to explore because understanding the conceptions of privilege is important for voters and for scholars. Political party crossover when a candidate is accused of sexual misconduct is important for policy makers to
understand because the voting patterns of certain people can be revealed. Policy makers can take advantage of this information to determine what types of people are more likely to vote for a candidate accused of sexual misconduct, and what groups are absolutely unwilling to vote for such a candidate or would cross political party lines instead. This phenomenon is also important for scholars to understand in order to conduct more research on the types of people who are more likely to cross political party lines in a time of such intense party polarization. Looking at how different women view sexual misconduct when it comes to potential political candidates can become a new predictor of how certain women vote.

White women who are heterosexual and of middle-class and higher socioeconomic statuses are more privileged than those who are not White, heterosexual, and of middle-class socioeconomic status. There are varying degrees of privilege, however; a woman can fit into one of the three categories and still be more privileged than a woman who does not fit into any of the categories. Besides privilege, partisan loyalty also plays a part in the crossover vote of women. Partisanship is the biggest determinant of voting habits. The prediction for this study is that with a woman’s increasing privilege, her likelihood to cross political party lines will decrease when a political candidate is accused of sexual misconduct.

Research Design

The goal of this study is to determine the relationship between a woman’s privilege and when or if she crosses political party lines when a candidate is accused of sexual misconduct. The focus is on women specifically because sexual misconduct is deemed a women’s issue (World Health Organization, 2002). Previous research and historical trends have indicated that a person’s voting behaviors can be predicted by several factors. First, a person’s partisanship is one of the most powerful predictors of voting habits (Bafumi & Shapiro, 2009; Berinsky et al., 2011; Campbell et al., 1960). People have strong allegiance and loyalty to the political party they identify with. Second is their privilege—defined in this study through race, socioeconomic status, and sexual orientation (Black & Stone, 2005; Blumenfeld, 1992; Bohan, 1996; Crenshaw, 1997; Dunlap et al., 2007; Jackson, 1999; Liu et al., 2007; Lucal, 1996). The aim of this study is to determine if there is a relationship between a woman’s privilege and her likelihood of crossing political party lines when the candidate from her political party is accused of sexual misconduct.

This study takes the existing research a step further by focusing on what types of women are more likely to cross political party lines when a candidate is
accused of sexual misconduct. It goes beyond the scope of regular voting habits and patterns by specifically including in the study a candidate accused of sexual misconduct. The independent variable is the level of privilege a woman has, and the dependent variable is a woman’s likelihood of crossing political party lines when a candidate is accused of sexual misconduct. The hypothesis is that as a woman’s privilege increases, her likelihood of crossing political party lines when a candidate is accused of sexual misconduct decreases.

More-privileged women will prioritize winning at the expense of supporting a candidate accused of sexual misconduct. Privilege allows a woman to overlook the issue of sexual misconduct and to prioritize winning an election. The less privilege a woman has, the more likely she will prioritize punishing a candidate accused of sexual misconduct by not voting for that candidate. Privileged women are more likely to see their family members or friends in candidates accused of sexual misconduct. A large majority of candidates running for office are wealthy White straight males. This fits the mold for a man who is present within a wealthy White straight woman’s life. That is, the more privileged a woman becomes, the less likely she may become to believe the accusations of sexual misconduct against a candidate because she can see a friend or relative in the accused candidate.

Increasing privilege → Decreasing likelihood of crossing political party lines when candidate is accused of sexual misconduct

In contrast, women with less privilege have an increased likelihood of crossing political party lines when a candidate is accused of sexual misconduct. A woman with less privilege can more easily relate to a candidate’s accuser. She may have more exposure to incidents of sexual harassment, assault, and violence. She may have more friends or family members who have been victims of sexual misconduct and may therefore be more willing to punish a candidate accused of sexual misconduct.

Decreasing privilege → Increasing likelihood of crossing political party lines when candidate is accused of sexual misconduct

To test these hypotheses, an online survey was administered through the online survey platform Qualtrics. The 61-question survey included questions designed to tap into political attitudes and behaviors, with each question being provided by a different student in an undergraduate research methods course (see Appendix A). The sample was recruited from Amazon's Mechanical Turk crowdsourcing program. Mechanical Turk, in spite of being a more recent tool for recruiting survey respondents, is inexpensive and is documented to produce reliable
data (Berinsky et al., 2012; Buhrmester et al., 2011; Mason & Suri, 2012; Levay et al., 2016). Although the convenience sample limits the ability to draw conclusions regarding the full American public, it provides sufficient leverage to test the hypothesis. The survey, available for one week during the spring of 2018, recruited respondents by paying them $0.50 upon completion of the survey. The survey had a total population of 751 respondents. The average time of completion for respondents was 13 minutes and 42 seconds. This translated to an effective average hourly rate of $2.24.

The survey includes an original survey experiment in which respondents are divided into a split sample of a fictitious Democratic primary race and a fictitious Republican primary race. Respondents were randomly assigned to one of the two hypothetical scenarios in which a candidate is accused of sexual misconduct. The wording for the fictitious Democratic-race scenario is “Imagine there is a contested Democratic primary election in your state for Congress. Quinn Johnson, one of the candidates running for this office, has been accused of sexual misconduct in their history. This candidate is the best chance for the Democratic party to win the general election against their Republican opponent. If this candidate wins, the Democratic party is guaranteed the majority in Congress. How likely are you to vote for Quinn Johnson?” The respondent could choose from four options: very likely, likely, unlikely, and very unlikely. For the fictitious Republican-race scenario, the question is “Imagine there is a contested Republican primary election in your state for Congress. Quinn Johnson, one of the candidates running for this office, has been accused of sexual misconduct in their history. This candidate is the best chance for the Republican party to win the election against their Democratic opponent. If this candidate wins, the Republican party is guaranteed the majority in Congress. How likely are you to vote for Quinn Johnson?” The respondents had the same options of choosing from very likely, likely, unlikely, and very unlikely.

This question serves to determine who crosses political party lines when a candidate is accused of sexual misconduct. The wording is the same for each, except one race is Democratic and the other is Republican. For the data analysis, only the people who self-identified as Republicans receiving the Republican experiment and the Democrats receiving the Democratic experiment were examined. This limits the data because people who identify as political independents, people who identify as Democrats and received the Republican experiment, and people who identify as Republicans and received the Democratic experiment were removed from the analysis. The name of the fictitious candidate is the same for both races and is purposefully relatively gender neutral, to prevent
explicit gender bias. The respondents have a choice to select if they are very likely, likely, unlikely, or very unlikely to vote for the fictitious candidate. The responses were codified as who voted for the candidate—those being the respondents who selected very likely and likely—and who did not vote for the candidate—those being the respondents who selected very unlikely and unlikely. Those who answered unlikely or very unlikely were considered as crossing the political party line and voting against the candidate accused of sexual misconduct. The people who chose likely or very likely were considered as unwilling to cross over, prioritizing winning by voting for the candidate accused of sexual misconduct.

Beyond assessing whether partisans are willing to cross over, this study also examines the extent of crossing over based on the degree of privilege based on a woman’s sexual orientation, socioeconomic status, and race. The survey questions that measure these demographic characteristics ask about participants’ gender identification, race/ethnicity, sexual orientation, and income. Because in this study, privilege is defined by these categories, these questions help assess the level of privilege of each respondent. The question that determined gender asked, “What is your gender?” with the options male, female, and prefer to self-describe. Because only one respondent chose the option prefer to self-describe, the current study cannot draw conclusions about this group of people, and this respondent was not included in the data analysis. Respondents included 371 males and 378 females.

The survey item that determined race was “Please specify your race or ethnicity (check all that apply).” The options given to respondents were White, Hispanic or Latino, Black or African American, Asian/Pacific Islander, Native American or American Indian, and Other (Please Specify). Race was codified as White or nonwhite. Six hundred twenty (620) respondents identified as White, and 131 identified as nonwhite.

The survey item to determine sexual orientation asked, “Do you consider yourself to be:” with the options heterosexual or straight, gay, lesbian, bisexual, or not listed above (please specify). This group was codified as straight or non-straight. Six hundred seventy-five (675) respondents identified as straight, and 75 identified as non-straight.

The survey item to gauge level of income was phrased, “Information about income is very important to understand how people are doing financially these days. What is your total household income?” The options were less than $10,000, $10,000 to $19,000, $20,000 to $29,000, $30,000 to 39,000, $40,000 to $49,000, $50,000 to $59,000, $60,000 to $69,000, $70,000 to $79,000, $80,000 to $89,000, $90,000 to $99,000, and $100,000 or more. Income was codified as less than
$50,000 and more than $50,000. Three hundred thirty-four (334) respondents had incomes less than $50,000, and 416 had incomes greater than $50,000.

The independent variable is the degree of privilege a woman enjoys. Privilege within the race category is defined as being White. Privilege within the gender category would be defined as being male, but this study focuses specifically on women and their likelihood to cross political party lines when a candidate is accused of sexual misconduct. Privilege within the category of sexual orientation is defined as being heterosexual. Privilege within the income category is defined as having a household income above $50,000, indicating a participant has reached middle-class or higher standing.

To compare privileged women to a group, data for privileged men was included in order to understand if there is a difference in how privilege affects men and women. The general hypothesis was that with increasing privilege, a person’s likelihood of crossing political party lines when a candidate is accused of sexual misconduct will decrease.

In this sense, partisanship was controlled for because it is such a strong predictor. A Democrat who received the Republican-primary experiment might be more willing to not vote for the candidate accused of sexual misconduct because crossing over would involve voting within the political party the participant identified with. The same logic applies to Republicans who received the Democratic primary experiment. Partisanship was originally measured in the survey with the question “Generally speaking, do you think of yourself as a/an …?” The options the respondent could choose from were Strong Democrat, Not so strong Democrat, Independent, Not so strong Republican, and Strong Republican. For the current study, partisanship was dichotomized as either Democrat or Republican. Independents were not included because their behavior was not of interest for this study. Respondents included 198 Republicans and 312 Democrats. By taking partisanship into account, this study is able to examine the relationship between partisanship and privilege and the extent to which that relationship varies by gender. To assess the how privilege affects voting amid scandals, this study relies on a series of regression tests discussed in the following section.

Results

The first test measured the impact of partisanship on political party crossover when the candidate is accused of sexual misconduct. This test did not account for privilege or gender. Being a Democrat makes someone more unwilling
to support a candidate accused of sexual misconduct compared to being a Republican. This test had a $p$ value of .004 and therefore is statistically significant. The coefficient for this value is 0.184, which means that Democrats are more likely than Republicans to punish candidates accused of sexual misconduct (Table 1). Thirty-eight percent (38%) of Republicans punish the candidate accused of sexual assault, meaning that most Republicans prioritize supporting the candidate with the best chance to win. Fifty-six percent (56%) of Democrats are willing to punish the candidate accused of sexual assault.

Table 1. Coefficients of Privilege, Party, and Crossing Over

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Full sample</td>
<td>Full sample</td>
<td>Men</td>
<td>Women</td>
</tr>
<tr>
<td>Democrat</td>
<td>.184***</td>
<td>.200***</td>
<td>.030</td>
<td>.335***</td>
</tr>
<tr>
<td></td>
<td>(0.063)</td>
<td>(0.064)</td>
<td>(0.091)</td>
<td>(0.087)</td>
</tr>
<tr>
<td>Privilege</td>
<td>.056</td>
<td>.123*</td>
<td>.135</td>
<td>.331*</td>
</tr>
<tr>
<td></td>
<td>(0.046)</td>
<td>(0.068)</td>
<td>(0.180)</td>
<td>(0.179)</td>
</tr>
<tr>
<td>Constant</td>
<td>.379***</td>
<td>.237*</td>
<td>.439***</td>
<td>.304***</td>
</tr>
<tr>
<td></td>
<td>(0.049)</td>
<td>(0.066)</td>
<td>(0.070)</td>
<td>(0.070)</td>
</tr>
<tr>
<td></td>
<td>254</td>
<td>254</td>
<td>121</td>
<td>132</td>
</tr>
</tbody>
</table>

***$p < .01$. **$p < .05$. *$p < .10$.

The next regression tested the impact of privilege, controlling for partisanship. In this test, being a Democrat still has an impact. When controlling for partisanship, privilege does not have an independent impact on the likelihood of support for a candidate accused of sexual misconduct. The $p$ value for being a Democrat and not supporting the candidate accused of sexual misconduct was .002, which is statistically significant. The coefficient for this value is 0.20, which means that Democrats are more likely to not support a candidate accused of sexual misconduct than are Republicans. Controlling for privilege, 24% of Republicans punish the candidate compared to 44% of Democrats. The $p$ value for privilege, a four-point scale that captures race, gender, and income, was .228 and was not statistically significant.

The third test examined differences between men and women. Being a Democrat has no impact among men; 44% of men, regardless of partisanship, continue to vote for a candidate accused of sexual misconduct. Among women,
being a Democrat makes someone more unwilling to support a candidate accused of sexual misconduct compared to being a Republican. The $p$ value was less than .0001, so it can be said with 99.9% confidence that partisanship affects women’s unwillingness to support a candidate accused of sexual misconduct if the women are Democrats. Here the gender gap is particularly notable. While only 30% of Republican women punish this type of candidate, more than 64% of Democratic women refuse to support the candidate accused of sexual misconduct.

The final test assessed whether privilege, controlling for partisanship, has an impact on the dependent variable separately among men and women. For men, privilege, not partisanship, matters. For women, partisanship, not privilege, matters. The $p$ value for privilege among men is .071, which is statistically significant with a 93% confidence interval. The $p$ value for men being Democrats having an effect on willingness to vote for a candidate accused of sexual misconduct was .531 and therefore was not statistically significant. Roughly 14% of men without any additional privilege, regardless of partisanship, refuse to support a candidate accused of sexual misconduct. For each one-unit increase in privilege (in terms of race, sexual orientation, or income), men become 12% more likely to punish that candidate. For comparison, a nonwhite, non-straight male making less than $50,000 is quite likely to prioritize winning, whereas a straight White male earning more than $50,000 has a fifty-fifty chance of crossing over and refusing to support a copartisan accused of sexual misconduct.

Among women, it is partisanship, not privilege, that shapes views toward the candidates. The $p$ value for women with the highest level of privilege is .871, making it statistically insignificant; however, it is in the direction predicted of women with more privilege being less likely to cross political party lines when their candidate is accused of sexual misconduct. The $p$ value for women being Democrats having an effect on willingness to vote for a candidate accused of sexual misconduct was .0001, making it statistically significant with 99.99% confidence. The coefficient of 0.332 reveals a massive partisan gap. Controlling for privilege, 33% of Republican women punish a candidate accused of sexual misconduct. For Democratic women, it is a different story; more than 66% of Democratic women refuse to support a candidate accused of sexual misconduct.

**Discussion**

Ultimately, this study found that privilege has a positive impact on males crossing party lines. As privilege increases among males, they are more likely to cross over and thus punish a candidate accused of sexual misconduct, although in
the expected direction, privilege among females does not have a statistically significant impact. For females, partisanship has an impact, not privilege. Female Democrats are more likely to punish a candidate accused of sexual misconduct, regardless of privilege. Partisanship has no impact among men. Among women, being a Democrat makes one more unwilling to support a candidate accused of sexual misconduct, compared to being a Republican. For men, privilege makes one willing to cross over.

Partisanship playing a preeminent role for women is of particular note, given the increasing gender gap. Women who self-identify as Republican appear to be unmoved on this particular issue. Republican women are less likely to self-identify as feminist and less likely to view feminism as empowering (Barroso, 2020). Issues that are often conceptualized as women’s issues push women to be sorted into Democrat and Republican camps. The Republican party has a certain platform on abortion, reproductive health, gender roles, and tradition, so perhaps it would be unreasonable to expect sexual misconduct to make Republican women support their party any less. As Freeman (1986) believes, Republican women think the best thing they can do for women is to elect Republicans. Another way to think of this is that women who self-identify as Republicans are already choosing a stance on sexual misconduct. There is an overwhelming partisan split among women. Democratic women are twice as likely as Republican women to cross over. Before taking privilege into account, 64% of Democratic women, compared to 30% of Republican women, were willing to punish a candidate accused of sexual misconduct.

The findings for males are particularly interesting in that they are opposite of what one might expect. The implications of a man’s privilege affecting his voting habits versus his partisanship perhaps requires further research into his marital status, or testing for specific education levels, which could be factors at play in this occurrence. What specifically about a man’s privilege allows him to forgo his partisanship? Does pressure from peers, employers, relatives, spouses, and friends win out in an attempt to promote his own reputation or clout? Additionally, a man with more privilege may have greater access and time to learn about sexual-misconduct scandals and form an opinion than might someone who is living from paycheck to paycheck, focusing on putting food on the table, and simply surviving. Perhaps this issue trumps partisanship for men, or perhaps men—as those most likely to engage in sexual assault (World Health Organization, 2002)—do not necessarily view sexual misconduct as a partisan issue. Men, having a more stable position in society, are afforded more freedom to decide, ponder, and debate this
issue. Men are able to discard other males who jeopardize their positions, or they are also able to forgive and to welcome predators back into the fold. When taking privilege into account, there is a 50% chance that men will punish a fellow male copartisan accused of sexual misconduct.

In terms of future research, the first step would be to replicate this study to increase the external validity of the findings. In particular, a larger sample size would be helpful in providing a broader demographic pool. The sample size for this study had a majority of White and straight respondents. A more diverse pool would provide the opportunity to examine how privilege affects voting behavior among people of color, people of sexual-orientation minorities, transgender people, and people of different socioeconomic classes.

A larger sample size could also compensate for the design of the split sample. A larger universe would allow all people who identify as Democrats to be exposed to the fictitious Democratic-race scenario and the people who identify as Republicans, the fictitious Republican-race scenario. This would increase the number of respondents and allow for drawing more-accurate conclusions regarding political party crossover, because this study did not examine Democrats who received the fictitious Republican-race scenario or Republicans who received the fictitious Democratic-race scenario.

Revising the experiment to incorporate political independents also would provide an opportunity to examine how those with less partisan loyalty respond to allegations of sexual misconduct. Political independents were not included in this study because the focus was on political party crossover and in today’s politics, a candidate is typically a Republican or Democrat, but it would be interesting to see how this group of people vote specifically in the instance of a candidate being accused of sexual misconduct. This could be a possible research topic to be pursued in future.

Additionally, this experiment was within a primary election. A similar experiment should be conducted within a general-election environment to see if attitudes are different than in a primary election. Voters often act differently in primary versus general elections; it could be expected that partisanship would play an even stronger role in a general election than a primary election.

This study also raises the question of when partisanship becomes obsolete. When is it too powerful? The most recent 2020 presidential election unfortunately brought forth a scenario not covered in my study: candidates of both political parties had been accused of sexual misconduct. In this instance, research should examine
how voters respond, including what types of groups or people are more willing to vote third party or not vote at all because of the candidates’ sexual misconduct. It could also be of interest to study the specific group of people who refuse to vote for both major-party candidates accused of sexual misconduct and who instead choose to vote third party or not at all. As highlighted in this study, when it comes to the ballot box, #MeToo is a gendered, partisan issue.
References


Appendix A

Select Survey Questions

1. What is your gender?
   a. Male
   b. Female
   c. Prefer to self-describe: ______

2. Please specify your race or ethnicity (check all that apply).
   a. White
   b. Hispanic or Latino
   c. Black or African American
   d. Asian/Pacific Islander
   e. Native American or American Indian
   f. Other (Please Specify):

3. Do you consider yourself to be:
   a. Heterosexual or straight
   b. Gay
   c. Lesbian
   d. Bisexual
   e. Not listed above (Please Specify): ______

4. Information about income is very important to understand how people are doing financially these days. What is your total household income? Drop down menu: Below options
   a. Less than $10,000
   b. $10,000 to $19,999
   c. $20,000 to $29,999
   d. $30,000 to $39,999
   e. $40,000 to $49,999
   f. $50,000 to $59,999
   g. $60,000 to $69,999
   h. $70,000 to $79,999
   i. $80,000 to $89,999
j. $90,000 to $99,999
k. $100,000 or more

5. Generally speaking, do you think of yourself as a/an …?
   a. Strong Democrat
   b. Not so strong Democrat
   c. Independent
   d. Not so strong Republican
   e. Strong Republican
   f. Other: (Please specify)

6. (Split Sample) Fictitious Democrat Race Format: Imagine there is a contested Democratic primary election in your state for Congress. Quinn Johnson, one of the candidates running for this office, has been accused of sexual misconduct in their history. This candidate is the best chance for the Democratic party to win the general election against their Republican opponent. If this candidate wins, the Democratic party is guaranteed the majority in Congress. How likely are you to vote for Quinn Johnson?
   a. Very likely
   b. Likely
   c. Unlikely
   d. Very unlikely

7. (Split Sample) Fictitious Republican Race Format: Imagine there is a contested Republican primary election in your state for Congress. Quinn Johnson, one of the candidates running for this office, has been accused of sexual misconduct in their history. This candidate is the best chance for the Republican party to win the election against their Democratic opponent. If this candidate wins, the Republican party is guaranteed the majority in Congress. How likely are you to vote for Quinn Johnson?
   a. Very likely
   b. Likely
   c. Unlikely
   d. Very unlikely
ROCKIN’ THE GRE: THE EFFECTS OF PREFERRED, NON-PREFERRED, AND CLASSICAL MUSIC ON COLLEGE STUDENTS’ COGNITIVE TEST PERFORMANCE

HANNAH B. BOLANDER & SEAN CALLAHAN, BUTLER UNIVERSITY
MENTOR: TARA LINEWEAVER

Abstract

This paper examines the effects of preferred popular music, non-preferred popular music, and classical music on college students’ reading comprehension. The reported study addressed shortcomings in the literature by presenting music to participants before rather than during testing and also by attempting to better equalize levels of arousal and mood that might be differentially affected by the various music types. After listening to each of the three music playlists, 33 undergraduate students at Butler University rated their enjoyment of the music, completed a music rating scale, self-reported their mood and arousal, and answered six GRE Reading Comprehension questions. Results showed that participants felt significantly more invigorated and less depressed after their preferred music compared to either non-preferred popular or classical music. Although participants enjoyed their preferred music significantly more than the other two music types, our hypothesis that preferred popular music would exert a greater positive influence on reading comprehension than would non-preferred popular music (which would, in turn, result in better reading comprehension than listening to classical music) was not supported. Reading comprehension scores across each of the three conditions were almost identical. Thus, the current study provides no evidence that various types of music have differentiated effects on reading comprehension. Instead, results suggest that personal song selections can be utilized to improve individuals’ energy levels and mood to a greater extent than either non-preferred popular music or upbeat classical music.

Researchers have been fascinated with the effects of music on cognitive abilities for years. The “Mozart effect” became popularized with the notion that music, specifically that of Mozart, enhances people’s general cognitive abilities (Cabanac et al., 2013; Perham & Currie, 2014; Rauscher et al., 1993). In support of the Mozart effect, Cabanac et al. (2013) found that studying music leads to short-term improvements in academic performance. They examined the effects of music
on high-achieving third-, fourth-, and fifth-year students in Canada. All students were required to take a music course during their first two years of school, but during years 3–5, these courses became optional. Cabanac et al. (2013) found that students who took music courses in years 3–5 performed better in a variety of classes, such as history, science, and English, than did students who chose to focus on art or theater.

Although classical music has been found to have a general effect on cognitive abilities, not all classical music has the same effect. In one of two experiments, Schellenberg et al. (2007) looked at how the music of Albinoni (with a slow tempo and minor key) and Mozart (with a fast tempo and major key) affected cognition and mood. The researchers gave participants a measure of processing speed after listening to each type of music and found that scores were better after participants listened to Mozart than after they listened to Albinoni, indicating that classical music in a major key with a fast tempo has a greater potential to improve performance on cognitive tests. They also found that participants scored higher in depressive affect after listening to Albinoni’s music than at baseline, in contrast to the decrease seen after listening to Mozart’s music. Finally, these researchers found that arousal scores increased while participants listened to Mozart but decreased after they listened to Albinoni. These results indicate that different types of classical music differentially affect both cognition and mood.

Although both Cabanac et al. (2013) and Schellenberg et al. (2007) have found that classical music increased cognitive performance, additional studies have found that preferred music has an even greater effect on cognition than does classical music. Eskine et al. (2018) had participants at a predominantly African American university in Louisiana listen to hip-hop, classical music, and background noise before completing a creativity test. They found that participants were more creative after listening to hip-hop compared to when they listened to classical music or background noise. A second experiment by Schellenberg et al. (2007) also found that children who listened to familiar music, compared to classical music, were not only more creative but also more persistent in the amount of time they spent creating a drawing relative to baseline. This suggests that listening to preferred music before completing a task can lead to even better performance than listening to classical music.

Conversely, not all studies have found positive effects of preferred music on cognition. Perham and Currie (2014) had participants listen to three different types of music—disliked lyrical music, liked lyrical music (preferred music), nonlyrical music—and to no music. As they listened, participants read four
passages and completed six questions about each passage. Participants who listened to liked or disliked lyrical music while performing the reading comprehension task did significantly worse compared to those listening to nonlyrical or no music. This suggests that listening to lyrical music may disrupt reading comprehension, even when it is the participant’s preferred music.

Although the literature suggests that various types of music can affect both cognition and mood, the relationship between these two responses to music is unclear. Schellenberg et al. (2007) suggested that increased arousal but not decreased depression might correspond with improvements in processing speed. In contrast, other studies have found little relationship between changes in mood and changes in cognition (Eskine et al., 2018; Gültepe & Coskun, 2016). Participants in the study by Eskine and colleagues (2018) reported feeling more excited and experiencing a more positive mood after listening to hip-hop than after listening to either classical music or noise, but the correlation between mood and creativity was not significant. Similarly, Gültepe and Coskun (2016) had participants listen to positive, negative, and neutral music for nine minutes at a time. After each type of music, participants completed a six-item mood scale. Participants then took 15 minutes to brainstorm solutions to a problem. Although positive, negative, and neutral music had different effects on cognitive flexibility, differences in the participants’ self-reported mood did not correspond with increases in cognitive performance.

The current study addressed some of the shortcomings in the literature. First, this study looked at the effects of listening to preferred music before, rather than during, reading comprehension tests. We expected that listening to music before the task would have a greater positive effect because it would be less distracting than music during reading, as utilized in the Perham and Currie (2014) study. Second, this study compared preferred versus non-preferred popular music. No study to date had examined whether one’s preferred music has a greater effect on cognition and mood than does highly similar music. We compared these two types of music to classical music. Third, this study design controlled for potential differences in mood and arousal that could be associated with popular versus classical music. Comparing preferred popular music to non-preferred popular music, which should be similar in their effects on mood and arousal, allowed us to differentiate whether differences in emotional responses account for the findings of previous studies or whether personally preferred music has a unique effect on cognition.
We had two primary hypotheses. Our first hypothesis was that popular music would exert a greater positive influence on cognitive performance than did classical music but that participants would experience a greater benefit from songs that they specifically selected (preferred popular music) than from other similar songs (non-preferred popular, nonclassical music). Our second hypothesis was that playlists comprising personally selected popular music or other participants’ popular song selections would both generate similar levels of arousal and positive mood that would surpass the affective response associated with classical music.

**Method**

**Participants**

Participants consisted of undergraduate psychology students at Butler University. The sample included 33 students (87.9% female; 84.8% white) between the ages of 18 and 22 ($M = 19.61$, $SD = 1.14$). On average, students were in their sophomore to junior year in college ($M = 2.30$, $SD = 1.02$). All students received extra credit in a psychology course for their participation.

**Materials**

**Profile of Mood States (POMS)**

This self-report questionnaire from McNair et al. (1992) was used to assess mood and arousal. The full measure included six subscales. As in the work of Schellenberg et al. (2007), participants completed the Vigor-Activity subscale, a measure of arousal, as well as the Depression-Dejection subscale, a measure of negative mood. There were 8 items on the Vigor-Activity subscale (e.g., energetic, cheerful) and 15 items on the Depression-Dejection subscale (e.g., sad, hopeless). Participants rated each adjective on a Likert scale of 0–4 ($0 = $ not at all, $4 = $ extremely) based on how they felt at that moment. The items on each subscale were averaged to provide a mean vigor score and a mean depression score that represented participants’ moods after listening to each of three playlists. Possible scores thus ranged from 0 to 4 for each subscale, with higher scores reflecting greater vigor and greater negative affect.


**GRE Reading Comprehension Practice Tests**

Each practice test included two reading passages followed by three multiple-choice questions about each passage. Each passage comprised approximately 7–9 sentences and was approximately 190 words long. The first three GRE practice tests from the website Graduateshotline.com (2021) were used for the purposes of this study. Participants could score between 0 and 6 on each of the three reading comprehension tests.

**Music Rating Scale**

This scale was created for the purposes of this study. Students were asked to rate how much they liked each of the three playlists immediately after listening to it, rating on a Likert scale of 1–5 (1 = strongly dislike, 5 = strongly like). Possible ratings thus ranged from 1 to 5 for each playlist, with a higher score reflecting a greater liking of the playlist.

**Demographic Questionnaire.** This questionnaire gathered participant demographics such as age, gender, race, and year in school.

**Procedure**

Students compiled a list of their five favorite songs for a pretesting packet administered through the psychology department at the beginning of the semester. Upon completion of the list, each student received an email invitation to participate in this study. The email included an access code that allowed students to select one of several group-testing sessions to attend. On the day of the study, participants completed the informed consent, a demographic questionnaire, and a baseline POMS. Next, all participants listened to one of three music playlists through headphones for 10 minutes. After listening, participants first rated the playlist on the Music Rating Scale and then answered the items of the POMS again. Next, they had 10 minutes to complete GRE Reading Comprehension Practice Test 1. Participants then used the following 10 minutes to complete a maze that served as a filler task. After the maze, the procedure was repeated with a second playlist followed by GRE Practice Test 2, and then the third playlist and GRE Practice Test 3.

The three playlists included (1) the participant’s personal playlist of preferred music, (2) a random selection of songs from other participants’ preferred musical selections, and (3) classical music consisting of Mozart’s compositions.
The order in which participants experienced these three playlists during their testing session was counterbalanced.

Upon completion of the testing, participants received extra credit in one of their psychology courses to thank them for their time.

Results

**Differences in Reading Comprehension Following Preferred, Non-Preferred, and Classical Music**

To address the primary hypothesis, we compared reading comprehension scores following preferred, non-preferred, and classical music. We utilized a within-subjects analysis of variance with the number of items answered correctly on the GRE reading subtest as the dependent variable and the type of music (preferred, non-preferred, and classical) as the independent variable. The main effect of music type did not reach significance in this analysis, $F(2, 31) < 1.0$, $p = 0.990$, $\eta^2_p = 0.001$. Thus, we did not find any significant differences in performance on the GRE reading comprehension practice tests after participants listened to playlists of preferred ($M = 3.97$, $SD = 1.10$), non-preferred ($M = 3.94$, $SD = 1.20$), or classical ($M = 3.94$; $SD = 1.20$) music (Figure 1).

**Differences in Affect Following Preferred, Non-Preferred, and Classical Music**

To examine our second hypothesis, we compared participants’ enjoyment levels and affect following preferred, non-preferred, and classical music. We conducted three within-subjects analyses of variance with music rating scale scores, vigor scores, and depression scores as the dependent variables and the type of music (preferred, non-preferred, and classical) as the independent variable.

When analyzing music rating scale scores, the main effect of type of music reached significance, $F(2, 31) = 84.29$, $p < .001$, $\eta^2_p = 0.85$. Follow-up within-subjects contrast analyses indicated that when rating the playlists, participants enjoyed listening to their preferred popular music ($M = 5.00$, $SD = 0.00$) significantly more than the non-preferred popular ($M = 3.36$, $SD = 0.96$) or classical ($M = 2.94$, $SD = 1.09$) music playlists, $p < .001$ for both. Participants’ ratings of the non-preferred and classical music did not significantly differ from each other, $p = .07$. 
Next, we looked at self-reported vigor levels after participants listened to each playlist (Figure 2). The main effect of type of music reached significance, $F(2, 31) = 29.54, p < .001, \eta^2_p = 0.66$. Participants described themselves as significantly more invigorated after listening to their preferred playlist ($M = 1.95, SD = 0.87$) than after listening to playlists of non-preferred ($M = 1.19, SD = 0.76$) or classical ($M = 1.00, SD = 0.67$) music, $p < .001$ for both. Participants’ ratings of their vigor after listening to non-preferred and classical music did not significantly differ from each other, $p = .20$.

Finally, we analyzed participants’ self-reported depression-dejection levels after they experienced each playlist (Figure 2). The main effect of type of music reached significance, $F(2, 31) = 4.19, p < .025, \eta^2_p = 0.21$. Participants described themselves as significantly less depressed after listening to their preferred music ($M = .09, SD = 0.27$) than after listening to non-preferred ($M = 0.16, SD = 0.26$) or classical ($M = 0.19, SD = 0.31$) music, $p < .05$ for both. Participants’ depression ratings after listening to non-preferred and classical music did not significantly differ from each other, $p = .29$.

### Correlations Between Participants’ Reading Comprehension Scores and Response to Music

Last, we analyzed the correlations between participants’ affect ratings after listening to each type of music and their scores on the corresponding GRE tests (Table 1). The only correlation to reach significance was the one between depression and GRE scores after listening to the playlist of non-preferred music. Interestingly, those who felt more depressed after listening to the non-preferred playlist performed better on the subsequent reading comprehension measure.

### Discussion

This study sought to address various shortcomings in the literature by comparing the effects of preferred popular, non-preferred popular, and classical music on reading comprehension. We had two primary hypotheses when conducting this experiment. Our first hypothesis was that popular music would exert a greater influence than classical music on cognitive performance but that participants would benefit more from specifically selected music (preferred popular music) than from other similar songs (non-preferred popular music). Our second hypothesis was that the playlists of preferred and non-preferred music, both comprising popular music, would generate similar high levels of arousal and low
levels of negative affect. At the same time, we expected that classical music would not influence affect as positively as the other two types of music.

Our hypothesis that popular music would exert a greater influence than classical music on participants’ cognitive performance was not supported by our results, as average reading comprehension scores were almost identical following each of the three playlists. This contrasts with the findings of Eskine et al. (2018) and of Schellenberg et al. (2007), who found that participants performed better on creativity measures after listening to popular music than after listening to classical music. To build on these previous studies, we also attempted to differentiate preferred popular music from non-preferred popular music by allowing participants to select their favorite songs. Even then, selections of preferred popular songs exerted no greater influence on reading comprehension than did other forms of popular music.

Considering the results of the current study within the context of the existing literature raises the possibility that popular music might differentially affect various cognitive processes. Previous studies have demonstrated music’s positive effect on divergent, or creative, thinking. Both Eskine et al. (2018) and Schellenberg et al. (2007) found that music enhanced the creativity and problem-solving abilities of participants. After listening to music, participants performed better on subsequent creativity tests and participated in creative activities, such as drawing, for longer periods of time. Similarly, Gültepe and Coskun (2016) demonstrated that various types of music differentially influence cognitive flexibility during a brainstorming task. Conversely, studies focused on analytical, convergent, thinking have not found positive benefits of popular music (Perham & Currie, 2014). As such, the results of the current study support those of Perham and Currie (2014), who similarly found that popular music did not improve reading comprehension. We originally hypothesized that the lack of positive effects of preferred music on cognition in Perham and Currie’s (2014) study was attributable to participants listening to music during reading comprehension tests. Unfortunately, playing music before rather than during the reading comprehension tests did not augment the effects of preferred music on cognition in the manner that we expected.

In the current study, although the three types of music did not have varied effects on reading comprehension, they did differentially influence both enjoyment and affect, providing support for our second hypothesis. Consistent with our expectations and past research (Eskine et al., 2018), students reported more enjoyment and experienced a greater improvement in mood after listening to popular music compared to classical music. Although we hypothesized that both
preferred and non-preferred popular music playlists would generate similar levels of arousal and positive mood, we instead found strong evidence supporting the effectiveness of preferred popular music over other popular songs toward improving emotional states. Participants enjoyed their specific selections significantly more than either non-preferred popular music or classical music. In addition, we found that preferred music led to significantly higher self-reported vigor levels and lower levels of depression. This exemplifies the importance of one’s personally selected music over similar popular music (such as a random selection of popular songs like that typically played on the radio) to positively influence mood. Although this study only compared preferred to non-preferred music within the popular music genre, this effect might extend to other genres as well. For example, preferred classical music might also have a greater effect on mood than non-preferred classical music, although further research would be necessary to confirm this.

When looking at the correlational results, we found little relationship between self-reported mood and subsequent performance on the associated reading comprehension test. This largely supports the results of similar studies finding that music’s effect on cognition is independent of its influence on affect (Eskine et al., 2018; Gültepe & Coskun, 2016). For example, the participants in Eskine and colleagues’ (2018) study reported a more positive mood after listening to hip-hop than after listening to either classical music or noise, but there was no correlation between affect and their creativity. In contrast, our study found a significant correlation between depressive affect and reading comprehension following non-preferred music listening. Interestingly, participants who reported more depressive affect after listening to their non-preferred music outperformed students who reported less depressive affect. This result raises the possibility of subtle relationships between mood and cognition following music listening, perhaps more apparent on convergent thinking tasks, such as ours, than on divergent thinking tasks, such as those of Eskine et al. (2018). Schellenberg et al. (2007) also documented relationships between post-music-listening mood and cognition, finding that participants who listened to fast-tempo, major-key classical music experienced both increased arousal and improvements in processing speed compared to when they listened to slow-tempo, minor-key pieces. This also suggests a direct relationship between music’s effects on mood and its effects on cognition after listening to non-preferred music. Although future research will be necessary to rectify these contradictory findings in the literature, the inconsistent relationships between mood and cognition do not diminish the positive effect of
music. These results indicate that when music affects mood and arousal, it does not always affect cognitive abilities as well.

Limitations and Future Directions

Although our results partially supported our hypotheses, limitations of this study should be considered when generalizing these results to other contexts. First, this study included only 33 undergraduate students. Although this is a relatively small sample size, the within-subjects research design increased the power to detect significant differences across conditions. Although this study could have included more participants, the small effect size associated with the differences between reading comprehension scores following listening to the three types of music ($\eta_p^2 = 0.001$) suggests that including additional participants would not have likely changed our results.

Participants in this study composed a homogenous sample. They included mostly white, affluent undergraduate psychology students at a private university. All of these students were enrolled in classes that require the regular use of reading comprehension skills. This may have limited the potential power of music to improve these well-honed cognitive abilities. A broader participant pool could potentially have resulted in a larger effect of music. Perhaps older adults who don’t frequently practice their reading comprehension skills would exhibit a greater improvement in cognition after listening to their preferred music. Future research should utilize a diverse population of participants representing a large range of ages. Comparing those who frequently employ their reading comprehension skills in the classroom to those whose jobs or daily activities require less-frequent utilization of this ability could help delineate individual differences in who is most likely to benefit from music interventions.

Another limitation of the current study is that participants listened to each playlist for only 10 minutes. This was designed to ensure that participants did not run out of music during testing sessions. Because of this short listening period, however, participants may not have experienced all of the songs on the three playlists, particularly the playlist of preferred songs. Had participants listened to more of each type of music, greater differences may have emerged between the various music types in terms of their effects on both mood and cognition.

Because each participant listened to their own unique playlist of preferred songs, these playlists comprised mixes of fast-tempo and slow-tempo songs as well as songs in major and minor modes. In addition, the classical music that participants
experienced consisted of four songs in major mode and one in minor mode, of varying tempos. Thus, this study could not determine the extent to which the tempo and mood of the music might have influenced students’ enjoyment and affective response to it. Designers of future studies may want to more carefully control these aspects to determine the extent to which these factors play a role in our current results.

Last, this study lacked a no-music control. Although no differential effect existed among the three types of music on cognition, it is possible that all three had an equally positive effect. The current study could not evaluate this possibility. Future research should add a fourth experimental condition that involves participants completing a reading comprehension test after listening to no music, to determine if the three types of music included in this study had no effect on cognition or if all three music types exerted an equally positive influence on reading comprehension skills.

Conclusions and Implications

Although reading comprehension scores did not improve after participants listened to preferred music, participants’ personally selected songs did have a significant and unique effect on enjoyment and mood. The benefit of personally selected songs therefore comes from the greater enjoyment, increased vigor, and diminished depression that participants experienced after listening to their favorite songs. The results of the current study provide no evidence that these three types of music have differentiated effects on cognition. In this sense, encouraging students to listen to particular types of music would not be a good intervention to help students maximize their cognitive abilities. In contrast, our results indicate that when individuals feel sad or would like to improve their moods, they might benefit from abandoning generic popular music, such as that played on the radio, in favor of personalized playlists full of their favorite songs.
References


Table 1. Correlations Between Vigor and Depression Following Preferred, Non-Preferred, and Classical Music and their Corresponding GRE Scores

<table>
<thead>
<tr>
<th></th>
<th>Preferred</th>
<th>Non-Preferred</th>
<th>Classical</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>POMS–Vigor</strong></td>
<td>$r = -0.147$</td>
<td>$r = 0.159$</td>
<td>$r = 0.175$</td>
</tr>
<tr>
<td><strong>POMS–Depression</strong></td>
<td>$r = 0.154$</td>
<td>$r = 0.352^*$</td>
<td>$r = -0.052$</td>
</tr>
</tbody>
</table>

* $p < 0.05$

Figure 1. Mean GRE Reading Comprehension Test Scores
Note. No significant differences emerged in participants’ performance on the GRE reading comprehension tests after listening to preferred, non-preferred, or classical music.
Figure 2. Mean Affect Scores After Listening to Each Playlist

Note. Participants described themselves as significantly more invigorated and less depressed after listening to their preferred music than after listening to either of the other playlists.
THE CAMPUS AS COMMONS: EDUCATIONAL ANCHOR INSTITUTIONS AND THEIR IMPLICATIONS FOR POLITICAL ECONOMY

CAMERON N. CONNER, WHITMAN COLLEGE
MENTOR: KEITH FARRINGTON

Abstract

In this paper, I analyze the growing call for institutions of higher education to better support the communities they are part of and so to act as anchor institutions. After establishing a precise definition of anchor institutions as they exist today, I draw on existing policy, reports, and academic papers to identify the common arguments used to advocate for anchor institutions. Dividing these into four main principles, I contend that the arguments for anchor institutions are fundamentally reliant on the theory of the commons and the underlying political economy it represents. By framing the pioneering work of anchor institutions as acts of commoning, I create a theoretical justification for the proliferation of anchor institutions; therefore, in connecting theory to praxis, this paper is an attempt to illuminate the evolutionary potential of the university as an anchor institution and to promote its capacity to act as a social asset aimed at achieving common abundance rather than at engineering scarcity.

Across the United States, a growing number of colleges and universities have begun to prioritize their relationships with the communities that surround them. At Portland State University, community engagement now counts toward the faculty tenure-review process.1 In 2016, Emory University committed to purchasing 75% of the food consumed on its campus from “local or sustainably grown sources” by 2025.2 Over the course of several years, the University of Minnesota has successfully integrated policies that prioritize local women- and minority-owned contractors for all university construction projects.3 Through these three programs alone, billions of dollars have been channeled into local economic and community development, and these initiatives are not unique. As of 2019, three-quarters of urban universities included commitments to public service within

1 Hodges and Dubb, Road Half Traveled, 45.
2 Emory University, 1.
3 Hodges and Dubb, Road Half Traveled, 123.
their mission statements. These factors represent a growing movement that seeks to harness the social, cultural, and economic power of higher education for the well-being of communities that exist beyond campus. As a result of this momentum, colleges and universities are increasingly expected to actively contribute to the communities in which they are located. In their 2019 report on the evolution of town-gown relationships, The Chronicle of Higher Education summarized the effect of this growing sentiment with a simple conclusion. “A college may be its own domain,” the introduction acknowledges, “but being part of a larger place increasingly means supporting it in numerous ways.”

Because of the potential impact that these educational institutions can have, colleges and universities have increasingly come to be thought of as one type of anchor institution. This term is used to describe institutions that are “grounded” in their community and so are generally unable to leave because of cultural, infrastructural, and economic investment in the location. In recent years, as these ties have been recognized, universities have grown to acknowledge that their institutional success is dependent on the health of their surroundings. Because of their relative size, spatial permanence, and corporate status, such educational anchor institutions also have the capacity to affect surrounding communities by generating and channeling significant resources. The notion of anchor institutions exemplifies the growing demand for universities to play roles in their communities beyond those which education and academia alone can provide.

As these town-gown partnerships have proliferated, a growing body of scholarship on the subject of educational anchor institutions has accumulated; however, much of the existing literature focuses on policy-based analyses of “best practices” by which these institutions can magnify their effects. Frequently left out of this conversation is an ideological analysis of where the pillars of this practice originated. Some national research institutes and nonprofit policy organizations have incorporated the notion of anchor institutions into their own theoretical visions.

---

4 Carlson and Bielmiller, “Campus as City,” 8.
5 Carlson and Bielmiller, 3.
6 For the sake of simplicity, I will use the term “university” through the rest of this paper to refer to all institutions of higher education.
7 In addition to colleges and universities, a number of other organizations have come to be seen as anchor institutions. These include hospitals, museums, community arts centers, faith-based institutions, and libraries.
8 Though research on education-based anchor institutions is often focused on urban universities that have large research components, all institutions of higher education—from research universities and community colleges to historically black colleges and universities and liberal arts institutions—are to be understood as anchor institutions in their own communities.
for a new political economy, but none have comprehensively unpacked the theoretical traditions that such town-gown partnerships are grounded in. The purpose of this paper is thus to bridge the gap between the policy proposals that have been put forth, on one hand, and the theoretical principles that justify their existence, on the other.

Illuminating this relationship is vital if educational institutions are to realize their full potential as anchor institutions. Underneath the argument that the university should consciously and strategically employ its resources for the welfare of surrounding communities exists a set of implicit principles and assumptions about the relationship among individuals, institutions, and the state within society. If anchor institutions are to be fully understood, let alone functionally realized, these underlying principles and their implications must be brought to light along with the theoretical genealogies underpinning them.

This paper is intended to situate the ever-increasing number of anchor institutions into a broader philosophical context; in essence, it is an attempt to theoretically justify why university-based anchor institutions should exist. I address this question by arguing that the contemporary justifications for anchor institutions stem from a notion of political economy based upon the theory of the commons. By interpreting the argument for anchor institutions through this commons-based framework, I seek to demonstrate that it is possible to illuminate the implicit assumptions and values that motivate these institutional practices and demonstrate why society truly needs both anchor institutions and the theoretical perspective they represent. The implications for this lens are substantial. If the growing popularity of anchor institutions can be interpreted as a manifestation of the commons, it means that key social institutions are beginning to stand at odds with the tenets of contemporary political economy. That is the larger conversation in which this paper is set.

The following pages are divided into four main sections. The first section is dedicated to developing a comprehensive understanding of the term anchor institution, specifically as it is applied to the university. The second analyzes the

---

9 To pursue this purpose, my argument is concerned with universities specifically, rather than with all manner of anchor institutions, for three reasons. First, the university has been the most extensively studied and documented type of anchor institution to this point. Second, because the academic nature of the university makes it prone to self-reflection, there is a greater understanding of its social mission and thus of its obligation to society more generally. Finally, as the engines of intellectual advancement, if universities are able to embrace their ability to cultivate community resiliency and so realize this new ethos of engagement, their impact has the potential to be the farthest reaching.
argument for why certain institutions bear an inherent obligation to better both society in general and—importantly—their surrounding communities in particular. The third section then introduces theoretical language of the commons as a productive framework through which to interpret these arguments for institutional engagement, and the paper concludes by exploring the broader implications this argument has for political economy as a whole.

“A New Paradigm”: Defining Anchor Institutions

The idea that institutions of higher education have an obligation to uphold the common good has prevailed since the advent of the earliest European universities in 1200 AD.10 This social-purpose credo has been manifested in many different ways throughout history. The primary goal of land-grant colleges, for example, was to “solidify the American economic infrastructure.”11 State-run cooperative extension programs, meanwhile, were begun in 1914 to have a more social impact by providing better “public and outreach services.”12 Likewise, “settlement houses” for underserved communities were originally staffed and supported predominantly by university students throughout the late nineteenth century and are exemplary of initiatives of community-driven work that likewise fulfilled the same social-purpose credo.13 These examples demonstrate that although the term anchor institution is relatively recent, the practices it represents are not.

Despite drawing on this long tradition, the contemporary manifestation of these principles under the label of anchor institution is unique. In the context of rising unemployment, inequality, and financial vulnerability during the 1960s, institutions of higher education emerged as a potential mechanism for cultivating economic stability and strength. Motivated by their public-service mission and the exigency of these deteriorating social circumstances, universities began evaluating how they could play larger roles in addressing local issues. To the extent that universities did seek to engage their host communities, however, most saw this primarily as “add-ons” far removed from their primary missions of education and research.14 The institutionalization of these practices was infrequent unless undertaken by strong leadership within the university.

10 Cuthill et al., “Universities and the Public Good.”
12 Hodges and Dubb, Road Half Traveled.
13 Hodges and Dubb, 4.
14 Taylor and Luter, Anchor Institutions, 2.
The term *anchor institution* as applied to American universities was first coined in 2001 by the Aspen Institute Roundtable on Comprehensive Community Initiatives.\(^\text{15}\) Responding to a need to name the growing trend in university (and hospital) engagement that was taking place across the country, the report generated by this assembly defined the phenomenon of anchor institutions as “central city institutions, ‘that have a significant infrastructure investment in a specific community and are therefore unlikely to move.’”\(^\text{16}\) The ability to name this new concept and identify its “best practices” catalyzed its popularity throughout the 2000s. This marked a turning point in the evolution of community-focused university engagement. In 2009, the Anchor Institution Task Force (AITF) was created to advise the U.S. Department of Housing and Urban Development (HUD) on how the federal government could better leverage anchor institutions as a means of alleviating “significant urban problems.”\(^\text{17}\) Simultaneously, an increasing number of universities intentionally began to assume the responsibilities expected of anchor institutions. On both the national and local levels, this concept soon grew to become “a new paradigm for understanding the role that place-based institutions could play in building successful communities and local economies.”\(^\text{18}\)

Under this new paradigm, anchor institutions were characterized by a relatively straightforward set of qualities. In their 2013 report intended to assess and synthesize the current state of knowledge on anchor institutions, the AITF concisely defined anchor institutions as “large, spatially immobile, mostly non-profit organizations that play an integral role in the local economy.”\(^\text{19}\) The potential influence these three qualities represent can be clearly seen in the case of American institutions of higher education. Cumulatively, the approximately 4,100 colleges and universities across the country employ more than four million people, serve 21 million students, possess $533 billion in endowments, and expend a similar amount in annual economic activity.\(^\text{20}\) The fact that these institutions individually have significant cultural, infrastructural, and economic investments in their particular locations means that these considerable economic resources are unlikely to disappear any time soon—as might happen, for example, with corporations of a similar scale. Size, spatial immobility, and corporate status are thus the defining

\(^{15}\) Fullbright-Anderson, Auspos, and Anderson, *Community Involvement*.

\(^{16}\) Fullbright-Anderson, Auspos, and Anderson as quoted in Harkavy et al., “Effective Governance,” 99.

\(^{17}\) “Anchor Institutions Task Force.”


\(^{19}\) Taylor and Luter, 8.

\(^{20}\) Snyder, *Digest of Education Statistics*.

133
features of anchor institutions as they have come to be known, features that a majority of American universities epitomize.

For a university to act as an anchor institution, however, these qualities must be employed for a specific purpose, referred to as an anchor mission. This mission is defined by Rita Hodges and Steve Dubb in their book The Road Half Traveled as “the conscious and strategic application of the long-term, place based economic power of the institution, in combination with its human and intellectual resources, to better the welfare of the community in which it resides.” The nature of this mission as articulated by Hodges and Dubb is unique. Unlike efforts at engagement by colleges and universities in the past, this mission does institutionalize the university’s commitment to community. This institutionalization expands the traditional understanding of a university by equally emphasizing three distinct missions: an education mission, a research mission, and now an anchor mission.

Where they have embraced this ethos, educational anchor institutions can adopt a number of economic, social, and/or cultural strategies to strengthen their surrounding communities. Economically, practices often include a commitment to hiring, procuring goods, and investing locally. For example, the university can choose to invest elements of its endowment in local community development financial institutions, as Harvard University and the University of Chicago have done. Schools can also choose to source a certain percentage of their produce locally, following the example of Middlebury College, University of Pennsylvania, and many others. Socially, universities have increasingly focused on utilizing their academic resources (e.g., graduate or medical students in need of experience, professors interested in making their research more applicable, and undergraduates seeking to develop professional skills) to address local issues. Stony Brook University in Long Island, New York, has encouraged professors and graduate students to prioritize “tackling local issues that support local industries and

---

21 Hodges and Dubb, Road Half Traveled, 147.
22 Hodges and Dubb, 147.
23 Community development financial institutions, or CDFIs, include credit unions, banks, loan funds, and venture capital funds that provide credit and financial assistance to programs and projects that contribute to the development of the local community.
24 In 1999, Harvard provided $20 million to local CDFIs in Boston. (For more information, see “Anchor Institutions.”) Likewise, in 2012, the University of Chicago decided to shift $1 million of its endowment into four community banks in a simple act that has since benefited local businesses and homeowners. (See Standish, “U of C Depositing $1 Mil.”)
25 Carlson and Bielmiller, “Campus as City,” 23.
interests." In the process, the school has harnessed these partnerships to become a well-respected research university.

Culturally, community colleges, research universities, and liberal arts institutions alike are expanding access to their facilities and are geographically integrating their campuses with cities in order to change the explicit and implicit dynamics between town and gown. Colby College’s newest residence hall is not located on or even adjacent to the campus itself but in the heart of downtown Waterville, Maine. Coincidentally, this facility, which was built to promote partnerships between the campus and city, is called the Main Street Commons.

Examples of institutional engagement abound, and although they differ drastically according to location, all stem from the recognition that universities and the communities they are part of have stakes in the success of each other. To again quote the Chronicle of Higher Education’s 2019 report “The Campus as City,” the popularity of educational anchor institutions is based on a growing realization that, “whether in small towns or in transforming cities, colleges have opportunities to stoke local economies, expand community resources, and prove their value to the public. At the same time, they can fortify themselves for years ahead.” This sentiment is the foundation upon which the term anchor institution is based. I turn now to the individual planks that constitute this platform and make up the argument for why its sentiment is relevant.

The Argument for Anchor Institutions

Having flushed out a general understanding of modern anchor institutions in the previous section, we can now identify the steps by which they are popularly justified. The main sources available in doing so are the policy proposals and evaluation reports that have been written to further institutions’ anchor missions. Using these documents as a starting point, this section identifies four main premises that underlie contemporary arguments for anchor institutions. Isolating these premises will make it possible for the next section to demonstrate how they are fundamentally grounded in a theory of the commons.

The argument for the existence of educational anchor institutions most often begins by articulating an unresolved problem. The picture painted of this problem is of American cities—and communities more generally—in a state of crisis. It calls

---

26 Carlson and Bielmiller, 31.
27 Carlson and Bielmiller, 31.
28 Carlson and Bielmiller, 9.
attention to cities across the United States where infrastructure is increasingly outdated and overwhelmed, poverty and homelessness have grown to epidemic proportions, and, because of declining industrial demand, those economies are failing.

The 2013 HUD report “Building Resiliency: The Role of Anchor Institutions in Sustaining Community Economic Development” begins with the example of Chester, Pennsylvania. Chester, this report claims, epitomizes the problems of “poverty, stagnation, and unemployment” that can be seen throughout urban America. The woes of the city, a former center for automobile manufacturing and shipbuilding, are said to stem “from the late 20th-century decline of an industrial economy in the United States.”

Similarly, in April of 2009, the AITF submitted to the HUD Secretary a report that aimed to demonstrate the power of “Anchor Institutions as Partners in Building Successful Communities and Local Economies.” This report, too, begins its argument by premising the problem of urban crisis, claiming that these deteriorating circumstances are a direct result of both deindustrialization and globalization. The AITF’s analysis of this disinvestment parallels that of the 2013 HUD report, stating, “Deindustrialization and globalization have undermined the traditional manufacturing-based economies in many American cities and metropolitan areas, leaving unemployment, poor schooling and general poverty in their place.” These two documents are emblematic of a much larger body of work that advocates for leveraging the power of anchor institutions. The problem they pose is primarily economic: cities are failing because traditional economies have been undermined by the social, technological, and global conditions of the latter part of the twentieth century. The logical conclusion left for the audience to draw is that a new economy of investment is desperately needed if these communities are to be saved.

If the first premise is that a crisis exists, the second premise asserts that so, too, do untapped assets for resolving the crisis. In 2002, the Initiative for a Competitive Inner City (ICIC) and CEOs for Cities summarized this idea in their report entitled Leveraging Colleges and Universities for Urban Economic Revitalization. The report concludes that, “despite their considerable size, colleges and universities are often overlooked as a component of urban economies.” Furthermore, not only are institutions a component of economic reinvigoration, but

30 Godsil and Brophy, Retooling HUD, 148.
31 Leveraging Colleges and Universities, 7.
their potential impact could be “enormous.” The report continues, “[Colleges and universities] have significant purchasing power, attract substantial revenues for their surrounding economies, invest heavily in local real estate and infrastructure, are major employers, and help to train the workforce and nurture new businesses.” Together, these two premises present what seems to be a perfectly balanced equation: with the deterioration of a manufacturing economy has come the rise of a knowledge economy, and as the former’s investment is lost, the latter’s should be expected.

ICIC and CEOs for Cities are not alone in identifying the potential present in universities. Alan Mallach, a Visiting Scholar at the Federal Reserve Bank of Philadelphia, has publicly extolled this same idea. Speaking on the subject of educational anchor institutions, Mallach expressed that “many of the same cities that are facing rampant disinvestment also ‘contain significant—although often underutilized—assets that may well be pivotal to the economic future of their metro areas and regions.’” While serving as a testament to the second premise of the justification of anchor institutions, Mallach’s use of the term asset is also rhetorically meaningful. Assets can be used to address a debt, restore financial or social stability in the community, and, in doing so, cultivate resiliency to combat future insecurity. Moreover, as the third premise will make evident, these assets are perfectly positioned to address the crisis already outlined.

Although these assets do not constitute a silver bullet and cannot fully replace the economic opportunities originally provided by the vanishing industrial sector, proponents argue that the assets are perfectly situated to begin addressing the issue. In the words of Ira Harkavy, founding director of the Netter Center for Community Partnerships at the University of Pennsylvania, “Universities and colleges … potentially represent by far the most powerful partners, ‘anchors,’ and creative catalysts for change and improvement in the quality of life in American cities and communities.” Again, this quote implies that the equation can be balanced so long as the right variables are manipulated.


---

32 Leveraging Colleges and Universities, 7.
33 Leveraging Colleges and Universities, 2.
34 “Building Resiliency,” 1.
35 Quoted in Hodges and Dubb, Road Half Traveled, xvii.
the most influential single work in the history of town planning,” Jacob’s text was one of the first to present a new theory of organic urban vibrancy that soon became the standard for all urban policy. The heart of her argument contends that cities need “a more intricate and close-grained diversity of uses that give each other constant mutual support, both economically and socially.” Viewed through this lens, a city’s failure to appropriately leverage the assets at its disposal can be read as a failure on the part of community leaders to integrate this system of mutual support—the logical result of which is likely to be imbalance and crisis. Although the concept of large public and private educational institutions taking on anchor missions is one step removed from Jacob’s discussion of sidewalks and neighborhoods, both nonetheless revolve around an inherent principle that “people must take a modicum of public responsibility for each other even if they have no ties to each other.”

This element of social and institutional integration is the third major premise underlying contemporary arguments for anchor institutions. That is, in order for the crisis-mitigating potential of anchor institutions to be tapped, this premise posits, the relationships that bind various elements of society together must be recognized. In this sense, specific mechanisms of governance are required for Jacob’s system of mutual support to exist. Only with such structure can the “assets” of universities be employed to address America’s economic crises. Whether financial, social, or cultural, assets can exist relatively isolated and unrealized, their relative benefits going without use. Likewise, that colleges and universities possess the potential to act as economic engines in their local communities does not mean they will always act on this potential. Without the appropriate knowledge, technology, or procedures of implementation, the resources that these assets represent cannot be accessed and channeled to their full effect—the effect in question being a university’s anchor mission to “better the welfare of the community in which it resides.” As such, to unlock the power that anchor institutions hold and, in so doing, resolve the crises that plague American cities, structures of governance must be in place to facilitate the institutional integration required for the realization of Jacob’s notion of mutual support.

Success, whether individual or institutional, is herein framed as a system of collaboration between the individual, society, and nature. No one person is solely

---

36 Fulford, “Jane Jacobs.”
37 Jacobs, Great American Cities, 14.
38 Jacobs, 83.
39 Hodges and Dubb, Road Half Traveled, 147.
responsible for their accomplishments in life; even the most “self-made” individual draws “upon a vast pool of knowledge and natural gifts they did nothing to create.”⁴⁰ In place of the individualist ideal, this blueprint for mutual support posits that all success arises from and depends continuously on the success of others. By mapping the relationships that bind university engagement to their social surroundings, contemporary advocates of anchor institutions implicitly emphasize an interdependence that constitutes a “culture of community”—the feeling that “we’re all in this together.”⁴¹ To summarize: in response to the crisis caused by globalization and deindustrialization, colleges and universities have risen to prominence as a means of combating disinvestment. Their institutional ability to do so, however, depends on a collaborative social dynamic that assumes the health of a community is co-constitutive and not a zero-sum war for individual survival.

The final premise of the argument for anchor institutions is that universities should, and in fact have an obligation to, address the societal crises that their communities face. This premise is founded on the assumption that acting in the best interest of the community is itself a form of enlightened self-interest for the university. For example, just as an institution’s educational mission correlates with its fiduciary success, so too does the emphasis on its anchor mission. In this sense, the university that chooses to act on its anchor mission is not doing so based on sheer altruism but rather on a broader appreciation of all the factors that contribute to its success. A student’s decision to enroll at the university, or the university’s ability to attract the best-suited professors, has as much to do with life beyond campus as life within. In other words, an institution, like an individual, does not exist in a vacuum. Furthermore, by encouraging these relationships and helping to solve real-world problems as they are manifested locally, universities can also further their other missions of research, teaching, and service that society expects of them.⁴² Institutions of higher education thus have “a strong economic stake in the health of their surrounding communities and——due to the scale and scope of their operations—the resources to make a genuine difference.”⁴³ Because universities have the ability to make a significant difference in the lives of those around them, the cumulative implication of these factors is that they have a strong incentive, and logical responsibility, to do so.

⁴⁰ Rowe and Barnes, Our Common Wealth, 72–73.
⁴¹ Alperovitz, Pluralist Commonwealth, 63.
As the premises that are commonly used to justify educational anchor institutions have been identified, a compelling overall theory for the existence of anchor institutions has begun to emerge. The equation is relatively simple: There is a crisis, and assets exist to address it. These assets, and those who traditionally manage them, bear an inherent responsibility to mitigate the crisis, but the proper relationships and procedures must exist so such assets can be “tapped.” Beyond highlighting the rhetorical logic of this argument, however, the identification of these four premises illuminates an underlying narrative about the role of educational institutions in society. This narrative holds that such key social institutions have a responsibility to serve the common good rather than just their own, and it is in this narrative that we find an inherent reliance upon the theory of the commons. It is to this theoretical connection that we now turn.

A Framework for Interpretation: The Theory of the Commons

Having illuminated the fundamental assumptions that dominate the field of anchor-institution research, we can begin to see how these assumptions fundamentally stem from a theoretical tradition of self-governance that is distinct from both market- and state-centered methods. This strain of political economy is often referred to simply as “the commons.” Interpreting university-based anchor institutions through this lens reveals a much more authentic image of their nature—an image that provides both a better understanding of these institutions overall and the tools with which to hold them accountable to their anchor missions. To establish the relationship between the theory of the commons and the basic arguments in support of anchor institutions, this section will offer a brief overview of the central pillars of commons theory before exploring relationship of those pillars to the premises enumerated in the previous section.

The origin of the term commons is often attributed to the system of agrarian land governance widely used in medieval England. Prior to the enclosure movement of the 1700s, significant tracts of English countryside were collectively owned by those who worked them. Although this land was most frequently part of an estate held by the lord of the manor under permission of the

---

44The term enclosure refers to the process by which collective or commonly held land is divided into individual parcels and privatized. The enclosure movement mentioned here occurred primarily between 1760 and 1820 throughout England. During this time, land use became restricted to those who claimed official ownership so that in village after village, common rights were lost. The resulting landless laborer (often pauperized) was “left to support the tenant-farmer, the landowner, and the tithes of the church.” For more information, see E. P. Thompson’s *The Making of the English Working Class* (1963).
Crown, those who worked the land managed its use. Under this system, these “commoners” dictated the number of grazing animals allowed to pasture, and/or the number of trees that could be cut for wood. Even the lord himself was bound to obey these rules. First practiced in the early 1500s and codified into law in 1827, the system expected that the lord “leave pasture enough to satisfy the commoner’s rights whether such rights are to be exercised or not.” In this way, resources were managed in common rather than privately. The system of political interaction that governed these communities grew out of this method of property distribution and the system of social organization it supported. Author, activist, and scholar Sylvia Federici describes this system in her book *Caliban and the Witch* as a series of “vast communalistic social movements” that, prior to implementation of feudalism, “offered the promise of a new egalitarian society built on social equality and cooperation.” Systems of “commoning” have existed successfully for millennia and belie the assumption that any resource accessible to a community at large will inevitably be exploited to the point of nonexistence.

Contemporary use of the commons as a theoretical framework began with concerns over natural resources. As these assets were increasingly depleted to the brink of collapse, politicians, scientists, sociologists, biologists, business owners, and others sought a new way of managing them. Alike, they observed that “neither the state nor the market is uniformly successful in enabling individuals to sustain long-term, productive, use of natural resource systems.” In *Governing the Commons*, one of the first comprehensive studies on the practice of commoning worldwide, American political economist and Nobel Laureate Elinor Ostrom introduced this concept as a “theory of self-organizing and self-governing forms of collective action.” This notion has been a fundamental principle within commons theory since the publication of Ostrom’s text in 1990. Since then, the term *commons* has evolved from referring exclusively to natural resources to referring to a combination of (1) resources, whether natural, social, or cultural, (2) the larger community within which these resources reside, and (3) the set of social protocols that governs interactions between the first two. The interplay between these elements underscores the fact that the practice of commoning is about much more

---

47 An argument first put forward by ecologist Garrett Hardin in 1968 as a critique of the commons. Since the publication of Hardin’s article, his critique has come to popularly be known as the tragedy of the commons.
48 Ostrom, *Governing the Commons*, 1.
49 Ostrom, 25.
50 Bollier, *Think Like a Commoner*, 15.
than shared resources; to quote David Bollier, founder of the Commons Strategies Group, “It’s mostly about the social practices and values that we devise to manage them.”

This emphasis on community does not ignore the undeniably powerful self-interest of human beings that liberal economists have for centuries posited as the driving force of individual and social prosperity. Rather, it seeks to contextualize this reality within a broader understanding of human nature. Humans may indeed have unattractive traits fueled by individual fears and ego, “but we are also creatures entirely capable of self-organizing and cooperation; with a concern for fairness and social justice; and willing to make sacrifices for the larger good and future generations.”

Looked at through this lens, a successful commons is simply a space in which individuals have learned to collectively govern their community so as to balance these ego-centric qualities for the well-being of all.

The successful management of a commons is based on principles of decentralized governance and collective accountability. Within this system of management, “ownership” is exercised at the level of the community rather than the individual. As with the pastures of fifteenth-century England, wealth in this sense is held in common. Though this concept of commonwealth may on the surface appear similar to philosophies of state-socialism, it is distinguished by a focus on scale and decentralization. Commons-based theory rests on the principle of subsidiarity, “the idea being that decentralized, small-scale solutions, should as far as possible be a default option.” In other words, when given an option, effective self-governance requires that decisions be made at the lowest possible level. This system allows wealth to be sustainably held in common because those who hold it know the unique context in which it exists and so are able to best regulate it. To return to the natural-resource analogy, the commons works “because people come to know and experience the management of a resource in its unique aspects. They come to depend on each other and love this forest or that lake or that patch of farmland.” As a result of this interdependence, individuals hold one another accountable for the stewardship of these resources. The underlying principle in this practice is that “the relationships between people and their

---

51 Bollier, 19.
52 Bollier and Helfrich, eds., Wealth of the Commons, xv.
53 Alperovitz, Pluralist Commonwealth, 59.
54 Alperovitz, 83.
55 Bollier, Think Like a Commoner, 12.
resources matter.” The intimacy of these relationships is what allows for—and in fact demands—self-governance and the collective management of resources, be they natural, cultural, digital, institutional, or otherwise.

In sum, the theory of the commons posits that all humans inherit innumerable “gifts” from society and the natural world. These assets are inherited jointly, and together, individuals hold them in trust for future generations. It is considered essential to both human and planetary well-being that we do so. Economists and many others often fail to account for how commons contribute to social and economic prosperity, however, because those contributions are difficult to monetize and also because the liberal economists’ archetypal *homo economicus* acts solely on a rationality defined by self-gain alone. Above all else, the language of the commons creates a unifying vocabulary that sheds light on the otherwise often ignored foundation of human existence—the social, natural, and cultural inheritances that all individuals share—so they can be passed on, undiminished and, hopefully, enhanced.

Despite relying on principles and rhetoric that can be easily interpreted as utopian, the commons is not a foreign or radically liberal idea. In fact, this theory owes as much to traditional conservative values as to more progressive thought. To quote self-described conservative Jonathan Rowe, “Conservatism is, or at least used to be, a way of thinking about society as a whole and the qualities that help maintain it.” Traces of commons theory thus can be found across the political spectrum. And far from being antithetical to the important societal aspects of market or state, this commonwealth system of political economy is important insofar as it maintains a balance between these elements. Markets are uniquely able to coordinate the actions of many decentralized social actors, connecting need (demand) to surplus (supply). Commons theory acknowledges that for any community to successfully fulfill the needs of its members, this ability to coordinate demand and supply must exist. It is only when the market is left unchecked that theorists critique it for reducing goods and services to the base denominator of commercial exchange and

56 Bollier, 12. In her text *Colonial Lives of Property* (Duke University Press, 2018), Brenna Bhandar develops a powerful and complementary critique of the systemic destruction of these intimate relationships among individuals, their surroundings, and the fabric of society. She terms this process “abstraction.” Bhandar’s argument lends a valuable perspective on the inextricable relationship between the way in which property is defined under neoliberalism and continuing practices of colonialism and imperialism. She terms this mutually constitutive relationship a “racial regime of ownership.”
57 Rowe and Barnes, *Our Common Wealth*, 11.
58 Rowe and Barnes, 54.
59 Alperovitz, *Pluralist Commonwealth*, 166.
for elevating pursuit of profit above the satisfaction of human and community needs as a result. The state likewise has the potential to catalyze or cultivate healthy commons by creating spaces in which they can flourish; its threat, however, lies in the tendency to infringe on the commons by centralizing governance and formalizing what should be intimate and place-based relationships. Together, the three elements of market, state, and commons are mutually constitutive and necessary for a flourishing society. The danger arises only when their balance is thrown off and communities risk losing the shared gifts that allowed them to thrive in the first place.

***

Long before they were labeled as anchor institutions, colleges and universities had a well-established relationship with commons theory. Typically, this was because of their unique ability to help diverse communities come together and produce knowledge. As Professors Madison, Fischmann, and Strandburg have put it, the university is inherently a “constructed cultural commons.” In this sense, it functions as a space dedicated to the cooperative generation of knowledge. Competition, rivalry, and profit-making doubtless exist in this process, but there is a general presumption “that knowledge should not become a proprietary product.” Very little in academia is based solely on the purchase or sale of these ideas; the practice of peer review exemplifies this tradition of collaboration for the sake of knowledge above all else. Such practices are typical in the university, for the academic community has become the cultivator of this particular resource. In this more traditional framing of the university as a commons, the university represents an inherited asset to be collectively stewarded because it holds the common gift of knowledge on which new generations can build.

In contrast to this framing, however, the argument for anchor institutions presents a new way of perceiving the university as a commons. The premises outlined in the previous section imply that although valuable, the depiction of the university as a constructed cultural commons is incomplete, in that universities have a larger role to play within their communities beyond facilitating the pursuit and integrity of knowledge. Culturally, an institution of higher education can act as an incubator of and reservoir for learning, but it also, socially and economically,

---

60 Alperovitz, 165.
61 Bollier, Think Like a Commoner, 72.
62 Bollier, Think Like a Commoner, 73.
has the potential to stimulate and stabilize the town, city, or region in which it is located.

The university is far more of a commons than academics often acknowledge. For example, the University of Pennsylvania not only provides a repository for knowledge and learning but also acts as a conduit for engagement between the institution (consisting of academic, human, and corporate resources) and the community of West Philadelphia. The school’s Comprehensive Neighborhood Revitalization plan engages other community organizations and affected community members to collectively govern how these resources will be employed. In a parallel vein, Miami Dade Community College embraces its anchor mission through similar programs that emphasize its surroundings,63 describing itself as an “opportunity pipeline” (which flows both ways) connecting stakeholders both on and off its various campuses in order to “effect community change and [community] wealth building.”64 These case studies demonstrate and embody a broader understanding of the university operating as a commons. The logic behind this is relatively simple: local communities provide vast tax subsidies and in many other ways contribute to the existence and success of universities, actively collaborating to steward a common resource. In return, anchor-institution advocates argue, these communities should be able to jointly draw upon these assets for their collective benefit.

In sum, the practices of anchor institutions rely on a very specific set of assumptions that are intimately linked to the theory of the commons. Whereas traditional depictions of the university as a commons highlight only the arena of knowledge stewardship, contemporary justifications of anchor institutions can and should be understood as expanding this model to social and economic realms, with universities constituting assets in more than one regard. Just as with commonly held irrigation systems that have existed in the south of Spain for centuries, or rapidly proliferating digital commons such as Wikipedia and open-access scholarly publishing platforms, a university is an intergenerational collaboration of individuals, society, and nature. As such, the outcomes of such collaboration are intersectional and reach far beyond a cultural, economic, or social impact alone. To

63 Examples of these programs include the college’s inclusive hiring practices, its goal of directing 20%–27% of purchasing dollars toward minority-owned businesses, projects that provide entrepreneurial and educational support to micro-enterprises, nonprofits, and individuals through the Meek Center, and focused neighborhood-revitalization initiatives. More information can be found in Hodges and Dubb, Road Half Traveled. 49.
64 Interview with Eduardo J. Padrón, President of Miami Dade College, in Hodges and Dubb, Road Half Traveled, 48.
analyze this relationship, let us now more thoroughly interpret the justification for university-based anchor institutions by applying the language of the commons.

The first premise underlying educational anchor institutions is that American cities are in crisis. This assumption stems directly from the principle of commons theory claiming that both state and market have failed—and always will fail—to sustainably manage humanity’s inherited common wealth. When left to their own devices, commons theorists argue, both systems of political economy will trend toward crisis. On the one hand, the state veers toward a Hobbesian Leviathan, wherein a sovereign, brutal force ensures public cooperation and order through raw coercion.65 On the other, the market habitually colonizes individuals, spaces, and relationships, reducing the world to the single dimension of monetary value.

Under the logic of commons theory, then, the negative externalities of globalization and deindustrialization that afflict American cities have resulted from the inherent limitations of contemporary market- and state-based systems of political economy. The crises that have resulted are natural results of a system based on an ever-expanding dispossession that is synonymous with the enclosure of the commons. Crisis has thus arisen because the resources within American cities have been poorly stewarded, because of both the market’s exploitation and the state’s oppressive centralization. Ostrom’s conclusion rings true: “Neither the state nor the market is uniformly successful in enabling individuals to sustain long-term, productive use of natural resource systems.”66 The forces of globalization and deindustrialization are the economic embodiments of a system that prioritizes profit over human well-being and exploits society’s collective inheritances. In response to these crises, advocates of anchor institutions are effectively proposing a return to the practices of the commons.

This solution presents universities as assets perfectly suited to confront the crisis precisely because of their ability to act as a specific kind of commons. Rather than turning toward traditional solutions grounded in a market or state paradigm, those who advocate for the application of anchor institutions are relying on a commons-based approach. As a means of combating economic crisis and addressing the basic needs of a community, anchor institutions are unique. Instead of implementing state-based social programs or mandating austerity measures and giving tax incentives to outside corporations, cities and towns that follow an anchor-institution model of economic rehabilitation are turning to the assets already

65 Ostrom, Governing the Commons, 9.
66 Ostrom, 1.
shared within their communities. By partnering with one of the 4,100 universities across the country and, by implication, that school’s portion of the four million employees, $533 billion in endowments, and 21 million students, a community is, in essence, beginning to reinstall a system of self-organization and governance at the local level.

Educational anchor institutions have been widely hailed as a solution to America’s economic woes precisely because of the qualities they share with the commons. First, they embody an intimate relationship between people and place, as the school is anchored in a specific community; second, the institution itself represents an inherited asset to be held in trust for future generations; and third, the success of any town-gown relationship relies on the principle of subsidiarity, which will necessarily prioritize small-scale, decentralized solutions. Depicted as assets, universities exist in a space between public and private with the potential to benefit a broader common good. To approach a local institution of higher education as a source of community investment in response to the crises of globalization and deindustrialization is to view it as a common social and economic resource, but even so, such an institution is clearly not an “asset” alone, as with any commons.

The university acting according to its anchor mission is simultaneously a resource for the community at large, a community in and of itself, and a unique matrix of social protocols. If the university is to serve as an anchor institution for a larger community, this complex interplay therefore requires a system of self-governance within the community. In order for resources to be employed at all, the university, its surrounding neighbors, local government, etc. must collectively agree on a method of governing (either formally or informally) their interdependent relationship. Only by effectively regulating the use of these resources can such relationships be justified by both the university and the community. These are the “certain forms of governance” that must be in place to promote the social and institutional integration required by the third major premise of anchor institutions. If Jane Jacobs’s ideal of “constant mutual support” between the diverse elements of society is to be realized, the intricate maze of these relationships must be navigated.67 This embodies the notion of a political economy “organized on principles of interdependence, designed with a coherent orientation towards the common good,” upon which the theory of the commons is founded.68

67 Jacobs, Great American Cities, 14.
68 Alperovitz, Pluralist Commonwealth, 58.
If the university is to operate as an anchor institution within the dynamics of a commons, not only must there exist a way of sustainably stewarding collective resources, but the university itself must have an incentive to act in accordance with an anchor mission—as specified by the fourth premise underlying educational anchor institutions. As stipulated in this premise, universities possess an “enlightened self-interest” that motivates their collaboration with outside actors in the community, a motivation stemming from the recognition that no institution can be successful in total isolation. This same realization has supported systems of commoning for centuries. As both a steward of the commons and a community reliant upon that common wealth, the university is dependent on its own reservoir of inherited assets: a functioning municipality that ensures the upkeep of public infrastructure, a healthy environment with minimal pollution and plentiful natural resources, and local institutions capable of attracting prospective students and professors. It is thus in the best interest of a university to steward these resources in turn. It is worth repeating that institutions of higher education “have a strong economic stake in the health of their surrounding communities and—due to the scale and scope of their operations—the resources to make a genuine difference.”

Broadly, it is this principle of incentivized collaboration that allows a community to combat the problems of free riding, lack of commitment, and compliance monitoring that typically confront a commons.

As the constitutive elements of the argument for educational anchor institutions are drawn out, a broader theory explaining their existence and necessity comes into focus. The theoretical principles used to justify this logic stem not from a political economy based on either market or state but rather from one grounded firmly in the notion of the commons. The policy-based analysis of “best practices,” to which much of the existing scholarship on anchor institutions has so far been dedicated, consistently prioritizes small over big, local over distant, and nonmonetary over monetary relationships. From the premise that a crisis of deindustrialization and globalization exists, to the conclusion that the solution is a return to the assets that are already present within a community, contemporary arguments on behalf of educational anchor institutions are firmly entrenched in the ideals of the commons. The argument for anchor institutions not only justifies an expanded interpretation of the university as a commons but also premises the pioneering work already being done to create an alternative political economy set

---

70 Ostrom, *Governing the Commons*, 27. For more on how commons mitigate these issues, see Ostrom’s third chapter, “Analyzing Long-Enduring, Self-Organized, and Self-Governed CPRs.”
71 Rowe and Barnes, *Our Common Wealth*, 2.
free from the binary of market and state. If the pragmatic practices of anchor institutions can truly be said to rest on this theoretical foundation, however, the question becomes, What are the implications of such practices of institutional engagement?

**Conclusion**

The historic design of the university has been to create walled cities dedicated to the pursuit of scholarship and knowledge—an educational fiefdom. What I am arguing is in fact the opposite. I advocate for the university adopting an alternative metaphor for its identity. In place of the citadel, a paragon of isolated, uncontaminated intellectualism, the academy must instead aspire toward the commons. Human beings develop, adopt, and live by metaphors. We have lived for decades with a certain model for the institutionalization of academia. It is time we consider another.

By realizing the potential of universities to act as social assets aimed at achieving common abundance rather than engineering scarcity, anchor institutions inherently disrupt the dominating narratives of individuality, autonomy, production, and commodification that mark our current paradigm. What do we want at the center of our world: growth and consumption, or sustainability and collective stewardship? One is based upon private wealth, the other on common wealth. One is based upon competition, the other on cooperation.

Commons theory premises a universal interdependence that directly contradicts liberal economics’ rational, autonomous, and self-sufficient *homo economicus*. As a groundbreaking manifestation of this, arguments for educational anchor institutions rely on a desire to cultivate community wealth, which will act as a foundation upon which individuals and future generations alike can build. This is not a novel idea; from Wendell Berry to Alexis de Tocqueville, scholars have noted the importance of “a proper community” and “self-interest rightly understood.” By drawing a connection between educational anchor institutions and the commons, this paper is meant as the beginning of a much larger conversation. The popularity of educational anchor institutions at home and abroad is a promising sign of what may come. Universities are still developing what it means to be an anchor institution. The concept has evolved substantially in recent history. My analysis provides universities that seek to embody the role of anchor institution with an aspirational ideal. As educational anchor institutions encounter new challenges and confront opposition from both market and state rationales, the challenge will be to hold the evolution accountable to the ideals that make it so revolutionary.
Bibliography


THE IMPACT OF NURSE STRESS, JOB DISTRESS, JOB SATISFACTION, AND MOOD ON PATIENT OUTCOMES

MADELYNNE WRIGHT & ABBEY COLLINS, BUTLER UNIVERSITY
MENTOR: TARA LINEWEAVER

Abstract

This study examines the intercorrelations among nursing home healthcare providers’ stress, job distress, job satisfaction, and mood, as well as the relationships between these factors and patient outcomes. Additionally, the study investigates the effect that Butler University’s Music First! program had on nurse job perceptions. A total of 63 nurses from nine long-term care facilities completed questionnaires asking about stress, job distress, and global job satisfaction. Nurses’ depression, anxiety, and hostility were also assessed, with the 15-item Profile of Mood States questionnaire. Of these 63 nurses, 20 completed the questionnaires at three time points during the implementation of the Music First! intervention. Patient outcomes included standardized assessments of agitation, psychiatric symptoms, cognitive functioning, and depression, as well as documented instances of patient falls and reportable behaviors. Intercorrelations among nurse stress, job distress, job satisfaction, and affect ranged from $r = -0.565$ to $r = 0.627$. All correlations except one reached statistical significance. Similarly, correlations between patient outcomes and nurses’ job-related self-reports documented strong relationships between nurses’ affective responses to their jobs and patient outcomes. For example, the patients of nurses reporting greater stress at work were more agitated ($r = 0.335$) and exhibited more severe dementia (SLUMS $r = -0.291$), more severe psychiatric symptoms ($r = 0.293$), and more frequent reportable behaviors ($r = 0.436$) than did patients of nurses reporting less stress.

Finally, the Music First! program was associated with a near-significant decrease in nurses’ job-related stress ($p = 0.054$). Although these results are correlational and cannot determine the direction of these relationships, these findings do suggest that improving nurses’ feelings about their jobs, improving patient behaviors, or implementing programs designed to improve patients’ quality of life can have a positive impact on the nursing home environment.

Many individuals face job-related stressors throughout their careers. Job-related stress is defined by the Centers for Disease Control and Prevention (CDC,
2014) as “the harmful physical and emotional responses that occur when the requirements of the job do not match the capabilities, resources, or needs of the worker.” Nurses experience job-related stress at an extremely high rate compared to people in other occupations (Redfern et al., 2002). Nurses who are newer to the field report higher rates of stress than do nurses who have been in the field for longer; this is possibly a result of better coping mechanisms by the more-experienced nurses (Zimmerman et al., 2005). Job-related stress is essential to understand because it affects the nurses who experience it as well as their patients.

Several research studies have documented the negative effects of job-related stress in the nursing profession. When nurses experience high amounts of stress, they are more prone to absences (Dugan et al., 1996). Additionally, increased job-related stress is correlated with lower job satisfaction (Redfern et al., 2002; Zimmerman et al., 2005). Finally, there is an association between stress and commitment. The more stressed nurses are, the less committed they are to their roles (Redfern et al., 2002).

One possible outcome of prolonged job-related stress is career burnout, which is defined as a “multidimensional, complex phenomenon that involves exhaustion, cynicism, and ineffectiveness” (Maslach et al., 2001). Burnout, like job-related stress, is linked to absenteeism, decreased job satisfaction, decreased occupational commitment, and increased stress levels (Maslach et al., 2001). In a comprehensive study, nurses were found to be particularly vulnerable to career burnout when compared to more-advanced healthcare practitioners, with this difference being upward of 20% (Lyndon, 2016). Burnout may be particularly problematic in nursing homes. Mukamel and colleagues (2009) found that staff in this setting experience turnover at a rate of 55% to 75%. For nurses who do not leave their positions, those who experience burnout take more short-term absence leaves as a result of their emotional exhaustion and also take longer absence leaves than do peers who experience less burnout (Anagnostopoulus & Niakas, 2010).

Beyond the effect on the nurses themselves, career-related burnout and job stress negatively influence the quality of care that the nurses’ patients receive. In a study examining nurse burnout and common infections (urinary tract and surgical-site) during hospital stays, higher rates of nurse burnout led to higher rates of patient infections (Cimiotti et al., 2012). Burnout can be associated with depersonalization and emotional exhaustion, and these in turn have also been demonstrated to affect patient safety as perceived by nurses (Halbesleben et al., 2008).

Beyond burnout, nurses’ stress levels in general also negatively affect patient outcomes. As stated previously, when nurses experience more job-related
stress, they are less committed to their jobs. This is an issue because nurse commitment is directly correlated with patients’ quality of life (Bishop et al., 2008). In units where nurses report higher stress levels, there have also been more patient incidents, patient falls, and medication errors (Dugan et al., 1996). Thus, nurse stress, as well as burnout, is important because of its consistent negative effect on patient care and outcomes.

Although several studies have examined nurse stress and patient outcomes in large, multi-unit hospital settings, few have been conducted within the setting of a nursing home. The current study investigated nursing home healthcare providers’ stress, job distress, job satisfaction, and mood, as well as the impact that these factors have on patient outcomes in the nursing home setting. The first hypothesis was that higher stress levels would negatively correlate with job satisfaction but positively correlate with job distress and negative affect. The second hypothesis was that these four factors would relate to patient outcomes, with high levels of stress, job distress, and negative affect being negatively associated with and high levels of job satisfaction being positively associated with patient outcomes. More specifically, higher levels of stress, job distress, and negative affect were expected to negatively affect patient outcomes, but higher levels of job satisfaction were anticipated to relate to more positive patient outcomes.

In addition to these relationships, the current study evaluated the impact of Butler University’s Music First! program on nursing home staff. Music First! examines the effect that listening to personalized music playlists has on sundowning behaviors of individuals with dementia. Prior research indicates that passive listening to individualized music selections improves both agitation and cognition of nursing home residents with dementia (Gerdner, 2000; Shiltz et al., 2015, 2018). In investigating the effect of this program on nurses’ job-related perceptions, the third hypothesis of this study was that Music First! would decrease nurses’ levels of stress and job distress as well as improve their mood and job satisfaction through its positive effects on the dementia-related symptoms of the patients in their care.

Method

Participants

Participants included 63 nurses recruited from nine nursing homes located in the Indianapolis area where the Music First! program was implemented. The average age of participants was 39 years (SD = 13.01), and 87.3% of participants
were female. The nurses had worked with elderly patients for an average of 12.64 years \((SD = 9.64)\) and had worked in their current nursing home for an average of 3.42 years \((SD = 4.59)\) at the initiation of this study.

**Materials**

*Nurse Stress*

Nurses rated their stress levels at work across the last week on a five-point Likert-type scale that ranged from 0 (very low) to 4 (very high). Higher scores thus reflected greater work-related stress.

*Nurse Job-Related Distress*

Staff completed an adapted version of the Job-Related Distress Survey (Lepnurm et al., 2009). In this survey, nurses rated statements about their job-related distress on a five-point scale from 0 (very untrue) to 4 (very true). The survey included three subscales: emotional reactivity (e.g., “I feel that work has desensitized my feelings”), fatigue (e.g., “I have enough time to devote to all of my guests”), and stress relief (“I receive the support I need from my supervisor”). Higher scores for the emotional-reactivity and fatigue subscales indicated more job-related distress, whereas higher scores on the stress-relief scale suggested less distress.

*Nurse Job Satisfaction*

The Physician Worklife Survey (PWS; Williams et al., 1999) was used to assess staff job satisfaction. The PWS examines global job satisfaction, career satisfaction, and global specialty satisfaction, but the present study focused only on global job satisfaction. Nurses responded to four statements on a five-point Likert-type scale that ranged from 0 (very untrue) to 4 (very true). Examples included “I find my present work personally rewarding” and “Overall, I am pleased with my work.” Higher scores indicated more satisfaction.

*Nurse Affect*

A shortened version of the Profile of Mood States (POMS; McNair, Lorr, & Droppleman, 1971) was used to evaluate nurses’ negative affect. This 15-item survey asked participants to self-rate their current feelings on a five-point Likert-type scale from 0 (not at all) to 4 (extremely). These feelings related to three
subscales (depression, anxiety, and anger/hostility), each comprising five items, with higher scores indicating more negative affect.

**Patient Falls and Reportable Behaviors**

Medical records provided data on documented patient falls and reportable behaviors.

**Dementia-Related Symptoms in Patients**

**Patient Agitation and Behavior.** The Cohen-Mansfield Agitation Inventory (CMAI; Cohen-Mansfield et al., 1989) and the Neuropsychiatric Inventory (NPI; Cummings et al., 1994) were used to measure patient agitation and behavior. The CMAI, a 29-item scale, measured aggression as perceived by a primary caregiver. Caregivers reported the frequency with which participants displayed physically aggressive, physically nonaggressive, and verbally aggressive behaviors. The NPI examines 10 subdomains of dementia-related behavioral symptoms (delusions, hallucinations, agitation/aggression, dysphoria, anxiety, euphoria, apathy, disinhibition, irritability/lability, and aberrant motor activity) as perceived by the caregivers. Within each domain, caregivers rated the frequency of the behavior on a four-point scale, the severity of the behavior on a three-point scale, and the distress that the behavior caused on a five-point scale. On both measures, higher scores indicated more agitation and problematic behavior.

**Patient Cognition.** The Mini-Mental State Examination (MMSE; Folstein et al., 1975) and the Saint Louis University Mental Status exam (SLUMS; Tariq et al., 2006) were used to measure patients’ cognitive abilities. The MMSE, an evaluation commonly used by clinicians and researchers, assessed residents’ cognition levels based on a 30-point scale. The SLUMS exam was also a 30-point evaluation and was used to identify individuals with cognitive impairments. Higher scores reflected stronger cognitive abilities, whereas lower scores reflected more severe dementia.

**Patient Mood.** The Patient Health Questionnaire-9 (PHQ-9; Kroenke et al., 2001) was used to examine patient mood. This 10-item questionnaire asked primary caregivers to identify the number and severity of depressive symptoms exhibited by their older adult residents.
Procedure

This retrospective study examined data collected as part of the Music First! program. Music First! is a research project that examines the effect on sundowning behaviors and dementia-related symptoms of older adult nursing home residents of listening to personalized music playlists one to three times a week for 30 minutes. This music intervention lasted six months. To qualify for the study, nursing home residents had to have a diagnosis of dementia and exhibit agitation or be on a medication to treat agitation. At the beginning of the Music First! study, residents completed several standardized neuropsychological tests that examined their dementia-related symptoms such as agitation, neuropsychiatric behaviors, cognition, and mood. These symptoms were reassessed at mid-study (three months into the music intervention) and post-study (at the completion of the six-month intervention). Relevant to the current study, as part of Music First!, the patients’ nurses filled out a questionnaire about their job-related perceptions. It was the original hope that the nurses would be followed throughout the study to complete this questionnaire at baseline, mid-study, and for post-study follow-up. Although some nurses finished all three stages of the study, the number of nurse participants decreased across time (baseline n = 63; mid-study n = 33; post-study n = 20) because of attrition. Baseline stress levels did not differ between those who provided data at all three time points of the study ($M = 2.20, SD = 0.89$) and those who dropped out ($M = 2.00, SD = 0.93$), $t(61) = 0.81, p = .42$.

Analysis

After the data-collection period ended, the first step was to categorize nurses by the nursing home facility they worked for. Because specific patients were not linked to specific nurses, the next step was to calculate an “average” patient in each facility. The final step was to enter the average number of reportable behaviors, number of falls, and scores on psychological measures of agitation, behavior, cognition, and mood of the patients living in the facility where each nurse worked.

Two sets of correlational analyses using SPSS software addressed the first goal of the study. The first set of analyses examined intercorrelations among nurse job perceptions. The second set of analyses calculated the correlations between nurse self-reports and patient outcomes.

Six analysis of variance (ANOVA) tests using SPSS software addressed the second goal of the study. These analyses examined changes in nurses’ job perceptions as a result of introducing Butler University’s Music First! program into
the nursing homes. Six outcomes (nurses’ stress, global job satisfaction, fatigue, emotional reactivity, stress relief, and mood) from baseline, mid-study, and post-study were compared. The within-subjects variable was time. For any analyses that reached or neared significance, follow-up analyses separately compared baseline to mid-study, and mid-study to post-study, to investigate when the change in nurses’ job perceptions occurred.

Results

Intercorrelations Among Nurse Perceptions

The first analysis examined the intercorrelations among nurse stress, job distress, job satisfaction, and mood. All correlations except one (that between the emotional-reactivity subscale of the distress measure and job-related stress relief ($r = -.190, p = .135$) reached statistical significance (Table 1). Nurses who reported more stress also reported significantly more job distress (fatigue: $r = .627, p > .001$; emotional reactivity: $r = .494, p > .001$) and negative affect ($r = .406, p = .001$) and were significantly less satisfied with their jobs ($r = - .392, p = .001$). In contrast, nurses who sought stress relief at work tended to experience less stress ($r = -.265, p = .036$), fatigue ($r = -.455, p < .001$), and negative affect ($r = -.401, p = .001$) as part of their jobs. Nurses who sought opportunities to relieve stress at work also reported greater job satisfaction, $r = .442, p < .001$. Finally, nurses who experienced greater global job satisfaction also reported less job-related distress (fatigue: $r = -.449, p < .001$; emotional reactivity: $r = -.529, p < .001$) and negative affect ($r = -.565, p < .001$).

Relationships Between Nurse Perceptions and Patient Outcomes

The second analysis examined the relationships between patient outcomes and nurses’ job-related self-reports (Table 2). The patients of nurses reporting greater stress at work were more agitated ($r = .335, p = .007$) and exhibited more severe dementia (SLUMS: $r = -.291, p = .021$), more severe psychiatric symptoms ($r = .293, p = .05$), and more frequent reportable behaviors ($r = .463, p = .012$) than did patients of nurses reporting less stress. Patients whose nurses reported more fatigue engaged in more reportable behaviors ($r = .340, p < .05$) and scored lower on dementia screening (MMSE: $r = -.271, p = .005$; SLUMS: $r = -.316, p = .007$), whereas those whose nurses reported more emotional reactivity were more agitated ($r = .248, p = .05$) and more cognitively impaired (MMSE: $r = -.294, p = .019$; SLUMS: $r = -.333, p = .008$) but were less depressed (PHQ-9 severity: $r = -.288$, $p = .010$).
The patients whose nurses were more satisfied with their jobs exhibited fewer reportable behaviors ($r = -.337, p = .013$) and had better cognitive abilities (MMSE: $r = .275, p = .004$; SLUMS: $r = .336, p = .007$). Finally, patients of those nurses who sought stress relief at work experienced more falls ($r = .319, p = .016$) but also had better cognitive functioning (SLUMS: $r = .276, p = .029$) and exhibited fewer reportable behaviors ($r = -.322, p = .017$). Nurses’ self-reported mood did not significantly correlate with any of the patient outcome measures.

Effect of Butler University’s Music First! Program

The final analysis examined the effect of the implementation of Butler University’s Music First! program on the job-related perceptions of the 20 nurses with data at all three time points (Table 3). The main effect of time did not reach significance for any of the job-distress subscales: fatigue [$F(2, 18) = 0.679, p = .519, η_p^2 = 0.070$]; emotional reactivity [$F(2, 18) = 1.604, p = .228, η_p^2 = 0.151$]; stress relief, [$F(2, 18) = 0.771, p = .477, η_p^2 = 0.079$]. Additionally, the main effect of time did not reach significance for job satisfaction [$F(2, 18) = 0.512, p = .608, η_p^2 = 0.054$] or mood [$F(2, 18) = 0.642, p = .538, η_p^2 = 0.067$]. The main effect of time did near significance for nurses’ job-related stress, however: $F(2, 18) = 3.451, p = .054, η_p^2 = 0.277$. The follow-up ANOVA indicated that nurses’ stress levels improved from baseline to mid-study [$F(1, 19) = 4.75, p = .042, η_p^2 = 0.200$] but worsened from mid- to post-study [$F(1, 19) = 5.444, p = .031, η_p^2 = 0.223$].

Discussion

This study had three goals. The first was to document the intercorrelations between nurses’ job-related perceptions. The second was to examine the impact of nurses’ job-related perceptions on patient outcomes. The final goal was to investigate the impact of Music First! on nurses’ job-related perceptions.

The results of this study supported the first hypothesis that job-related stress would positively correlate with job-related distress and negative mood but would negatively correlate with job satisfaction. Nurses who were more stressed at work also experienced more job distress, had more negative affect, and were less satisfied with their jobs. This relationship between job-related stress and lower job satisfaction aligns with findings in the studies conducted by Zimmerman et al. (2005) and Redfern et al. (2002). Both studies found that job satisfaction and job-related stress are negatively correlated. Whereas Zimmerman et al. looked at these
relationships across several types of care settings and Redfern et al. focused on nurses working within only one nursing home, the current study expanded on their findings by assessing these relationships across multiple nursing home facilities within a large metropolitan area.

The second hypothesis of this study, that nurses’ negative job perceptions would be associated with worse patient outcomes, and nurses’ positive job perceptions would relate to better patient outcomes, was partially supported. Nurses with greater stress cared for patients who displayed more severe dementia, more severe psychiatric symptoms, more reportable behaviors, and greater agitation. Additionally, nurse fatigue and emotional reactivity correlated with more negative patient outcomes such as more frequent engagement in reportable behaviors and more severe depression. In contrast, nurses with high job satisfaction had patients with fewer reportable behaviors and better cognitive functioning. Of the many variables included in this study, only nurses’ negative affect failed to correlate with patient outcomes.

The strong relationships documented between nurses’ job perceptions and patient outcomes parallels the results of Cimiotti and colleagues (2012), who found that increased nurse burnout led to worsened patient outcomes. At the same time, the current study suggests that the potential effects of nurses’ feelings about their work extend beyond burnout, with less extreme negative perceptions (e.g., everyday stress levels, fatigue) also possibly affecting patients. The current study additionally indicates that the influence of these negative job perceptions reaches beyond typical infections to dementia-related symptoms such as reportable behaviors and agitation. These findings also align with those of Halbesleben and colleagues (2008), who found increased nurse depersonalization and emotional exhaustion to be associated with decreased patient safety, and those of Bishop et al. (2008) and Redfern et al. (2002), who documented significant relationships between nurses’ levels of job commitment and patient quality of life. At the same time, current results do not parallel those of all past studies (e.g., Dugan et al., 1996). Although the current study did not find a positive correlation between nurses’ stress and the frequency with which their patients experience falls, it did document significant relationships between other patient characteristics, such as agitation levels, cognitive function, psychiatric symptom severity, and reportable behaviors, and nurses’ self-reported stress levels at work.

Finally, the third hypothesis of this study, that implementing Music First! would positively affect nurse job perceptions, was weakly supported. Only job-related stress showed a near-significant improvement with the implementation of
the music intervention. Follow-up analyses indicated that nurses’ stress levels improved from baseline to mid-study but then reverted back to baseline levels from mid-study to post-study. Preliminary analyses of patients’ responses to Music First! have also shown that residents’ psychological well-being and agitation tend to improve during the first three months of the music intervention and then gravitate back toward baseline levels across the next three months (Vitelli et al., 2018). The trend documented in the nurses’ stress levels in the current study may thus mirror that of their patients’ psychological responses to the music program.

Although the results of this study largely support the hypotheses, the generalizability of these findings may be limited by several factors. The first limitation of this study is the correlational nature of its results. Because the results are correlational, the direction of these relationships is unclear. One possibility is that nurses’ job perceptions influence patient outcomes, consistent with the second hypothesis. Alternatively, patients’ symptoms and behaviors may influence nurses’ perceptions of their jobs. Future studies could introduce policies or interventions designed to improve nurses’ job-related stress or distress and observe the resulting changes on patient outcomes to help determine the direction of these relationships. Nonetheless, the current study does highlight the critical relationships between nurses’ job perceptions and patients’ outcomes; thus, improving either aspect of the nursing home environment may have the potential to improve the other.

A second limitation of this research was nurse attrition across the six-month study. Although the missing data that resulted from attrition is unfortunate, the study did accurately capture the high turnover rate seen all too often in nursing homes. The high turnover rate of nurses who experience burnout is also reflected in studies by Mukamel and colleagues (2009) and Anagnostopoulus and Niakas (2010). Interestingly, the baseline stress levels of nurses who dropped out were not different from those who completed the study. This lends validity to the study findings despite the high attrition rate.

A third limitation is that this study captured only the beliefs of nurses who worked the first or second day shifts. The nurses who worked the night shift were not included in the study because they were not present when data collection occurred. Because of different job responsibilities, nurses who work various shifts may have different job perceptions. This study targeted only nurses who worked the day shifts because they had the most interactions with the patients during daytime hours when the Music First! program took place. Although nurses working the second shift may encounter more challenges than their first-shift peers because of patient sundowning symptoms being more common in the afternoon, differences
between shifts could not be compared as part of this study because data were not available regarding which shifts the nurses worked. Future research should examine the job perceptions of nurses working during different shifts to determine whether the current findings generalize to nursing home staff more broadly.

Finally, the inability to link specific patients with their nurses when examining the relationships between nurses’ job-related perceptions and patient outcomes was a limitation. Additionally, patient data included only those nursing home residents enrolled in the Music First! program. Because of the qualification criteria for inclusion in Music First!, these patients may have had more severe symptoms of dementia than typical patients in each nurse’s care. Collecting more comprehensive data on the patients of each nurse in future studies could help more accurately capture the correlation between patient outcomes and nurse job perceptions. For example, future research could determine whether the number or proportion of severe dementia patients in a nurse’s care affects job-related stress, job-related distress, mood, and job satisfaction, as well as whether programs like Music First! might increase the nurse’s ability to effectively manage a greater number of dementia patients.

In conclusion, these findings suggest that improving nurses’ feelings about their jobs or improving patient outcomes can have a positive impact on the nursing home environment. These improvements can be made through the implementation of programs such as Music First! or through policies that better support nursing staff in their roles. Although these and other types of workplace interventions may be effortful to implement, the time and energy invested may be well worthwhile, given their potential to significantly improve the lives of both the nurses who work in nursing home settings and the patients they care for.
References


Vitelli, S. E., Bolander, H. B., Ladd, K., Johnson, H. M., Brimmer, T., & Lineweaver, T. T. (2018, May). *Personalized music listening improves*
agitation and mood of nursing home residents with dementia
[Presentation]. Association for Psychological Science, San Francisco, CA, United States.


Table 1. Intercorrelations Among Nurse Self-Reports

<table>
<thead>
<tr>
<th></th>
<th>Stress</th>
<th>Global job satisfaction</th>
<th>Fatigue</th>
<th>Emotional reactivity</th>
<th>Stress relief</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stress</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global job</td>
<td>–.392**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>satisfaction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fatigue</td>
<td>.627**</td>
<td>–.449**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emotional</td>
<td>.494**</td>
<td>–.529**</td>
<td>.526**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>reactivity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stress relief</td>
<td>–.265*</td>
<td>.442***</td>
<td>–</td>
<td>–.19</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.455**</td>
</tr>
<tr>
<td>POMS total</td>
<td>.406**</td>
<td>–.565**</td>
<td>.671**</td>
<td>.610**</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.401**</td>
</tr>
</tbody>
</table>

Note. POMS = Profile of Mood States.
*p ≤ .05. **p ≤ .01. ***p ≤ .001.
### Table 2. Correlations Between Nurse Self-Reports and Patient Outcomes

<table>
<thead>
<tr>
<th></th>
<th>CMAI</th>
<th>MMSE</th>
<th>SLUMS</th>
<th>PHQ-9</th>
<th>PHQ-9 frequency</th>
<th>NPI severity</th>
<th>NPI symptom frequency</th>
<th>NPI frequency</th>
<th>Falls</th>
<th>Reportable behaviors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stress</td>
<td>.335**</td>
<td>−.141</td>
<td>−.291*</td>
<td>−.093</td>
<td>−.03</td>
<td>.212</td>
<td>.293*</td>
<td>.146</td>
<td>.036</td>
<td>.463**</td>
</tr>
<tr>
<td>Global job satisfaction</td>
<td>−.162</td>
<td>.275*</td>
<td>.336**</td>
<td>.064</td>
<td>.011</td>
<td>.031</td>
<td>−.236</td>
<td>−.094</td>
<td>−.019</td>
<td>−.337*</td>
</tr>
<tr>
<td>Fatigue</td>
<td>.188</td>
<td>−.271*</td>
<td>−.316*</td>
<td>−.116</td>
<td>−.064</td>
<td>−.02</td>
<td>.216</td>
<td>.026</td>
<td>−.124</td>
<td>.340*</td>
</tr>
<tr>
<td>Emotional reactivity</td>
<td>.248*</td>
<td>−.294*</td>
<td>−.333**</td>
<td>−.288*</td>
<td>−.268*</td>
<td>.073</td>
<td>−.039</td>
<td>−.235</td>
<td>−.145</td>
<td>.121</td>
</tr>
<tr>
<td>Stress relief</td>
<td>−.184</td>
<td>.241</td>
<td>.276*</td>
<td>.244</td>
<td>.213</td>
<td>.15</td>
<td>−.086</td>
<td>.131</td>
<td>.319*</td>
<td>−.322*</td>
</tr>
<tr>
<td>POMS total</td>
<td>−.001</td>
<td>−.208</td>
<td>−.176</td>
<td>−.049</td>
<td>−.016</td>
<td>−.161</td>
<td>.196</td>
<td>−.063</td>
<td>−.094</td>
<td>.05</td>
</tr>
</tbody>
</table>

Note. CMAI = Cohen-Mansfield Agitation Inventory; MMSE = Mini-Mental State Examination; NPI = Neuropsychiatric Inventory; PHQ-9 = Patient Health Questionnaire; POMS = Profile of Mood States; SLUMS = Saint Louis University Mental Status.

*p ≤ .05. **p ≤ .01. ***p ≤ .001.
Table 3. Changes in Nurses’ Job-Related Perceptions During the Implementation of Butler University’s Music First! Intervention

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>Mid-Study</th>
<th>Post-Study</th>
<th>F(2, 18)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stress</strong></td>
<td>2.20 (.894)</td>
<td>1.80 (.616)</td>
<td>2.15 (.587)</td>
<td>3.451</td>
<td>.054</td>
</tr>
<tr>
<td><strong>Global job satisfaction</strong></td>
<td>2.76 (.763)</td>
<td>2.95 (.637)</td>
<td>2.89 (.656)</td>
<td>0.512</td>
<td>.608</td>
</tr>
<tr>
<td><strong>Fatigue</strong></td>
<td>1.84 (.557)</td>
<td>1.87 (.902)</td>
<td>1.98 (.713)</td>
<td>0.679</td>
<td>.519</td>
</tr>
<tr>
<td><strong>Emotional reactivity</strong></td>
<td>1.48 (.663)</td>
<td>1.28 (.645)</td>
<td>1.41 (.611)</td>
<td>1.604</td>
<td>.228</td>
</tr>
<tr>
<td><strong>Stress relief</strong></td>
<td>3.18 (.766)</td>
<td>2.95 (.887)</td>
<td>2.93 (.816)</td>
<td>0.771</td>
<td>.477</td>
</tr>
<tr>
<td><strong>POMS total</strong></td>
<td>0.64 (.589)</td>
<td>0.57 (.642)</td>
<td>0.51 (.581)</td>
<td>0.642</td>
<td>.538</td>
</tr>
</tbody>
</table>

*Note. POMS = Profile of Mood States.*
Abstract

Brown dwarfs, objects within the mass range of 13–80 $M_J$, exhibit similarities with both low-mass stars and giant planets. Despite great progress in both detection and theory in the past few years, the main formation mechanism of brown dwarfs remains unclear. To understand their origin, I tested whether low-mass ($\leq 42.5\ M_J$) and high-mass ($\geq 42.5\ M_J$) brown dwarfs came from distinct populations, as has been claimed in the literature. I collected all the available data on brown dwarfs from the NASA Exoplanet Archive and the Extrasolar Planets Encyclopaedia. Then, using SPSS Software, I performed nonparametric tests to analyze their statistical properties. Finally, I compared the results to the theories of gravitational collapse, core accretion, and disk instability.

Results showed that the two mass groups did not come from distinct populations. In fact, inside the brown-dwarf desert, stellar and planetary processes met and produced almost indistinguishable objects. Despite the hybrid origins of brown dwarfs, disk instability seems to be their main formation mechanism.

Brown dwarfs, starlike objects, are common yet mysterious. The Milky Way contains about 25–100 billion brown dwarfs (Mužić et al., 2017) and approximately 60 billion stars (McMillan, 2011), suggesting that there might be one brown dwarf for every star in the Milky Way—perhaps even more. Despite the frequency of brown dwarfs, their origin is still unknown. Kumar theorized their existence in 1963, but not until 1995 did Oppenheimer and colleagues observe the first brown dwarf. Unlike stars, brown dwarfs are unable to fuse hydrogen (Nakajima et al., 1995), but they do share several characteristics with stars (Luhman, 2012), such as radial velocity dispersions, spatial distributions in young clusters, and outflows.

Because of this overlap, Chabrier et al. (2014) argued that stars and brown dwarfs must have a common origin; still, a clear consensus is lacking in the literature. For instance, Mollière and Mordasini (2012) reasoned that brown dwarfs and planets could form correspondingly. Since 1995, at least seven theories for formation have been proposed, ranging from turbulent fragmentation (Padoan &
Nordlund, 2002) to accretion-ejection (Reipurth & Clarke, 2001); however, because the formation process in each of these theories can be classified as a variation of either a stellar or planetary process, I simplify the scenarios by discussing only the major theory of star formation—gravitational collapse—and the two main theories of planetary formation—core accretion and disk instability—the latter two operating at different distances.

Debate is still underway to determine which of the three is the main mechanism responsible for brown-dwarf formation. The theory of gravitational collapse argues that a molecular cloud, which consists of gas and dust atoms bound together, collapses under its gravity to create a star (Bodenheimer, 2011). On the one hand, in the theory of core accretion, small particles collide to form a solid core that then gathers or accretes mass and thus eventually grows into a planet (Armitage, 2010). On the other hand, in the theory of disk instability, the gaseous disk surrounding a star gravitationally fragments into a planet (Armitage, 2010). While core accretion is associated with close companions, disk instability can effectively produce wide companions (Luhman, 2012). The answer might also be closely related to the brown-dwarf desert—the theorized absence of brown dwarfs that have low orbital periods (≤100 days) within short separations (≤3 AU) of their main-sequence stars (Carmichael et al., 2019)—as the desert is seen as evidence not only of different mechanisms in operation (Maldonado & Villaver, 2017) but also of the boundary between them.

If we succeed in understanding the nature of brown dwarfs, we will improve our knowledge of galaxies and their composition, determine the minimum stellar mass (Pinochet, 2019), and reassess the boundary between stars and planets. To serve this aim, this paper collects all the available data about brown-dwarf binaries (i.e., systems in which a brown dwarf orbits a host star or another brown dwarf) in the NASA Exoplanet Archive and the Extrasolar Planets Encyclopaedia. Next, it divides the sample into low-mass and high-mass groups to test the hypothesis that these, particularly inside the brown-dwarf desert, come from different populations. Then, the results are compared with the theories of gravitational collapse, core accretion, and disk instability. Finally, based on these results, this paper suggests the main formation mechanism of brown dwarfs.

**Methodology**

This paper considers objects within the mass range of 13–80 $M_J$ ($M_J$ is the mass of Jupiter, or $= 1.898 \times 10^{27} \text{ kg}$) as brown dwarfs (Sahlmann et al., 2011). I obtained the brown dwarf data by accessing the NASA Exoplanet Archive and the
Extrasolar Planets Encyclopaedia and then filtering the search for the indicated mass range. This generated 184 results, all checked against the literature. The variables obtained for each object included mass or m*sin(i), in Jupiter masses; eccentricity; orbital period, in days; semimajor axis, in astronomical units; effective temperature, in Kelvin; and the stellar host mass, in solar masses.

I divided the brown dwarfs into low-mass (13–42.5 MJ, N = 142) and high-mass (42.5–80 MJ, N = 42) groups to assess the claim that these objects belong to different populations (Ma & Ge, 2014). Then, I used the SPSS software to perform various tests. I began by checking the normality of the distribution with a Shapiro-Wilk test because of the small sample size. None of the variables were normally distributed (p > .05), so the rest of the tests were nonparametric to avoid the requirement of normality in the data. I next performed the Mann-Whitney U, Kruskal-Wallis H, and Kendall’s tau-b tests, the first two of these tests identifying statistically significant differences between groups, and the third providing an indication of how strongly two variables are monotonously correlated. Lastly, I produced three scatter plots and two bar graphs to summarize these data.

Results

The Mann-Whitney U test exhibited two significant differences between the low-mass (13–42.5 MJ, N = 142) and high-mass (42.5–80 MJ, N = 42) groups (Figure 1). First, the mean rank orbital period was dissimilar (U = 587, p = .029), with values of 54.75 and 38.95, respectively; hence, the low-mass group had a higher mean orbital period compared to the high-mass group. Second, the mean rank temperature varied significantly (U = 278, p = .11), with values of 43.16 and 26.53, respectively. The high-mass group had a lower mean temperature than the low-mass group. Besides these, the other measured variables of the sample—eccentricity, semimajor axis, and stellar mass—did not show a significant difference; therefore, while the low-mass group was related to hot and close objects, the opposite was true for the high-mass group.
Although eccentricity is not statistically different between the low-mass and high-mass groups, I would like to highlight two things. First is the paucity of brown dwarfs with high eccentricities ($\geq 0.60$) within a mass range of 25–55 $M_J$ (solid line,
Figure 2) as well as across all eccentricity values within a narrow mass range of 42.5–49 $M_J$ (dotted line, Figure 2). Second is that a Kendall’s tau-b test revealed a positive medium correlation, between the orbital period and the eccentricity, that was statistically significant ($\tau_b = .318; \ p = .000025$). Figure 3 shows that this correlation holds for both groups.

*Figure 2. Mass Distribution of Brown Dwarfs Depending on Eccentricity*

*Note. The dotted and solid rectangles are the identified gaps of brown dwarfs.*
Figure 3. Positive Correlation Between Orbital Period and Eccentricity

As Figure 4 shows, 10 objects within the brown-dwarf desert were classified: EPIC 219388192 b, WASP-128 b, PSRJ2055+3829 b, 2M1510A b, Kepler-492 b, 2M1510A a, SDSSJ1411+2009 b, EPIC 212036875 b, TOI-503 b, and AD 3116 b. Sixty percent (60%) of these objects were in the low-mass group, and the remaining 40% were in the high-mass group. A Kruskal-Wallis H test revealed that, apart from mass, there is an insignificant difference ($p > 0.05$) in the orbital period, eccentricity, semimajor axis, effective temperature, and stellar mass of the brown dwarfs inside the desert.
Discussion

This study identified two types of brown dwarfs: hot, close, low-mass objects, and cold, distant, high-mass objects. Between these two groups, the orbital period and effective temperature were significantly different, but the eccentricity, semimajor axis, and stellar mass were not. As Figure 5 suggests, the temperature variation could be explained through the dependence of the brown dwarf’s temperature on its distance from the host star (Kutner, 2003); thus, the key difference lies in the orbital period. This pattern was even stronger inside the brown-dwarf desert, where the 10 brown dwarfs, located in the driest region of mass range of $35–55 \, M_J$ (Ma & Ge, 2014), did not show statistical differences in orbital period, eccentricity, semimajor axis, effective temperature, and stellar mass. This study thus rejects the hypothesis that the low-mass and high-mass groups belong to completely distinct populations.
Another similarity between the groups was a positive correlation between the orbital period and eccentricity, with the most distant objects having higher eccentricities. This trend, however, is interrupted in a mass range of 25–55 $M_J$ at high eccentricities ($\geq 0.60$). This can be explained by the fact that lower-mass companions, with high eccentricities, are more likely to be ejected through dynamical interactions (Chabrier et al., 2014) and can be uncommon. Likewise, there is an absence of brown dwarfs across all eccentricity values in a mass range of 42.5–49 $M_J$—without a clear explanation; this would be an interesting region to explore and in which to further test the formation mechanisms. The previously mentioned gaps have not been reported in the literature yet.

The previous findings should be interpreted using the three major models of brown-dwarf formation: gravitational collapse, core accretion, and disk instability. Luhman (2012) highlighted the similarities in radial velocity dispersions, spatial distributions, and outflows between stars and brown dwarfs. The numerical simulations of Bonnell et al. (2008) support the claim that brown dwarfs can form similarly to low-mass stars. Despite this, only 10% of their objects ended up as brown dwarfs. Stamatellos and Herczeg (2015) pointed out another complication in the theory: to collapse, the low-mass core needs to be very dense and compact. Marks and colleagues (2017) stressed that those conditions are rather unlikely to occur, and they rejected that similarity in spatial distributions, between

---

Figure 5. Relationship Between Orbital Periods and Effective Temperatures
stars and brown dwarfs, points toward a common origin, because scattering and ejections also come into play. Consequently, gravitational collapse cannot be the only mechanism involved. Mollière and Mordasini (2012) put forward the idea that core accretion can produce brown dwarfs. This theory can explain brown dwarfs at very close separations (Ma & Ge, 2014) and, to a lesser extent, at wide separations (≥ ~10 AU), through outward migration or scattering (Murray-Clay, 2010). Nevertheless, as the numerical simulations of Stamatellos et al. (2007) have shown, disk instability is more effective at producing brown dwarfs at larger separations. In fact, Kratter and Lodato (2016) argued that disk instability is more strongly linked to brown dwarfs than to planets, provided that fragmentation occurs.

Ma and Ge (2014) first proposed to divide brown dwarfs into low-mass (≤42.5 $M_J$) and high-mass (≥42.5 $M_J$) groups because of the conflicting evidence and plausibility of both the stellar and planetary models. The division was motivated by the brown-dwarf desert, which was thought to be evidence for the presence of two independent processes. Ma and Ge claimed that two distinct brown-dwarf populations existed: the low-mass, formed by a planetary mechanism, and the high-mass, by a stellar one. Nevertheless, Carmichael et al. (2019) pointed out that their first claim, based on a Kolmogorov–Smirnov test, was flawed. Moreover, the detection of brown dwarfs in that region (Grieves et al., 2017; Nowak et al., 2017; Persson et al., 2019) challenged the mere existence of the brown-dwarf desert. There is increasing evidence for the accuracy of their proposed formation mechanism for each mass group, however (Li et al., 2015; Riaz et al., 2018; Rilinger et al., 2019; Stamatellos and Herczeg, 2015). This denotes that the stellar and planetary mechanisms form objects in the brown-dwarf desert range (35 $M_J$ ≤ mass ≤ 55 $M_J$, 0 days ≤ orbital period ≤ 100 days, 0 AU ≤ semimajor axis ≤ 3 AU), at the very least, but, as Maldonado and Villaver (2017) suggested, these processes might operate with different efficiencies.

Because there is no clear division between these two formation mechanisms in nature, we must acknowledge the hybridity of brown dwarfs. Instead of being classified by their formation processes (Carmichael et al., 2019; Luhman, 2012) or on individual bases (Vorobyov, 2006), brown dwarfs should be placed in a continuum where, depending on their mass, they share characteristics with both stars and planets to a greater or lesser extent. This paper supports this idea because of the lack of significant differences between the low- and high-mass groups within the brown-dwarf desert, even though brown dwarfs in that region must have formed by core accretion/disk instability and gravitational collapse, respectively. From this, it is clear that stellar and planetary mechanisms produce almost indistinguishable
objects in the region of the brown-dwarf desert. Only when we compare objects that are far from each other in this continuum—with a mass difference of $\geq 30 \, M_J$—do significant differences in the orbital period and effective temperature emerge. This raises interesting questions about the boundary between stars and planets. If objects were defined by their formation mechanisms, the boundary between stars and planets would disappear. If we instead acknowledge the hybrid origins of brown dwarfs and define objects depending on their inherent physical properties, we can make a distinction among stars, brown dwarfs, and planets.

Brown dwarfs sit in a continuum between stars and planets, although disk instability seems to be their main formation mechanism. The great majority (77%) of the sample belongs to the low-mass group, which shares a common origin with planets, of which 46% most likely formed through core accretion ($\leq 3$ AU) and 54% probably through disk instability ($\geq 3$ AU). Moreover, gravitational collapse cannot explain the brown dwarfs with very close separations ($\leq 3$ AU), which represent 45% of the total sample, as a system formed in this way would quickly migrate inward and merge with the star (Ma & Ge, 2014). Formation mechanisms might not be mutually exclusive (Rilinger et al., 2019), but gravitational collapse should be relegated to a secondary role.

Nonetheless, these results should be interpreted by taking into account three major limitations. First, although the sample size of this study ($N = 184$) is considerably larger than usually seen in the literature ($N = 65$; Ma & Ge 2014), it remains small because the population of brown dwarfs in our galaxy was estimated to be 25–100 billion (Mužić et al., 2017). For this reason, the study considered only nonparametric statistical tests although they are less sensitive, causing small effects to go undetected (Garth, 2008). Second, the two main detection methods—radial velocity and transit—are biased toward observing objects in close orbits (Planetary Society 2002a, 2002b), meaning that wide-orbit objects are underrepresented. Finally, there is a fundamental limit in our capability of inferring from current values the exact primordial conditions and their relationship to formation mechanisms (Chabrier et al., 2014)—for example, due to dynamical interactions (i.e., mergers, ejections, migrations).

Independent studies are required to confirm the statistical difference in the orbital period and effective temperature between the low- and high-mass groups, as well as the homogeneity of brown dwarfs inhabiting the brown-dwarf desert. Additionally, more observational data of brown dwarfs are needed to gather a more robust sample. We can expect major contributions in this area from the upcoming James Webb Space Telescope, Atmospheric Remote-sensing Infrared Exoplanet
Large-survey, and Wide Field Infrared Survey Telescope space missions. As we obtain more data, particularly of high-mass brown dwarfs, the hypothesis that disk instability is the main formation mechanism of the overall population will be tested. Finally, to understand the positive correlation between orbital period and eccentricity, the newly identified gaps need to be explored, including a mass range of 25–55 $M_J$ at high eccentricities ($\geq 0.60$) and across all eccentricity values in a mass range of 42.5–49 $M_J$.

**Summary and Conclusions**

Determining the main formation mechanism of brown dwarfs is possible only by comparing and contrasting the brown dwarfs’ statistical properties with theory. For this paper, I collected all the data on brown-dwarf binaries, available in the NASA Exoplanet Archive and the Extrasolar Planets Encyclopaedia, and divided the sample into low-mass ($\leq 42.5 \, M_J$) and high-mass ($\geq 42.5 \, M_J$) groups to test if they came from the same population. I discovered that the only significant differences between these groups were the orbital period and effective temperature, although the latter is dependent on the distance of the object from the host star. I also classified 10 objects inside the brown-dwarf desert that came, almost equitably, from the high-mass and low-mass groups. I not only challenge the existence of the desert but also prove that, besides mass, all the other properties—orbital period, eccentricity, semimajor axis, and stellar mass—show no significant difference. Because the low-mass group is associated with core accretion and disk instability, and the high-mass group with gravitational collapse, this result suggests that the stellar and planetary processes meet in nature and produce almost indistinguishable objects. This paper therefore calls for the recognition of the hybrid nature of brown dwarfs.
Works Cited


Abstract

This paper aims to present Ottoman, Arab, and British dynamics in the Arabian Peninsula. The paper highlights British exploitation of political circumstances to gain presence in Iraq and the Gulf (more particularly Kuwait) throughout the late 19th and early 20th centuries. As pan-Arab movements began to gain traction within Arab territories during the final years of the Ottoman Empire, Gulf tribes sought to move away from Ottoman influence and to establish greater sovereignty, control, and border integrity. This desire for Arab tribes to form independent polities was utilized by the British to competitively gain strategic presence against France, the Dutch, and Russia in the Arabian Peninsula and the Middle East on a wider scale for various reasons.

This paper first covers internal conditions within the Ottoman Empire, as well as Ottoman-Arab relations. Second, utilizing historical texts, the paper specifies the nature of British-Arab relations to lay the historical contextual framework for shifting sentiments in the Gulf as the Ottoman Empire weakened. The paper then moves on to explicate geopolitical factors for continuous border disputes between Kuwait and its northern neighbor, Iraq, which eventually culminated in the 1990 Iraqi invasion. Secondary sources are consulted to present the consequences of British involvement in heightening tensions in Iraq and, vicariously, Kuwait due to the catalyst effect of their diplomatic missions and establishment of protectorates parallel to growing pan-Arab sentiment. Finally, the events discussed are reviewed through power-balancing theory to assess the effects of political decisions on both Iraq and Kuwait.

Keywords: Pan-Arabism, tribal relations, British diplomacy, Ottoman Empire, power balancing, Arabian Gulf

In the political climate of the 19th and 20th centuries, the Ottoman Empire found itself a waning empire surrounded by European powers who intended to
increase their influence in a highly competitive environment fueled by imperial expansion. In a more magnified version of urban sprawl, the Ottomans struggled to secure the integrity and unity of their more distant provinces into the Ottoman Empire politically and nationally.

By the 19th century and after tumultuous periods of reform, the Ottoman Empire’s frontier extended to North Africa, Mesopotamia, the Balkans, and parts of the Arabian Peninsula. With European powers such as France, Russia, and England encroaching on the Ottoman Empire, Abdulhamid II was faced with the need to be careful with his alliances. Sultan Abdulhamid II’s willingness to cooperate with the British soon fueled the disgruntled to speak out with anti-British sentiment (Tallon, 2019). Eventually, the Young Turks (primarily the Committee of Union and Progress) deposed Sultan Abdulhamid II in 1908 (Al-Hamdi, 2015). This deposition left the Young Turks with the burden of further securing Ottoman borders through various agreements and increased military presence around some borders (Tallon, 2019). Not long after, however, an Ottoman–Balkan war throughout 1912 and 1913 stripped the Ottoman Empire of its last remaining territories in Europe (Tallon, 2019). Along with all of this tension, throughout the late 19th and early 20th centuries in the Ottoman Arab provinces, particularly aggressively in the “fertile crescent,” an opposition movement—Pan-Arabism—had been brewing (Dawn, 1988).

The growing pan-Arab movement proved to be a tool for opposing imperial powers to exploit, especially the British. With a second government stationed in India, the British were highly involved in attempting to shift the balance of power in the Arabian Peninsula to further secure their position in India and to deter other European powers from establishing spheres of influence with the Arab peoples. This led the British to utilize tactics with tribal leaders of both Saudi Arabia and Kuwait that they had previously used in India, as well as to interfere with Baghdad’s and Basra’s finely tuned political structure, leading to various consequences. Although British intervention in supporting Arab movements during the 1880s and throughout World War I assisted in achieving autonomy from the Ottomans, it fueled bilateral and internal conflicts in the post-WWI period, eventually setting the foundation for the 1990 Iraqi invasion of Kuwait.

The Ottomans under Sultan Abdulhamid II and after the Young Turk Revolution, however, were privy to the European interest in the more distant provinces such as those in present-day Iraq and the Gulf. Furthermore, the Ottoman administration recognized vulnerability to European powers (Çetinsaya, 2003). For instance, various governors, ambassadors, and viziers reported a British threat in

186
Iraq and in the Gulf, with particular regard to Kuwait. The emboldened efforts of the British to approach ambitious Arab leaders were attributed to a lack of a “forward policy” and to local Ottoman officials being “neutral bystanders” to Britain’s actions in the region throughout the 1880s (Çetinsaya, 2003, p. 199).

Further, ambassadors noticed the “neglect” of residents in Arab provinces in what is now Iraq, as Ottoman officials considered those residents to be “ignorant and uncivilized” (Çetinsaya, 2003, p. 201). Because of this “incivility,” Ottoman administrators feared a conquest of Basra beginning with British control and use of Kuwait’s harbors (Çetinsaya, 2003). Because of this and other factors such as instability in the vilayets of Basra, Mosul, and Baghdad, in the years after the 1878 Berlin Treaty, the Ottomans sought ways to pull in the periphery more aggressively, by planning for the Berlin–Baghdad railway, increasing naval presence, and implementing policies on the ground in a rather unwelcomed process of “Turkification” in an attempt to re-regularize administration empire-wide (Simon, 2004, p. 40).

The Arab Political Climate

Overtures by the British in Kuwait, Qatar, Yemen, and Bahrain were not isolated events. As British authorities (more specifically, the government of India) recognized a potential to exploit tensions in the region throughout the 19th century and the interwar years, a “scramble for Arabia” was instigated (Tallon, 2019, p. 98). With the Ottoman Empire turning its attention to more direct rule in Mesopotamia after the pacification of Mohamad Ali in 1841, discontent in local communities began to brew (Blumi, 2012). As Arabs began to desire more autonomy, the pan-Arab movement mobilized. Additionally, the increasingly

---

1 “Turkification” refers to the systematic process of limiting expression of variance in ethnicity, language, and religion within Ottoman territories in order to enforce Ottoman uniformity. The unionist ideology of the Young Turks paid little attention to diversity, as Ottomanism was largely synonymous with being a Muslim Ottoman Turk. This ideology led to underrepresentation of Arabs and other ethnic groups in the Ottoman Parliament. Because Arab identity was based predominantly on religion and language, Arabs felt increasingly alienated by the Young Turk administration imposing Turkish language policy, for instance. Further, Muslim Arabs associated Turkification with the phasing out of Islam from society, particularly because of the irreligious character of non-Arabic-speaking Ottoman government officials assigned to administrate their areas. The disruption of traditional social and political norms in Arab territories eventually encouraged Muslim Arabs to form literary organizations (e.g. al-Muntada al-Adabi) rejecting Turkification, in an attempt to reinforce their version of group identity. Consequently, as Arabs found their cultural, political, and social rights continually suppressed, an Arab Revolt ensued (Kayali, 1997).
aggressive racial Turkification policies of the Young Turks further urged and fueled separationist, nationalist ideas in provinces such as Basra, eventually culminating in the 1914 Great Arab Revolt (Simon, 2004). Pan-Arab sentiments occurred most prominently in tangible form within Iraq and Syria (Dawn, 1988), where publications of pan-Arab and nationalist thinkers were used in schools and were circulated heavily. In comparison, the more tribal dynamics of the Arabian Peninsula manifested these sentiments in expansionist strategies by various tribes, such as Al Saud.

Writers such as al-Shurayqi, al-Khatib, and Miqdadi explored common themes of the Arab nation being a “living body” (Dawn, 1988, p. 69) and of Arabs as being the last of a series of Semitic migrations (i.e., Semitic wave theory), making them “heirs of the Semites” (Dawn, 1988, p. 70).

With regard to external powers, pan-Arab/nationalist texts often contrasted the “noble” Arabs to the exploitative, economically driven, “hateful” Persians and “innately mean” Europeans (Dawn, 1988, p. 76). The texts also regarded the trade routes going through Arab territories to India as a way for external powers to weaken Arabs in the face of the Ottomans (Wagner, 2015). Despite this idea—which later set the groundwork for the Ba’athist and Nasserist regimes—there was an ideological dissonance with pan-Arabism, as Arabs were still willing to deal with external powers to consolidate autonomy and greater agency from the Ottomans (Dawn, 1988).

Furthermore, Arab movements also had a religious dimension represented in Pan-Islamism: ideas of an Arab caliphate that would preserve Arab society from the “second Jahiliyya” of Sufism (Dawn, 1988, p. 74), a religious sect highly associated with Ottoman tradition. This was showcased with the increasing control of Al Saud over Mecca and Medina through their use of the Wahhabis and Ikhwan to consolidate influence in frontiers surrounding them, which threatened Ottoman religious legitimacy and hajj revenues (Tallon, 2019). This later led to skirmishes between Al Saud and the Ottoman-backed Ibn Rashid. Other figures began to harness the Ottomans’ crumbling religious legitimacy as well, including Husayn ibn Ali (sharif of Mecca), who called for a separate Arab Muslim state for himself (Yaphe, 2004).

Later in the post-WWI period, however, the once-quelled conflicts among movements in Arab territories gained even more traction; the pan-Islamists and nationalists were in almost constant conflict. The British were quick to attempt to take advantage of the tumult without properly understanding the complexities of
the interconnections of these movements, leading to major instability in areas they have occupied in what later became Iraq.

**British Policy and the Arabs**

Although the British did not fully comprehend political and social intricacies in regions such as the vilayets of what later became Iraq, early endeavors to explore the Gulf and surrounding areas navally in the 1700s and 1800s, mail links, and the presence of political residents allowed for the British administration to recognize the potential for a minimally challenged strategic encroachment into the area (Simon, 2004). Serious consideration for mobilizing more direct control in the Gulf by the British region began in 1913, however, when Ottoman neglect of the area allowed piracy, arms sales, slavery, and German presence to go unchecked, according to the British, threatening future British hegemony in the Gulf (Simon, 2004). Additionally, this was an opportune time to increase presence in Ottoman territories in the Gulf, as the war-ravaged Ottoman Empire was focused elsewhere after the yearlong Balkan conflict mentioned previously.

As World War I wore on, Britain surveyed the general situation of the region and consequently also surveyed societal divisions that it might utilize accordingly. According to a 1913 British governmental document, British policy was multidimensional and included with regard to “Arabia” establishment of protectorates in Southern Syria and Mesopotamia, a protectorate in Kuwait, recognition of the Hijaz as independent, encouragement of an Arab caliphate led by the sharif in Hijaz, and a coordinated military effort with other allied European powers to quell unrest. Furthermore, the press would be used to politicize Sunni Islam further throughout other areas in “Arabia” and even India. These plans would also serve to put pressure against the unstable Young Turk administration from within (Hurewtiz, 1914; Wagner, 2015). These policies aimed to achieve a concrete severance of aspiring Arab leaders from the Ottoman sphere of influence while also further divorcing religious credibility heralded by the Ottoman sultan as protector of Mecca and Medina.

Furthermore, the British realized that to garner Arab support, their propositions would have to be framed to appeal to existing movements (primarily Pan-Islamism and Arab nationalism/Pan-Arabism, as previously mentioned). In a 1914 British governmental proclamation to “the natives of Arabia and the Arab provinces,” Britain denied a desire to conquer, possess, occupy, or protect any Arab territories (Burdett, 1998, p. 99). Further, the text heavily references God and refers to the Ottomans as “Turks” who have “laid upon” the Arabs a “heavy burden” that
the English would “cast” with “God’s help” (Burdett, 1998, p. 100). In another proclamation a year later, the British maintained the same rhetoric, insisting on complete independence and that “please God, [the lands of Arabia] return along the paths of freedom to their ancient prosperity” (Burdett, 1998, p. 101), echoing the sentiments of Pan-Arabists and Pan-Islamists as they appealed to them.

The British consequently increased their presence “on the ground” in the Arabian Peninsula and in vilayets that later made up Iraq, attempting more “hands-on” policies. Britain took on an active role in resolving both tribal and political conflict in the case of Kuwait, for instance. Additionally, the extravagance of the buildings that housed British representatives became more prominent throughout the 20th century, symbolizing the growing British influence in the area and the local acceptance of their authority (Muir, 2008).

Real alarm in the Ottoman Empire was heightened, however, after what was termed the British–Ottoman confrontation over Kuwait, spanning from 1896, with Muhammad Al-Sabah’s assassination, to 1904. The Ottomans considered British policy with the Arab provinces and areas such as Kuwait as another English “civilizing” mission like that with Egypt in the past (Çetinsaya, 2003). Eventually, the warnings and intuitions of the various Ottoman officials concerning the Arab provinces and the Gulf were realized, as Kuwait began to pull away from the Ottoman sphere of influence more definitively and ultimately succeeded in doing so—meaning that Ottoman policies to control dissent were not enough.

Kuwait

Geographically, Kuwait (“Grane,” as termed by the British) was situated south of the Ottoman province of Basra, strategically bordering Shatt al-Arab, which made it a valuable land- and sea trade route for the Ottoman Empire and a significant port between Basra and Bahrain, where Indian ships came to stop (Muir, 2008). Relatedly, Kuwait was a municipality in the Ottoman Basra province. Kuwait’s location eventually made it a point of contention as imperial powers competed to absorb it into their spheres of influence. Kuwait exploited its location to create a commercial-friendly environment, even taking a toll on caravans that passed through it to transport supplies and weaponry to the Ottoman provinces of Damascus and Aleppo, eventually even rivaling Basra as a “trans-desert route” (Muir, 2008, p. 171). The flow of weapons from and to the area was especially problematic for the Ottomans and, later on, the British, as it gave rebel forces means of battle and altered the status quo (Blumi, 2012). This exploitation of movement was further optimized with Mubarak Al-Sabah’s increasingly strong ties with Al
Saud and rivalry with Ibn Rashid (Blumi, 2012). Additionally, new trade routes were established in the mid-1800s, excluding major merchants in the area and British correspondents in nearby Bushire, giving Kuwait further leverage as a trading post (Blumi, 2012).

The importance of Kuwait’s location of course did not go unnoticed by larger powers. As the Ottoman Empire sought to invigorate its suffering economy and link its provinces to the Anatolian metropole, Kuwait became the proposed end of the planned Berlin–Baghdad railway (Blumi, 2012). The proposed railway would dually attract European powers to its locality by the late 1800s in an attempt to secure a foothold and more favorable economic shares, as well as cause Kuwait to seek integrity even more aggressively (Rush, 1991a).

**Beginnings of Kuwaiti–British Relations**

Before the factors and events that led to Kuwait distancing itself from Ottoman control are presented, context regarding this Ottoman–Kuwaiti relationship must be detailed. As mentioned above, Kuwait’s geographic location provided a political environment that would later allow Kuwait to propel itself further from sole Ottoman influence through strategic balancing of major powers and local alliances. Historically, however, Kuwait had always enjoyed a degree of autonomy compared to areas in Ottoman Mesopotamia. In an administrative report by the local British agency in Kuwait, contextual political and historical information were provided to the British metropole. The document detailed that the sheikh was regarded a de facto ruler, but it documented a de jure governor by the Porte in the region (Archive Editions, 2020). This distinction would later allow the British to surpass higher powers in the Ottoman “chain of command” to deal with the Kuwaiti sheikh directly in order to manipulate this Ottoman “communications problem” when the right time came to sway Kuwait into their sphere of influence (Kumar, 1962, p. 71).

Although the Kuwaiti sheikh Muhammad Al-Sabah embraced Ottoman relations, he was eventually assassinated by his ambitious half-brother, Mubarak Al-Sabah, in 1896. With Mubarak in power, realizations of Kuwait pulling away from Ottoman influence began. During his reign, Mubarak ran various military campaigns to extend his tribal reach further into the Arabian Peninsula (Archive Editions, 2020). This is also because his coup did not go completely unchallenged, as the late Muhammad Al-Sabah’s allies, such as Yousef Al-Ibrahim, attempted to counter Mubarak’s expansion militarily for more than two decades (Blumi, 2012).
As Mubarak sought to fortify himself with regard to tribal alliances and the seeking of British help, Ottoman officials aimed to prevent the British from creating a toehold in Kuwait and saw the necessity of Mubarak’s “ejection” (Çetinsaya, 2003, p. 201). Realizing that Kuwait was too autonomous for the Ottoman Empire’s integrity, officials planned to incorporate Kuwait in a “reconstitution” of Basra into a single kaza more effectively (Çetinsaya, 2003, p. 201). By 1899, Kuwait had secretly struck a deal with the British to guarantee its protection from Ottoman forces, effectively severing it from Basra (Rush, 1991a). Despite previous attempts by Mubarak to secure an agreement with the British, the British finally considered dealing with Kuwait as they realized the Russian and German interests in the land.

Furthermore, once Curzon had assumed position as viceroy of India in 1898, he had insisted that Kuwait was instrumentally important to British interests with regard to India, trade routes, and ports in the Gulf (Muir, 2008; Pillai & Kumar, 1962). This British interest in Kuwait is undoubtedly also linked to the fact that Kuwait was to be the endpoint of the Berlin–Baghdad railway, which pushed Britain to interfere with construction indirectly to increase its bargaining power against Germany. In this way, the British would ensure greater trade benefits once the Berlin–Baghdad railway was completed, especially that the tracks would lie precariously close to English–Iranian oil fields (Kumar, 1962).

In 1903, with Curzon’s visit there, Kuwait became a de facto British protectorate, especially as Curzon recognized the threat of Kuwait falling back into Ottoman influence if Al Saud would be on the losing side of tribal spats with the Ottoman-backed Ibn Rashid (Al-Hamdi, 2015). By 1904, the British government tasked political resident Knox to monitor the Khor Abdallah as well as relations between Al Saud and other tribes in the vicinity. Knox’s reports had the dual purpose of protecting trade interests and maintaining the status quo that marked the official British presence in Kuwait (Muir, 2008). Consequently, Mubarak exploited the competition of foreign powers in gaining access to Kuwait to achieve his own ends.

Despite the secret British deal and the rather autonomous dealings of the Kuwaiti sheikh, however, Kuwait continued accepting Ottoman titles, providing tax to the Ottoman metropole, receiving revenues from Ottoman-provisioned date farms to the north, and even providing aid for the 1912–1913 Ottoman–Balkan War (Rush, 1991a). This was to perhaps prevent open conflict, along with the added benefit of securing continuous and significant revenues from the date farms in a power-balancing act.
In 1913, however, the Anglo–Ottoman convention was signed. The convention delineated the rather previously ambiguous northern borders between Basra and Kuwait, granted Kuwait the islands of Bubiyan and Warbah, and declared Kuwait autonomous but under “Ottoman suzerainty” (Pillai & Kumar, 1962, p. 118). During World War I, Anglo–Ottoman correspondences continued to detail Kuwait’s cession diplomatically. True recognition of the Kuwaiti state by the British Empire eventually occurred in 1914, with the outbreak of WWI, when Kuwait symbolically adopted a red flag with the Arabic word for its name (Rush, 1991b). Consequently, Britain achieved its goal in establishing a protectorate in Kuwait, as mentioned previously.

British interests in Kuwait continuously morphed throughout the 20th century as Kuwait went from being a key port and land post to a source of oil with the establishment of the KOC (Muir, 2008), meaning that Kuwait would have British support until its official independence. Kuwait’s nationalist endeavors to maintain its autonomy and the integrity of its borders continued until formal independence in 1961 in the face of Al Saud and Iraqi efforts, and even then, tensions remained with Iraq, later culminating in the 1990 Iraqi invasion.

Mosul, Baghdad, and Basra

The discussion of imperial relations with Ottoman Basra and, later on, Iraq better explicates the border disputes between Kuwait and its neighbor throughout the 20th century.

Contrary to their rather clear-cut dealings with territories and tribes in the Gulf, British representatives found frequent difficulties in establishing a presence in the Basra province because of a far greater Ottoman grip represented by the regional administration’s jurisdiction and normative societal structures that were more complex compared to the very tribal relations in the Gulf (Clark, 2008). Additionally, Ottoman Basra represented a more highly multifaceted society, with deep variations in culture, religion, language, and local political affiliation (Simon, 2004).

The vilayets of Basra, Baghdad, and Mosul came into Ottoman control under the reign of Sultan Suleyman I in 1534 (Yaphe, 2004). As discussed earlier, with regard to the importance of Kuwait’s geographic location, the vilayets were also instrumental to the Ottoman Empire because they were hubs for trade routes. Additionally, the vilayets were a separating frontier between the Ottomans and the Persians, which made them particularly challenging to manage and maintain long
before European powers showed interest. These tensions were a result of the multireligious and multiethnic nature of the vilayets’ residents, who found similarities with their Persian neighbors, especially in Baghdad (Simon, 2004). This diversity led to frequent revolts by Shi’i sects, which the Persians used to their advantage. Arab Lakhnids and Ghassanids, Kurds, and Eastern Christians also represented factions in society that neighboring empires utilized to incite unrest, particularly throughout the 18th century (Simon, 2004). Social and political disturbance became a feature of the area throughout World War I and well into the late 20th century, especially in areas such as Kerbala and Baghdad (Yaphe, 2004).

The British and Iraq

Prior to British occupation of Iraq in 1917 and the establishment of military rule, the British had entered Ottoman Basra through the British Mesopotamian Expeditionary Force in 1914 (Al-Hamdi, 2015; Simon, 2004). At the time, the British had identified the desire of local Arab Ottoman associates to become autonomous, albeit within the Ottoman system (Yaphe, 2004). Prior to World War I, the British had accumulated intelligence to infer certain dynamics within the Iraqi provinces among the various religious, nationalist, and ethnic movements. As the British began to settle in the recently conquered Iraq, however, their perceptions were found to be largely false; movements became more heterogenous and began to clash more violently and frequently, as mentioned previously (Wagner, 2015).

Policies to “civilize” the Arabs in Iraq into an independent state after World War I were deliberated rather haphazardly and chaotically by several British governmental entities: the War Office, the Foreign Office, the Arab Bureau, and the India Office (Yaphe, 2004). In the vein of their methods of Indian rule, the British attempted to create cleavages in Iraqi society by giving tribal leaders considerable agency with tax collection and dispute resolution based on tribal customs and land ownership, for instance, which had been largely absent during Ottoman rule. This favoring of tribal provincial rule was an attempt to counter the increasing influence of city folk, who had overwhelmingly nationalist ideas (Yaphe, 2004); however, because Britain was implementing a one-size-fits-all strategy in Iraq, domestic stability soon evaporated and tribal sheikhs became increasingly authoritarian and brutal, thwarting political development (Yaphe, 2004).

Unrest became especially pronounced after World War I, and Foreign Office officials sensed nationalist sentiment growing at an “unstoppable momentum” (Simon, 2004, p. 36). Not long after, the 1920 revolt began and was a
unifying event, joining the multiple religious sects and residents of differing socioeconomic status to combat hypertaxation, unemployment due to the appointment of British officials in the British Civil Administration, and even calls for an independent Arab Islamic state (Wagner, 2015).

After quelling the rebellion by force, the British government worked to reconsolidate control by abolishing military rule and setting up a Western-inspired constitution in 1921, promising representation, checked power, and democracy (Wagner, 2015). Following this came British deliberations on whom to appoint, these being easily manipulated, pliable leaders and officials. As a result, an inexperienced Sunni religious official leader was chosen, as well as a Hashemite Arab, King Faysal I, who had no particular affiliation with any faction of Iraqi society at the time (Simon, 2004). The result was something termed an institutional façade (Yaphe, 2004, p. 33), a method for more indirect British control of Iraq.

Now with an established government, treaties, such as the binding 1923 Treaty of Lausanne, detailed the formation of the country of Iraq in the joining of the three vilayets of Basra, Baghdad, and Mosul—with no mention of Kuwait (Pillai & Kumar, 1962). Upon Iraqi independence in 1932, another frontier-management agreement was drawn, with Sir Percy Cox issuing a memorandum with Prime Minister al-Askari, to detail Kuwaiti–Iraqi borders once more based on the 1913 Anglo–Ottoman convention (Al-Hamdi, 2015), reiterating its legitimacy despite its ratification being previously interrupted by World War I.

Once a very Western-influenced and maladjusted Iraqi government was in place, certain dynamics began to play out both domestically and bilaterally, with Kuwait, throughout the 20th century.

**Kuwait and Iraq**

As discussed previously, Pan-Arab/nationalist movements continued with fervor even after the British had assigned an Iraqi constitutional government (Al-Hamdi, 2015). After Faysal I’s death following Iraqi independence, however, a very staunchly nationalist Ghazi I was king and was highly maximalist in his diplomatic stance. To disseminate his ideology broadly, Ghazi I ran a radio station; propaganda was rampant in the press. Using his media channels, Ghazi I illustrated ideas such as that Kuwait was an illegitimate state and was instead part of Iraq. To support this view, he cited Ottoman-era relations despite border delineations spanning back to 1913 (Pillai & Kumar, 1962). By 1938, Ghazi I’s claims reached their height; a
military attack was planned, but when he unexpectedly died, the plan was abandoned (Muir, 2008).

The advocacy of nationalist and pan-Arab sentiment endured with politicians as well. Prime Minister al-Said, under Faysal II, promoted secular pan-Arab sentiments and threatened to annex Kuwaiti islands when Kuwait refused to join a Hashemite union (i.e., AHU) to contribute funds (Al-Hamdi, 2015). Additionally, al-Said’s fervent desire for Kuwait to join a Hashemite union might have been an implicit diplomatic gesture to imply that Kuwait belonged within Iraqi borders.

As the monarchy was overthrown in 1958 and a republic was established under Qasim, however, nationalistic policies and diplomacy increased in aggressiveness and frequency, with Qasim refusing to acknowledge Kuwaiti independence in 1961, publicly stating that Kuwait was “an integral part of Iraq” (Clark, 2008, p. 9). At this time, Kuwait had already established its position in the international arena—at times even with help from the Iraqi government to join international organizations—and had considerable influence on the British market economy (Muir, 2008). This implicit threat by Qasim led to Kuwait requesting British assistance preemptively, as well as a UNSC meeting being held. Iraq took this opportunity to harness growing anti-British sentiment among Arab nations, but it received little support from fellow Arab nations in opposing Kuwait (Al-Hamdi, 2015). As Qasim’s successor, Arif of the Ba’athist regime, continued to push the claim on Kuwait, Iraq suffered domestic political unrest and a significant loss of diplomatic ties (Muir, 2008).

Claims to Kuwait were based on historical Ottoman links, despite the technical autonomy of Kuwait during that time and despite conventions detailing borders, as mentioned previously. This was in part because Kuwait could provide Iraq closer access to the Gulf and, after the discovery of oil, to more than 15% of the world’s oil reserves (Pillai & Kumar, 1962). Furthermore, Iraq had experienced tensions and border disputes with its neighbor Iran spanning from the days of Ottoman rule, eventually leading to an economically devastating war with Iran under Saddam Hussein (Yaphe, 2004). The Iraq–Iran War brought an already politically turbulent and economically feeble Iraq to its knees, instigating the 1990 Iraqi invasion of Kuwait to reconstitute losses on the grounds of reclaiming the “rogue” state of Kuwait back into Iraqi control (Simon, 2004).
Power Balancing: Consequences and Effects

Ilai Saltzman’s 2011 book, *Securitizing Balance of Power Theory: A Polymorphic Reconceptualization*, reimagines common theories and strategies of power balancing in international relations—such as soft- and hard balancing, buck-passing, and bandwagoning—in a revised, multifaceted framework with security at its core. Relatedly, Randall Schweller’s 2006 book, *Unanswered Threats: Political Constraints on the Balance of Power*, explores the causes for underbalancing in terms of both domestic and international threats by discussing elements such as social/elite cohesion and regime vulnerability. In this section, the theories presented in the two books will be used to discuss and tie in the factors that led to Britain gaining influence in the Ottoman Gulf and contributing to Iraqi domestic unrest, which encouraged border disputes and later, the 1990 Iraqi invasion.

The Ottomans

In the context of the Ottoman Empire in the mid- to late 1800s, Schweller’s ideas of underbalancing are very much relevant, as mentioned earlier, with the complaints of Ottoman officials regarding a lack of “forward” policies, along with the Ottoman policy neglecting the British presence. What prompts a state to delay reactive policy toward governmental threats or concerns? According to Schweller (2006), incoherent policies result from factors such as a lack of elite consensus, governmental regime vulnerability, and threats to social cohesion. Additionally, an important fact to consider is that decisions by individual policy makers occur after assessments of perceived threat rather than of what is, which could lead to mishaps in deciding power-balancing policies (Saltzman, 2011). All of the previous factors have been exhibited, with the rise of Pan-Arabism and other social divisions in Ottoman provinces, the threat of European powers, and the tumultuous reforms in Ottoman administration, all discussed previously, leading to what Schweller termed policy paralysis, which is caused by a weakened government generally having less policy capacity (2006, p. 57). Furthermore, a compromised sense of governmental legitimacy as a result of fragmented social cohesion would limit a state’s options to enforce hard-balancing policies for fear of antigovernmental action in response. Relatedly, limited domestic social cohesion often means that outsider threats do not have the usual effect of increasing cohesion but rather disband ingroups further, as with Iran and Shi’i groups in Basra, for instance.

Jack Snyder, however, has also suggested that instability in a state’s regime would bring about another reaction in policy making: overexpansion (as cited in
Schweller, 2006). Overexpansion was what the Young Turks attempted after their assumption of power in the years before World War I. It should be noted that overexpansion could refer to both spatial expansion and expansion of executive power. Synder identifies two elements to a government’s decision to implement overexpansion (which must not be confused with overbalancing): a weak central authority and several concentrated interest groups (as cited in Schweller, 2006). In the Ottoman Empire under the Young Turks came the continual development of pan-Arab sentiment, conflict among the Young Turks themselves, and conflict between Ottoman liberals and the Young Turk administration (Hurewtiz, 1914), which led to policies that Arabs viewed as increasingly antagonizing (e.g., tighter military and naval presences, and Turkification policies), eventually encouraging Arab leaders to seek soft-balancing policies with European powers (e.g., Kuwait’s agreement with the British).

The British

Although the Ottomans in the 19th century and early 20th century did not engage in overbalancing with regard to Arab provinces, the British in many ways did so as they gained more control of the region through the early and mid-20th century. Along with overbalancing both during and after World War I, the British initially implemented various soft-balancing techniques to gain a larger presence in the Gulf in their quest to establish protectorates in Kuwait, Qatar, Bahrain, etc.

Although still a confrontational method of power balancing, soft balancing often does not feature open conflict. Instead, soft balancing can include diplomatic maneuvering, nonmilitaristic policies, and institutional binding, and hard balancing may be only a plan B to avoid widening conflict within the current balance of power (Saltzman, 2011). In this instance, one example is the British cooperating with the administrations of Sultan Abdulhamid II and the Young Turks while dealing with local leaders in the Arabian Peninsula to secure personal interests and undermine the Ottoman Empire from within. A prime example, of course, is the previously discussed British deal with Kuwait, in which Kuwait also played an expertly tuned game of soft balancing with the Ottomans and rivaling European powers. Although the British’s soft-balancing measures were sound in upholding British interests with regard to the political climate in the Arabian Peninsula and the insufficient, underbalanced Ottoman response, the British began to significantly misinterpret and misperceive the status quo as time wore on. In decision-making, state officials must judge the level of threat domestically and internationally, as well as determine the “resource extraction potential” at their disposal (Saltzman, 2011, p. 33)—that
is, how capable their state is politically, militarily, diplomatically, and economically to undertake certain policies. On one hand, it can be assumed that by 1917, as the British moved into Ottoman Mesopotamia, British officials considered their resource-extraction potential high. This perception was due to their successes in the Gulf and their advantage in World War I. On the other hand, the threat of the occupied society dissenting is also perceived as high because of British militaristic rule and drastic policies aiming to dramatically assert British control in an effort to occupy the power vacuum left by the Ottomans. Snyder regards overbalancing policies as both very costly and very likely to take away privileges from various factions in society (as cited in Schweller, 2006). That being said, the overbalancing policies of the British in what later became Iraq repressed significant portions of society and highly privileged a select few, fueling dissent and instability as the public refused the new balance of power that the British were attempting to instill, and the British left a maladjusted, volatile Iraqi political environment in their wake as they retreated by 1932.

Conclusion

As demonstrated in the previous section, it is important to note the crucial connection between domestic politics and external policies throughout the discussions of Ottoman policy (or lack thereof), Kuwaiti alliances, and British presence in the Gulf. That is to say, domestic shifts in power have all led to definitive events connected to Kuwaiti independence and Kuwaiti–Iraqi border conflicts: Muhammad Al-Sabah’s assassination, Curzon’s assumption of the position as viceroy of India, Pan-Arabism’s gaining of support, and the Ottoman Empire’s suffering through various unsuccessful reforms.

Building on the Ottoman Empire’s internal political tumult, it becomes clear that the once finely tuned system to govern and administer provinces has become ineffective. This ineffectiveness was exhibited by the numerous movements dissenting from Ottoman cohesion, as well as the failure of Ottoman officials overseeing these areas to appease the growing ethno-religious movements (e.g., Orthodox Christianity, Shi’i Islam, Pan-Arabism) or to recognize that they were a threat that might need intervention by the Ottoman metropole, in the aforementioned “communications problem” (Kumar, 1962).

As discussed above, these growing separationist sentiments were attractive to European imperial powers aiming to gain further influence around the Ottoman Empire’s borders and slowly chip away at its core. Meanwhile, leaders in areas such as Kuwait took advantage of their territories’ political, geographic, and
economic importance to achieve their desires of autonomy by balancing various powers appropriately over time. Once imperial powers gained further access and influence into the Gulf through agreements such as this with local leaders, however, their policies were not as predictable, as seen with Iraq and the British.

Britain’s miscalculations in power balancing during World War I and throughout the early 20th century in what was once Ottoman Mesopotamia reiterated Ottoman convictions that British presence in the Arab provinces was akin to a civilizing mission—the British disregarded the complexities of the occupied community and assumed that all those from a general geographic location required similar policies (which were mostly built on a more aggressive divide-and-rule method once the British assumed power).

Legacies of imperialism include governmental infrastructures that the now “free” people are left to grapple with to define their new national identity. In the case of Iraq, this proved a detrimental move by the British, later inspiring major upheavals domestically as well as bilaterally. Relatedly, as previously mentioned, domestic conditions are a great determinant of political decisions, meaning that the raging clash of movements within Iraq left leaders with limited options for maintaining control and nationalistic identity. Further, outsider threats (whether real or perceived) are one method of improving cohesion, further encouraging Iraqi authorities to instigate conflict over border integrity throughout the 20th century, with heightened claims in times of economic distress or war, culminating in the 1990 Iraqi invasion on Kuwait, a territory largely autonomous before its independence and fully sovereign at the time of the attack. Power-balancing shortcomings by both Ottoman officials and the British government with regard to Arab territories hence contributed to major political changes in both Kuwait and Iraq, with some being quite detrimental.
References


201


EFFECTS OF THE COVID-19 PANDEMIC ON U.S. EMPLOYMENT RATES

JORDAN NEESE, OLIVET NAZARENE UNIVERSITY
MENTOR: KAYLA BRUCE

Abstract

The current coronavirus pandemic has caused a multitude of jobs and lives to be lost globally. While COVID-19 has affected everyone in some way, certain demographics have been affected more than others in terms of employment. Industries, women, LGBTQ+ people, and people ages 16–25 have been most economically affected by the pandemic. COVID-19, a respiratory disease, has spread widely in a short time, and the spread of the virus was predicted to get even worse in the winter of 2020–2021. With a large second wave coming, more jobs were expected to be lost and the to be hit hard once again. Although regulations were in place, not everyone followed them consistently, or at all. With a president who downplayed the virus tremendously and did not consistently wear a mask—and was even infected by SARS-CoV2, the virus that causes COVID-19—Americans were in a war with each other about the severity of the virus.

Although the pandemic has had many negative impacts on everyone, this paper discusses the impact of COVID-19 on employment rates in the United States of America. It addresses the groups who have been most economically affected by the coronavirus, and examines unemployment rates prior to COVID-19 and in late 2020.

According to the World Health Organization (WHO, 2020), SARS-CoV-2, the virus that causes COVID-19, spreads mainly through droplets of saliva or discharge from the nose when someone who is infected coughs or sneezes. Most people who are infected with the coronavirus experience mild symptoms that affect the respiratory system (WHO, 2020). Older people and those who are immunocompromised are more likely to experience harsher symptoms and are more likely to die from being infected with the coronavirus. The WHO states that the best way to prevent the spread of COVID-19 is to frequently wash or sanitize your hands, wear a mask, and participate in social distancing.
This coronavirus has affected people globally, with 1.15 million deaths and 42.8 million cases worldwide as of November 2020. As these numbers change daily, Americans can easily find updated statistics on the New York Times’ coronavirus updates page. In the United States as of November 2020, there had been 225,000 deaths and 8.4 million cases, meaning that America’s COVID-19 cases and deaths made up almost 20% of global cases and deaths at the time (Almukhtar et al., 2020). These numbers were predicted to only get larger as the winter went on, which is why everyone must take the proper precautions to help prevent the spread of coronavirus.

As of October 7, 2020, an estimated 14.7 million jobs had been lost in the United States because of the pandemic (Fronstin & Woodbury, 2020). This paper will address how COVID-19 has affected American employment, as well as which groups of people have been most affected by unemployment. This paper addresses three research questions:

1. What were the employment rates in the United States prior to COVID-19?
2. How has COVID-19 affected Americans’ employment?
3. Who have been the main groups of people affected by the coronavirus with regard to unemployment?

The coronavirus has affected everyone in some way, myself included. With most classes moving online, students were forced to try to adapt as quickly as possible to the plethora of changes. Teenagers and those without compromised immune systems began doing the grocery shopping for those who no longer felt safe venturing outside of their homes and into crowded stores and public places. Online shopping has been at an all-time high, with large corporations like Amazon thriving while local businesses go out of business because of their inability to keep up supplies to meet demand. In short, the coronavirus has affected everyone in similar yet different ways, and with how the coronavirus has affected local communities and loved ones, everyone should be willing to learn about and properly educate themselves on the impacts of the pandemic.

**How COVID-19 Has Affected Unemployment Rates in the United States**

With the spread of COVID-19 across the nation, implications for American workers have been staggering. In February 2020, the unemployment rate in the United States was 3.8%, which skyrocketed to 13.0% by the end of May after reaching its all-time high in April, when 14.4% of Americans were unemployed (Kochhar, 2020). The unemployment rates fell further after April 2020, with only
7.9% of Americans being unemployed in the month of October. With COVID-19 spreading rapidly in the autumn of 2020, however, the percentage was expected to only increase throughout winter (Kochhar, 2020).

Because of the coronavirus pandemic, the FFCRA (Families First Coronavirus Response Act) was implemented by the U.S. Department of Labor. This act provides regulations to the workforce in light of businesses and schools reopening. According to the U.S. Department of Labor (2020), “The FFCRA helps the U.S. combat and defeat the workplace effects of the coronavirus by giving tax credits to American businesses with fewer than 500 employees to provide employees with paid leave for certain reasons related to the coronavirus.” Under the FFCRA, workers are qualified to use paid leave “only if work would otherwise be available to them” (U.S. Department of Labor, 2020). This essentially means that employed workers who have been infected with the coronavirus may receive paid leave. The act also states that an employed worker who must provide care to someone in their household who has been infected with the coronavirus may receive paid leave. Nothing is definite, because it is up to employers to decide, but workers have the right to receive paid leave because of the coronavirus pandemic.

The CDC (Centers for Disease Control, 2020) provided a list of suggestions for businesses during the COVID-19 pandemic. Some of these suggestions consist of conducting daily health checks, conducting a hazard assessment of the workplace, encouraging employees to wear cloth face coverings when appropriate, implementing social-distancing policies, and improving buildings’ ventilation systems. These suggestions are not followed by all businesses and restaurants, however, specifically small, locally owned businesses, given the mass amounts of profits they have lost because of COVID-19 and resulting restrictions. This does not mean that all small businesses do not follow these restrictions, but the fact is that chain businesses, including restaurants, can afford to limit the capacity of their consumers, while small businesses cannot. Although the CDC has provided some tips for small businesses during this pandemic, those tips are hard to successfully enact because of the loss of profit. Small business owners must keep their employees in mind and keep them afloat while also keeping their businesses running. Employee compensation can be hard to maintain because of the loss of profit, and many small business owners refuse to lay off their employees because they do not want their workers to become unemployed and face even harder financial consequences of COVID-19.

In Illinois, the state where I reside, Governor Pritzker announced that many of the state’s regions would face restrictions on restaurants and bars due to the spike
of COVID-19 in these regions. Many small businesses refused to close the indoor dining areas of their restaurants, however, because their businesses would not be able to survive. Just the other day, I went out to lunch with my grandma at one of the restaurants that refused to close. For purposes of anonymity, I will not disclose this restaurant’s name, but the restaurant did appear on the news, with the owner saying the restaurant was going to remain open because the owner had a business and 30 employees to take care of. When my grandma and I showed up, there was a wait to even enter the little restaurant. We waited for about 15 minutes, and while we waited, a man waiting outside with us said, “I really hope this works out for her. She’s a good person, and I just really hope this all works out for her.” This is the attitude that a lot of people have in regard to local businesses. Regardless of political affiliations, communities are coming together to help their local businesses in whatever way they can. Quite frankly, this was the busiest I had seen this restaurant in years! Although the coronavirus has had more negative effects than positive, it has brought communities together in some ways.

Depending on who you talk to, you will receive different responses on this question, which is why this section as brief as possible. The information I have read about how President Trump handled the coronavirus pandemic comes from former national security advisor John Bolton, including from his book The Room Where It Happened, in which he touches on how Trump reacted to COVID-19, but the majority of the information in this section comes from an interview of Bolton on CNN. The lack of quotes from Bolton’s book in this section is a result of the book being criticized for releasing classified information and because of future court dealings, but I suggest that everyone read the book before it is possibly too late.

In his book, Bolton criticizes President Trump for many different actions, but he also praises Trump for others. One of the actions that Bolton continues to criticize President Trump for is his handling of COVID-19. In a June 2020 interview, Bolton told CNN:

I think there is an empty chair in the Oval Office, because the President did not want to hear bad news about Xi Jinping, his friend. He did not want to hear bad news about the cover-up of the virus in China, or its potential effect on the China trade deal that he wants so much. And he didn't want to hear about the potential impact of a pandemic on the American economy and its effect on his reelection. Turning a blind eye to all these early signs I think hampered the country's ability to deal with this, and continues to do so. (Cohen & Atwood, 2020).
According to Bolton, President Trump was worried only about the American economy, not the health of the American people. This was proven true by Trump himself. Trump did not like listening to the doctors and instead wanted to reopen the country so the economy could recover. Experts, however, had predicted that even if the country were to reopen, the American economy would not recover that quickly.

Throughout his book, Bolton discusses foreign policies that Trump enacted, as well as Trump’s friendships—including failed friendships—with other countries’ leaders. Bolton discusses his time in the White House and how sporadic dealings could be at times. Although Bolton’s credibility has been debated by those who support Trump, that does not change the fact that he was in close contact with President Trump for many months.

**Employment Rates in the United States Prior to COVID-19, and Whom COVID-19 Has Affected Most**

Certain states’ employment have been affected by the coronavirus pandemic more than others, as recognized by *Forbes*. In May 2019, the unemployment rate in Nevada was 4.0%, but a year later, during the coronavirus pandemic, it rose to 25.3%. In Hawaii, the unemployment rate in May 2019 was 2.7%, which had risen to 22.6% a year later because of the coronavirus pandemic. Other states were not doing nearly so badly, including Nebraska, whose unemployment rate rose from 3.1% to 5.2% (Patton, 2020).

Typically, the states that had not faced serious consequences regarding unemployment due to the pandemic do not rely heavily on industries that have been negatively affected, or they have more diverse economies. The leisure and hospitality industries have been most negatively affected, with the unemployment rate for these industries rising 618%, from 5% in May 2019 to 35.9% in May 2020 (Patton, 2020).

With regard to age groups, the greatest rise in unemployment coincided with workers under the age of 25. The largest spike in unemployment belonged to the 16–19 age group, and the second largest spike belonged to the 20–24 age group. This is because younger workers tend to be in minimum-wage jobs because these jobs are easily accessible for students and those without high school and college degrees. Younger workers tend to have fewer marketable skills and less seniority and are therefore typically the first group of people to be laid off, according to *Forbes* (Patton, 2020). Minimum-wage jobs are typically defined as “unskilled
jobs,” meaning that anybody can participate in them because they require no special degree. These jobs have therefore been hit the hardest during the coronavirus pandemic; businesses can afford to lay off workers in these positions because they are not necessarily essential to the business. Nursing, teaching, and other jobs requiring degrees, as well as jobs that belong to unions, have not been hit as hard during this pandemic. That does not mean they have not been affected, however. Nurses and teachers are defined as essential workers, while McDonald’s employees are not.

The coronavirus pandemic has also had a large impact on employment rates by race and ethnicity. Prior to COVID-19, the unemployment rate for African Americans was at a record low, only 5.1% in October and November 2019. That rate as of May 2020 was 16.8%. As of May 2020, the unemployment rate was 17.6% for Hispanics and Latinos, 15% for Asians, and 12.4% for Whites (Patton, 2020).

Women have been more affected than men by the coronavirus pandemic in terms of unemployment (Madgavkar et al., 2020). Women’s jobs are 1.8 times more vulnerable to being lost during this pandemic. One reason is that the virus has greatly increased the burden of unpaid care, which is carried mainly by women. Another reason that women are more likely to be negatively economically affected by the coronavirus pandemic is preexisting gender inequalities in the workplace. Prior to COVID-19, women made up 46% of the workforce (Madgavkar et al., 2020); it is estimated that 54% of the total jobs lost during this pandemic have been those worked by women. A third reason that women are losing jobs at a higher rate than men is that women and men tend to work different jobs. I am not saying that this is all the time, but for reference, most people think of nurses as being women and not men—that is just a small example of the big picture here. One study shows that jobs women are working are 19% more at risk than those worked by men because “women are disproportionately represented in sectors negatively affected by the COVID-19 crisis. We estimate that 4.5 percent of women’s employment is at risk in the pandemic globally, compared with 3.8 percent of men’s employment, just given the industries that men and women participate in” (Madgavkar et al., 2020). With this information in mind, we will next move on to how COVID-19 has economically affected LGBTQ+ people.
Regardless of COVID-19, LGBTQ+ people face higher unemployment rates during economic crises than do people not of sexual or gender minorities (Gruberg & Madowitz, 2020). Figure 1, from Gruberg and Madowitz’s article, shows how same-sex couples have been affected economically compared to opposite-sex couples, with green representing opposite-sex couples and blue representing same-sex couples. This figure shows that for 2014–2019, households with opposite-sex couples faced lower unemployment rates than did households with same-sex couples. Households with same-sex couples are also more likely to receive Supplemental Nutrition Assistance Program (SNAP) benefits than are households with opposite-sex couples. Discrimination is also a large part of why LGBTQ+ people face higher unemployment rates than do people not of sexual or gender minorities. Many businesses, whether large or small, discriminate against LGBTQ+ people, not allowing them to even enter the building, let alone work there. Some businesses, such as Hobby Lobby, have been known to discriminate against LGBTQ+ people, and some small business owners have been known to not allow LGBTQ+ people to enter their businesses because of “religious freedoms.” The unemployment rate of LGBTQ+ people compared to that of people not of sexual or gender minorities has not been greatly affected by COVID-19 per se, because economic crises in general affect LGBTQ+ people more. Although COVID-19 has caused a large economic crisis, this impact is unfortunately not entirely new to LGBTQ+ people.
My Interviews With People Working During the Coronavirus Pandemic

I conducted an interview on October 25, 2020, with two people I know who were working from home for the most part during the coronavirus pandemic. Kim P. is an executive assistant for a small company in Chicago, and Darin N. is a signals and communications manager for the Canadian National (CN) Railway.

To start the interview, I first asked, “What is it like working during a global pandemic?” Kim responded, “It is a unique experience, given everyone’s working from home. Trying to keep the team morale amongst employees virtually has been difficult but doable.” Darin answered, “At first, it was a lot harder because trying to get field people to work with safer conditions was difficult. Locating where to purchase hand sanitizers and masks in mass amounts was hard.” Working during a pandemic is hard, according to both. Luckily, it has become easier for some, but that does not mean that others are not still struggling.

For my second question, I asked, “Is it easier working from home or in the office?” Kim responded, “Personally, it is better [at home] because you don’t have the long daily commute. Professionally, it’s not easier because the days become longer. You’re expected to be online 24/7 because you happen to be working from home. There’s no real balance—it’s not like a 9-to-5 job anymore.” Darin responded simply, “There is no real difference.”

I followed up by asking, “Is it more stressful working from home?” to which Kim responded that it is and Darin responded, “No, because there are less face-to-face interactions with coworkers.” As read above, the working aspect of working at home is harder, but personally, being home is easier. Or, if you’re like Darin, there is no real difference.

For my fourth question, I asked, “Do you feel as if you’ll go into the office eventually, or can you see your job being mainly at home from now on?” Kim answered, “I’ll be going back into the office post-vaccine. For now, I will be working at home during COVID-19.” Much like Kim, Darin answered, “I’ll be going into the office post-vaccine. I’m already in the office once in a while, but some of our offices are shut down right now because they don’t want people who don’t need to be going in there going in. They want the people who actually need to be there to be safe in their own offices, COVID-19 free.” As read above, many workers will be returning to their workplaces when the coronavirus pandemic has ended or when a vaccine is released to the general public. Some workers were already in the office occasionally.
The next set of questions I asked pertained only to Darin. I first asked, “Do you feel safe going to CN sites?” He responded, “Yes, I feel safe because everyone is working outside, and we are all separated and wearing masks.” I followed that question with “What precautions has the CN taken?” Darin responded, “Masks are mandatory unless you’re working by yourself. Hand sanitizer is necessary, too. We’ve moved everyone out to the field, and we’ve split operation centers into multiple buildings so one outbreak at a train site doesn’t take out a bunch of our workers.” Many businesses have taken the proper precautions to protect their workers, as the CN Railway has: masks are mandatory when one is working with others, hand sanitizer is necessary, and so on. It is important to note that not all businesses are taking these precautions, though. Large businesses, such as the CN Railway, can afford to take these precautions and pay for their workers’ safety. Small businesses, in comparison, may not be able to afford everything necessary to keep their workers safe.

As read above, working from home during this pandemic has not necessarily been easy. Many workers will be entering their workplaces again post-vaccine, if the vaccine is deemed safe by health officials. Working from home amidst the coronavirus pandemic is different for everyone, as seen above. While Kim said working from home is more stressful, Darin sees working from home as being less stressful, and many of their other answers differ greatly.

**Conclusion**

The coronavirus pandemic has affected everyone in one way or another. In general, this pandemic has had a negative impact on the U.S. economy and has affected some people, including women, LGBTQ+ people, and people of color, more than others. While some states had large spikes in unemployment rates, others had seen increases of only 3% or so by late 2020. Unfortunately, unemployment rates were expected to spike once again with the impending second wave of COVID-19 cases. Although precautions have been taken by many businesses, some businesses, such as small businesses, cannot afford to follow them. The coronavirus has caused people to lose loved ones, businesses, and steady sources of income. If we did not take proper precautions, everything would only get worse in winter 2020–2021.
References


THE POLITICAL AND ECONOMIC IMPACT OF MEDIA OWNERSHIP STRUCTURES AND CONCENTRATION ON GLOBAL COMMUNICATION

YUXI WEN, UNIVERSITY OF NOTTINGHAM NINGBO CHINA
MENTORS: XIAOLING ZHANG, ANGELA LEWIS

Abstract

The global media is not limited to disseminating information but also plays the role of fostering ideas, diffusing stories, and exchanging culture. The content produced by global media, however, may be subject to different ownership structures. The ways that different media ownership structures influence global communication in both economic and political significance are therefore worth exploring. This paper first introduces two types of global media ownership structures—state-owned and privately owned—and cites specific examples to parse how these two types of ownership structures may influence global communication economically and politically. The paper then critically discusses issues of media ownership concentration in western democracies.

Keywords: global communication; state-owned media; free media; authoritarianism; ownership structure; ownership concentration

Global media is not limited only to disseminating information in modern society but also plays the role of fostering ideas, diffusing stories, and exchanging culture alongside technological developments and mass communication (Flew, 2018). The availability and accuracy of information are therefore of utmost importance for the massive public, including voters, investors, consumers, and executives, to make rational decisions on daily matters in modern economies and societies. The content produced and diffused by global media may be subject to different ownership structures, however (Hamelink, 2015). It is therefore important to understand how different ownership structures of global media have influenced global communication in both economic and political significance. This paper first introduces two types of global media ownership structures—state-owned and

1 The author offers sincere thanks to Prof. Xiaoling Zhang and Ms. Angela Lewis for their valuable guidance and support in this research.
privately owned—and cites specific examples to compare political and economic impacts of these two types of ownership structures. The paper then critically discusses media ownership concentration in western democracies. This paper adopts the political-economy approach to consider the essential relationship between power and wealth while analyzing media ownership structures and employs egalitarian thinking to assess the inequality and social injustices caused by ownership concentration.

**Literature Review**

State-owned media is fully controlled by the government financially and is editorially at the government’s service, whereas privately owned media can be fully owned by individuals or by private corporations distinctly for commercial revenues. Moreover, governments may fund two noticeably different types of media, which are state-owned media and public-service media. Most public-service media, such as the BBC in the United Kingdom and the CBC in Canada, are directly or indirectly funded by the government, with a few exceptions receiving funds from foundations and businesses. Despite the financial tie with government, public-service media enjoys independent editorial rights and serves the interests of the general public. State-owned media, in comparison, is not only financially dependent on the government but also strictly controlled by the state in both financial decisions and editorial operations, thereby serving the interests of the state (Webster, 1992). The state media discussed in this paper is state-owned media rather than public-service media.

In the modern world, although state-owned media may be advantageous over privately owned media economically, it undermines political and market freedom by exploiting a nation’s soft power, controlling the flow of stories, and manipulating information dissemination in global communication. Public interest theory, established under the premise that governments maximize the welfare of citizens, is notably the first media theory that compares the role of state-owned media and private media in mass communication. According to this theory, state-owned media appear to be more favorable than privately owned media on several grounds (Begoyan, 2009). First, information is a public good that is costly to keep from unpaid groups if it has been distributed to paid consumers. Additionally, high fixed costs and low marginal costs can be found in the media industry, as collecting and distributing information are significantly expensive, while the cost of reproducing can be relatively low. These two characteristics fundamentally determine that the ownership of global media may be by either wealthy individuals
or the state. As a result, state-owned media may not only possess the economic strength to distribute information but also be able to maintain the neutrality of the information disseminated instead of serving the exclusive interest of a few wealthy individual owners (Begoyan, 2009).

In contrast, public choice theory holds that state media are more likely to distort information, manipulate public opinions, and preclude the public from accessing facts in order to entrench the power of the incumbent ruling class while private media tend to diversify public narrative by supplying alternative perspectives, verify information authenticity by engaging in independent investigation, and ultimately lead the general public to make informed decisions. In addition, the control mechanism of the free market—open competition between different private media—may also effectively ensure that the information circulated is unbiased and carefully verified (Djankov et al., 2003). Admittedly, given the public-good nature of information, state media may be better placed than private media in economic significance. In reality, however, state-owned media may ultimately undermine the free market and modern democracy because of the coercive dominance of the government and the absence of independent editorial accountability, while private media with market competition can sustain the checks and balances of modern democracy and serve the general public with truthful information.

In a study examining the media ownership structures of 97 countries, researchers found ample evidence favoring the public choice theory; no empirical evidence has been established to support the public interest theory. The research found that a greater number of state-owned media tends to result in a lower level of freedom of press, which is represented by a higher number of jailed journalists and a greater number of media closed by the government (Djankov et al., 2003). Moreover, heavy internet restrictions and censorship can be found in countries with a substantial number of state-owned media, such as China, Iran, and Venezuela, which casts doubts on the premise of the public interest theory that state-owned media serve benevolent purposes (Djankov et al., 2003). Additionally, state-owned media are more likely to reshape global communication that favors the state by expanding its soft power. Nye (2006) defines soft power as the key to success in world politics and demonstrates that politics becomes a race for legitimacy, awareness, and credibility after digitalization vanishes national boundaries. The ability to spread stories and to convince story recipients is becoming vital to global power grip; therefore, through the expansion of soft power, state media would be
able to justify the legitimacy of totalitarianism, gain international support, and promote national branding globally.

Originally, the information revolution in global communication appeared to be an opportunity for western democracies to gradually liberalize the autocratic states by broadcasting democratic values through global media, but authoritarian regimes have turned the table on democracies by imposing domestic censorship and exploiting the openness of the democratic societies they broadcast to (Walker, 2016). The combination of China’s domestic censorship and international broadcasting through state-owned media, for instance, has created an alternate reality domestically and beautified China’s image internationally (Lim & Bergin, 2018).

Empirical Findings and Discussion

State-Owned Media in Global Communication

The rise of China’s soft power in Africa may notably exemplify how state-owned media exploit soft power to influence global communication in economic and political discourses, which will be introduced in infrastructure, service, and content sectors. Two types of Chinese media outlets—fully state media and ostensibly private media—can be found in China’s media presence in Africa. China Global Television Network (CGTN), the global English-language media outlet previously known as CCTV-9 and CCTV News, for example, is fully owned by the Chinese government and was officially launched in 2016 as a propaganda machine based in Beijing with coverage of more than one hundred countries. In addition, China Mobile, a Chinese state-owned telecommunication company providing mobile voice and multimedia services, is also an active presence in Africa, serving the interest of the state.

Huawei, ZTE, and StarTimes, in comparison, are ostensibly privately owned and, by definition, pursue profit maximization, although in fact, they all receive substantial financial support from the Chinese government and may be obliged to comply with Chinese laws and regulations. According to China Daily (2014), $163 million was provided to StarTimes, a Chinese electronic and media company, to expand its businesses in Africa by the Export-Import Bank of China from 2012 to 2014, and another $60 million loan was approved for further development in 2014. Additionally, Huawei, a Chinese telecommunication-equipment company providing information and communications technology infrastructure and smart devices, has received as much as $75 billion from the
Chinese government to grow from a start-up to one of the world’s largest global telecom-equipment companies, which presumably results in strong state control over its business operation (Yap, 2019). Similarly, ZTE, a Chinese technology-manufacturing company specializing in telecommunication, is also propped up as a leading global company by receiving heavy subsidies from the Chinese government for its research and development expenditures (Atkinson, 2020). As a result, despite the fact that Huawei, ZTE, and StarTimes are ostensibly privately owned, their strong financial ties with the Chinese government and their obligation to comply with Chinese law have led them to being state-controlled in nature.

In the information technology-infrastructure sector, according to the Center for International Media Assistance (CIMA), China invested more than $5 billion in telecommunication infrastructures for dozens of African countries from 2000 to 2013 (Yudico, 2017). Moreover, Chinese telecommunication companies, including Huawei, ZTE, and China Mobile, have dominated telecommunication infrastructure and undersea cable all across Africa, fundamentally creating a solid basis for state-owned media to control the content broadcast in Africa, promote China’s model, and serve China’s military presence in Djibouti (Yeophantong & Wang, 2019). The Zambian and Ugandan governments, for instance, contracted with Huawei to install network surveillance and censorship instruments to monitor their own citizens as a result of the countries duplicating China’s authoritarian model (Yudico, 2017). Furthermore, China’s undersea cable in Africa, built by a state-owned telecommunication company, also enables encrypted data transmission from Africa to Beijing, serving China’s Djibouti military base. Additionally, in the media service sector, StarTimes, China’s primary television service provider in Africa, offered pay-television channels to more than 30 African countries with 10 million subscribers. The economic affordability (only $4 per month) and adequate quality secured comparative advantages for StarTimes over western television service providers, allowing Chinese state media to further enhance its capability to propagate China’s version of stories (Marsh, 2019). Finally, in the content sector, content broadcast by state-owned media may always rest with the political interests of the ruling power because of the lack of independent editorial accountability. State media may therefore be able to redefine the norm of freedom of the press, induce self-censorship, create an alternate reality, and ultimately justify the legitimacy of totalitarianism in the international community (Walker, 2016). Content provided by Chinese state media in Africa, such as CGTN, may therefore favor Chinese authority and create a positive image of China abroad. Additionally, even outside of Africa, CGTN has frequently appeared on the public stations of foreign countries, such as Peru, to disseminate state propaganda abroad, which is
known as the “borrowed boat” strategy (Cardenal et al., 2017). Hence, although state-owned media may be economically advantageous over privately owned media, its economic benefits cannot offset its erosion of political freedom and the fact that it jeopardizes democratization in global communication.

Privately Owned Media in Global Communication

According to public choice theory, introduced above, privately owned media play a positive political role in global communication, unlike state-owned media. First, private media, if free from government intervention, may function as a watchdog to monitor government officials who are in power and may thereby render the government more responsible. Consequently, individuals who consume such media would be well informed in choosing political candidates, selecting goods, and making investment decisions (Djankov et al., 2003). Moreover, private media is considered an important component of democracy and an effective way to promote democracy globally, as it reinforces freedom of the press and preserves freedom of speech. For instance, The Washington Post, one of the oldest privately owned newspapers in the United States, uncovered the Watergate scandal, the most infamous political scandal in the history of the United States. Two exceptional watchdog journalists from the Washington Post revealed how President Nixon used federal agents to negligently influence his reelection, and their revelation led directly to the impeachment and resignation of President Nixon (Schudson, 2004). It is presumably impossible to uncover presidential wrongdoing without the involvement of private media.

In addition, Van Belle (1997, cited in Whitten-Woodring, 2009) suggested that two countries with privately owned media are less likely to enter into military conflicts with one another because the legitimacy of private media in both countries facilitates the dissemination of trustworthy information and creates a recognition of shared values. Private ownership may therefore play a positive role in fostering the story of multilateral trustworthiness and in maintaining peace at the global level.

Ownership Concentration: Phenomena, Causes, and Consequences

Although privately owned media may positively influence global communication in the above aspects, the profit-maximizing nature of private media in the era of digitalization tends to gradually raise social inequality and reduce information pluralism because of ownership concentration. The emergence of ownership concentration may be traced back to the early history of all relevant
industries. Since the 1920s, the global film industry has been dominated by eight primary companies—Paramount, Warner Bros., 20th Century Fox, Loew’s Inc., United Artists, Universal Pictures, RKO, and Columbia—that have controlled the production and distribution of movies around the globe (Hamelink, 2015). The concentration issue in the film industry is now even more serious, such that Disney, with its large scale of acquisition (including 20th Century Fox), accounted for nearly 40% of the U.S. box office in 2019, whereas its closest rival, Warner Bros., accounted for only about 14% of the box office in the same period (Whitten, 2019).

Furthermore, in the global news industry, the production and distribution of international news has been controlled by the “big four” news agencies—United Press International, Associated Press, Reuters, and Agence France-Presse—since the late 19th century (Hamelink, 2015). In addition, the telecommunication manufacturing industry has also been significantly consolidated since the 1960s, with 13 telecommunication manufacturers accounting for 90% of international supply in 1978 and only 5 telecommunication manufacturers accounting for 76% of the global market in 2019 (Hamelink, 2015; Weissberger, 2020). The above examples serve as compelling evidence of the unprecedented large scale of media ownership consolidation. The causes and consequences of ownership concentration may therefore be worth exploring.

From the economic viewpoint, ownership consolidation may be attributed to developments in global digitalization, high rates of merger, and the interlocking interests of companies. The emergence of new technologies seems to reduce the high fixed costs required to produce and distribute information under conventional technology and thereby facilitates the establishment of new media outlets to end the dominance of existing media tycoons, whereas in fact, digitalization reinforces the dominating capacity of existing media giants. As digitalization allows media content to be spread and stored online without physical interactions, the marginal cost of producing additional copies has been reduced to nearly zero, so media tycoons with large numbers of consumers are more likely to share the first-copy costs with more individuals than are new entrants, thereby reaching efficient economies of scale (Goodwin, cited in Brown, 1999). Moreover, in a highly digitalized world with mass communication, the attractiveness of media content becomes a key element in gaining investments from advertising agencies, which are considered the primary source of revenue for media outlets and consequently decide the fate of media companies. Large global media companies with more resources may therefore secure the comparative advantages over new entrants of absorbing investments from advertising corporations by substantially expanding
investment in production and distribution to ensure a higher level of attractiveness (Herman & Chomsky, 1994).

The second major reason for ownership concentration is likely to be the high merger rate. Economically speaking, the survival chances of information companies, especially in the era of digitalization, appear to be greater if small businesses become part of large conglomerates, as tremendous investments with enormous financial risks and advanced mass communication technology are required to reach a large enough audience for a company to financially survive in the global communication industry. As a result, the mass communication industry has a tendency to be consolidated into a handful of dominant media conglomerates that operate primarily for profits (Herman & Chomsky, 1994).

Given the above analysis, existing global media giants and advertising corporations exhibit more common interests than anticipated. The communication industry seemingly enjoys the free market and fair competition, but in fact, the interests of dominating tycoons are closely intertwined through various connections, such as joint ventures, joint ownership, joint directorates, and mutual agreements, in order to exploit monopolistic advantages and profit maximization (Herman & Chomsky, 1994). For instance, joint directorates can be found between IBM and Time, and mutual agreements can be seen between IBM and AT&T (Hamelink, 2015). As main competitors may reach a multilateral consensus regarding production, supply, sales, distribution, and price through those interlocking connections, it is safe to conclude that an authentically competitive market may be absent in the communication industry, which deters new entrants and exacerbates ownership consolidation.

Now that the phenomena and causes of ownership concentration have been discussed, attention can now be turned to the discussion, from an egalitarian perspective, of two primary consequences of ownership concentration: inequalities in the public sphere and loss of expression diversity. It is reasonable to believe that when a small group of very wealthy people are in charge of many media conglomerates, the combination of wealth and power may lead to significant inequalities in public decision-making and social justice, as unelected wealthy owners are more likely to prioritize their own interests by controlling the flow of stories to favor themselves than to hold themselves accountable to society (Hamelink, 2015). In addition, a high level of media ownership concentration may result in a unanimous single voice on certain issues and thereby reduce the pluralism of opinions and information.
Rupert Murdoch, who is one of the world’s wealthiest and most influential men and controls 57% of newspapers in Australia, at least four national newspapers in the United Kingdom, substantial shares in Sky Group worldwide, and several high-profile U.S. newspapers (including WSJ), for example, undeniably plays a leading role in global politics, economics, and social issues (BBC, 2011; Evershed, 2020). It has been reported that Murdoch abused his power to gain commercial benefits by ordering his editors and journalists to run lobbying campaigns to weaken media ownership regulations and influence political candidates and government cabinets (Lidberg, 2019). Moreover, his son James Murdoch has publicly criticized his father’s media outlets for disseminating misinformation about the ongoing climate-change discussion and for downplaying the severity of the climate crisis despite overwhelming scientific evidence that favors the urgent necessity of addressing climate change; this downplaying consequentially deters mutual consensus on climate change within the Australian government (Waterson, 2020). Additionally, the loss of voice diversity can be evident from the fact that media controlled by Murdoch spread a single voice supporting the Iraq War in 2003, even though the existence of weapons of mass destruction in Iraq—used as justification for the Iraq War—was never proven (Waterson, 2020). The above examples may justify the necessity of addressing the problem of ownership concentration in private media. Government regulations or laws that set maximum levels of individual media ownership appear to be indispensable.

Conclusion

Using the political-economy approach, this paper has comprehensively explained how ownership structures and concentration have influenced global communication. On the one hand, although state ownership appears to be economically superior, state-owned media is politically toxic, as it undermines the universal value of democracy by controlling the flow of stories, serving state interests, and propagating justification for totalitarianism. On the other hand, privately owned media tends to play the role of watchdog over governments and safeguard the world’s peace, although its ownership-concentration problem may ultimately reduce information pluralism and increase social injustice. Hence, the transition of state-owned media to privately owned media, along with the mitigation of the ownership-concentration problem through legislation restrictions, is suggested globally.
References


GOODPASTURE SYNDROME: AN INVESTIGATION OF DISEASE PROCESS, DIAGNOSIS, TREATMENT, AND CONTRIBUTION TO INTRAPULMONARY HEMORRHAGE

JULIA SHIELD, QUEENSLAND UNIVERSITY OF TECHNOLOGY
MENTOR: MARK B. PLENDERLEITH

Abstract

Goodpasture syndrome is a rare autoimmune disorder affecting approximately one person per million. It is caused by the development of anti-glomerular basement membrane (GBM) antibodies. Since the description of Goodpasture syndrome in 1958, understanding of the syndrome’s pathophysiology has significantly increased with advances in molecular biology, delineation of immune function, and research into the structure and function of collagen, which forms the basement membranes in the alveoli and glomeruli. Although the exact trigger for the development of these autoantibodies is unknown, there is a clear genetic predisposition in affected individuals with a correlation to human leukocyte antigen (HLA) alleles in anti-GBM disease. Environmental factors such as viral infection and smoking then damage the alveoli, causing increased permeability and increased access to the basement membrane for the autoantibodies. The autoimmune response attacks the basement membrane of both alveoli and glomeruli, causing pulmonary hemorrhage, hemoptysis, anemia, respiratory failure, glomerulonephritis, and renal failure. It is diagnosed by a combination of respiratory function tests, chest X-ray, serology for anti-GBM antibodies, and renal biopsy. A renal biopsy reveals crescentic glomerulonephritis and linear immunofluorescent staining from autoantibody deposition in the capillaries. Disease progress is monitored by using these tests serially. Treatment involves immunosuppressive medications such as corticosteroids and cyclophosphamide to prevent autoantibody production, as well as plasmapheresis to remove autoantibodies from circulation. Novel treatment methods currently under investigation include immunoadsorption, cryofiltration, and enzymatic degradation. Immunoadsorption is a developing therapy with early results showing high effectiveness, but further trials are necessary. New developments for using enzymatic proteins from the Strepococcus pyogenes bacteria to remove anti-GBM autoantibodies that are bound to the actual basement membrane and not just those in circulation are currently under investigation. This is facilitated by further recent
advances in delineating collagen molecular structure composition in the basement membrane. Provided Goodpasture syndrome is diagnosed early, treatment regimens are very effective in controlling the disease; however, late-stage disease requires dialysis and lung or kidney transplantation. As advances in molecular biology and genetics continue, the immune function is mapped in greater detail, and genetic susceptibility is better understood, more opportunities to treat autoimmune conditions such as Goodpasture syndrome will develop.

Goodpasture syndrome is a rare autoimmune disorder characterized by the destruction of the basement membrane in both the lungs and kidneys (Zhong et al., 2020). Circulating anti-glomerular basement membrane (GBM) antibodies attach to collagen-fiber networks in the membrane, which triggers the membrane’s destruction, causing alveolar hemorrhage, resulting in anemia and respiratory failure due to impaired gas exchange (Greco et al., 2015). Alveolar hemorrhaging from basement-membrane breakdown causes dyspnea, hemoptysis, and chest pain (Huart et al., 2016). Given the potential severity of these clinical manifestations, pulmonary abnormalities usually become evident before respiratory failure develops; however, a rapid diagnosis is essential, as hemorrhage is the leading cause of death in individuals with Goodpasture syndrome (Nasser & Cottin, 2018). Although the exact trigger of this antibody-induced autoimmune response remains unknown, certain environmental factors (e.g., chemical or infectious exposure) coupled with genetic susceptibility may lead to its development (Pedchenko et al., 2018). Diagnosis is achieved by a combined interpretation of respiratory function testing, antibody detection through serology, chest X-ray, and lung or kidney biopsy. Patients diagnosed with Goodpasture syndrome undergo an array of treatment depending on the severity of their condition. Commonly used therapies include plasmapheresis, immunosuppression, renal dialysis, and organ transplantation. Emerging research has provided numerous opportunities for alternative methods of treatment in the near future. The following review will utilize current literature and research to investigate the history, epidemiology, etiology, and pathophysiology of Goodpasture syndrome. The most common forms of diagnostic and treatment procedures will be analyzed. Additional insights for emerging research in the field will also be discussed.

**Literature Review**

This review will begin by highlighting the initial investigations of Goodpasture syndrome, proceeded by more modernized approaches by relating
incidence of alveolar hemorrhage to renal failure. Discussion will turn to further analysis of contrasting clinical manifestations, diagnostic methods, and treatments.

Goodpasture syndrome was first described in 1958 by Australian scientists Stanton and Tange (Pedchenko et al., 2018), who reported nine cases of rare glomerulonephritis with close association to events of idiopathic pulmonary hemorrhage. One case study described a man admitted to hospital with anemia, hemoptysis, and signs of bronchopneumonia in the right lung and who died after three days (Collins, 2010). Urinalysis showed traces of albumin, which is indicative of glomerulonephritis. A postmortem examination showed that the terminal bronchioles filled with erythrocytes, in addition to necrosis of the alveolar walls (Collins, 2010). The authors named the disease in recognition of similar case reports described by Dr. Ernest Goodpasture in his 1919 publication on the etiology of influenza (Greco et al., 2015). At that time, similar clinical presentations of lung hemorrhage linked with glomerulonephritis were attributed to atypical influenza infection. Not until the detection of anti-GBM antibodies in the 1960s was a clinical description of the pathogenesis for autoimmune-induced Goodpasture syndrome possible (Gulati & McAdoo, 2018).

Since its establishment in 1986, the Vasculitis Foundation has provided research funding, educational resources, and networking opportunities for patients affected by Goodpasture syndrome and other closely related diseases (National Organization for Rare Disorders, 2019). Given the rarity of the disorder, archives like these allow continued investigation into the pathophysiological processes and controversial etiological implications of the disease. Consequently, the best possible diagnosis and treatment options for patients can be provided as research progresses.

**Epidemiology and Incidence**

Goodpasture syndrome is a rare disorder, with 0.5 to 1.8 cases per million individuals, according to a study published in 2016 that monitored disease numbers across 26 countries (Canney et al., 2016). This is similar to previous approximations of 1 to 2 cases per million individuals each year across Caucasian populations (Gulati & McAdoo, 2018). Some studies report that the disease is more common in males than in females, with a 6:1 occurrence ratio (Kaewpu et al., 2020), but an investigation from the University of Bari found no significant difference in incidence between gender for affected individuals (Dammacco et al., 2013). Goodpasture syndrome has a bimodal age distribution, primarily affecting individuals between 20 and 30 years and between 60 and 70 years (Betsy et al.,
Furthermore, it is most common in adults with Caucasian or European descent, with African populations less likely affected (Marques et al., 2020).

**Etiology**

Goodpasture syndrome is caused by the production of anti-GBM autoantibodies that travel through the bloodstream and mediate the destruction of the alveolar and glomerular basement membranes (Nasser & Cottin, 2018). The specific trigger for the development of these autoantibodies is not fully understood. Certain genetic and environmental influences may contribute to the development of this disorder, however (Pedchenko et al., 2018).

Cases of Goodpasture syndrome are often reported occurring among families, which suggests a genetic susceptibility (Gulati & McAdoo, 2018). There is a clear correlation with human-leukocyte antigen (HLA) alleles in anti-GBM disease, with the antigen HLA-DR15 being found in 88% of affected individuals (Nasser & Cottin, 2018). Environmental factors including viruses may contribute to the development of the disease by enacting antibodies to fight infection, which then cross-react against the basement membrane. Consequently, the body’s immune defenses can attack surrounding healthy tissue (Marques et al., 2020). Pulmonary hemorrhage is also closely associated with smoking, as cigarette contents break down the thin alveolar lining and surrounding capillaries (Nasser & Cottin, 2018).

**Pathophysiology**

As discussed, Goodpasture syndrome is caused by the destruction of alveolar and glomerular basement membrane by circulating anti-GBM autoantibodies. To understand how this process occurs, we must first understand the structure and function of antibodies and basement membranes. Investigation will turn to the development of secondary outcomes and disruption of the physiological processes involved in respiration.

The autoantibodies reactive with the alveolar and glomerular basement membranes are from the IgG class. Pathological antibodies also belong to this class, but anti-GBM antibodies belong to subclasses IgG2 and IgG4, which enable variation in the fragment antigen-binding domains and therefore specificity in binding (Gulati & McAdoo, 2018). Such observations show the significance of evaluating the structural and binding features of autoantibodies to understand their
ability to trigger an autoimmune response and therefore the pathogenesis of Goodpasture syndrome.

Alveolar basement membranes are extracellular matrices comprising type IV collagen, heparan sulphate proteoglycan, laminin, and nidogen (Gulati & McAdoo, 2018). Six alpha chains compose the collagen IV family, which bind to form unique triple-helical protomer structures (Pedchenko et al., 2018). These polymerize to form hexametric structures that provide support to overlying cells. In the basement membrane, type IV collagen networks in their native form are arranged into triple-helical protomers from α3, α4, and α5 chains (Zhong et al., 2020). These domains associate to produce the hexameric-NC1 domain. The quaternary structure becomes reinforced by hydrophilic, hydrophobic, and disulphide bonds across separate domains (Gulati & McAdoo, 2018), and these polymerize with other like units to form a lattice network of collagen, as shown in Figure 1 (McAdoo & Pusey, 2017). The NC1 domains of collagen protomers arise in specific combinations that produce linear arrays found in both the lungs and kidneys.
Figure 1. Collagenous Structure of the Alveolar and Glomerular Basement Membrane

Note. Panels A and B: Structure consists initially of α3, α4, and α5 chains to produce a triple helical protomer. Panel C: Association of these protomers produces the hexametric NC1 domain. Panel D: Multiple domains bind by hydrophilic and hydrophobic interactions to form the lattice structure found in α3(IV)NC1 collagen networks. From "Anti-Glomerular Basement Membrane Disease," by S. McAdoo and C. Pusey, 2017, Clinical Journal of the American Society of Nephrology, 12(7), 1162-1172 (https://doi.org/10.2215/CJN.01380217).

The core target for circulating anti-GBM antibodies is the noncollagenous domain (NC1) in the α3 chain of type IV collagen, or α3(IV)NC1 (Gulati & McAdoo, 2018). This molecule contains two epitopes, E_A and E_B, where antibodies bind, allowing T-cell recognition to promote a humoral immune response. This is also known as the self-antigen, which triggers the autoimmune response in Goodpasture syndrome after antibody binding is initiated (Zhong et al., 2020). Specific binding of antibodies to the alveolar basement membrane is enabled by the
epitopes on self-antigens as well as the arrangement of collagen chains allowing easy access to traveling antibodies. In a healthy individual, endothelia in alveoli provide a protective barrier so anti-GBM antibodies are unable to bind to these epitopes (Hudson et al., 2019), but antibodies can more efficiently bind to the epitope units in collagen domains of the basement membrane because of increased permeability from either increased capillary hydrostatic pressure or environmental factors including smoking and hydrocarbon exposure or pathogenic infection in the alveoli (Pedchenko et al., 2018).

A strong correlation has been shown with Goodpasture syndrome and the HLA system. This system aids in human immunity by determining the difference between self- and non-self-antigens. HLA-DR15 is one such HLA located in about 88% of people affected by Goodpasture syndrome (Nasser & Cottin, 2018). The high prevalence of these alleles within individuals with the disease highlights a potential-genetic predisposition for this rare autoimmune disease; however, certain alleles involved with the HLA system are fairly common among unaffected populations, suggesting that additional factors are involved in development of the disease (Pedchenko et al., 2018). Additionally, a study released in 2016 investigated spatial and temporal clustering of cases, which indicates that the presence of environmental factors may contribute to provoking disease (Marques et al., 2020). Some preexisting damage to the alveolar lining is essential to allow the increased permeability that provides access for autoantibodies. Binding can then occur on the epitopes on the self-antigens of membranous subunits and activate cascade reactions that ultimately result in injury to the lining (Nasser & Cottin, 2018).

The binding of antibodies to the basement membrane triggers a response via two effector mechanisms. In the first, inflammatory processes are activated, which enables the mobilization of leukocytes and inflammatory mediators, forming the lytic membrane attack complex. The second response involves binding of Fc receptors (FγR) by anti-GBM antibodies (Dammacco et al., 2013). This binding action activates phagocytosis and antibody-dependent cellular cytotoxicity to produce a strong autoimmune response. There remains a distinguishable correlation between the presence of HLA and anti-GBM antibodies in Goodpasture syndrome. HLA is a gene complex for major histocompatibility complex (MHC) proteins, which help regulate human immunity (Hudson et al., 2019). This shows that T cells must be activated before autoantibodies can be generated. After being activated by T-helper (Th) lymphocytes, macrophage cells arrive at the site of antibody accumulation and produce interferon-(IFN)-γ and interleukin (IL)-12 (Dammacco et al., 2013). These cytokines, along with all accumulated cells, induce structural
change within the matrix of proteins in the membrane, causing mechanical weakness, which leads to eventual irreversible damage. Th1 specific memory cells also migrate to the site of inflammation to initiate a delayed hypersensitivity response (Zhong et al., 2020). This contributes to most of the damage over time once the initial response has been established.

The complete pathogenic process by which basement membrane destruction arises in lungs and kidneys affected with Goodpasture syndrome is illustrated in Figure 2.

Figure 2, Pathogenic Procedure of Goodpasture Syndrome Leading to Both Pulmonary and Renal Damage

Significant damage to the collagenous microstructure in the alveolar basement membrane in turn damages the surrounding alveolar capillaries, causing mass hemorrhaging into the alveolar space. The high capacitance but low pressure in the pulmonary circulation causes chronic low-grade hemorrhage (Greco et al., 2015). The secondary manifestations of hemoptysis and anemia may consequently arise. Gas exchange still occurs with the blood pooled in the alveoli, although oxygenation for blood still traveling throughout the pulmonary circulation is no longer possible. This can lead to alveolar collapse, ultimately causing hypoxemia (Kusunoki et al., 2018). In more-severe cases, this can lead to respiratory failure and death.

Clinical Features and Symptoms

The most significant feature of Goodpasture syndrome is pulmonary hemorrhage leading to hemoptysis (Huart et al., 2016). Severity can range from coughing up specks to large volumes of blood. Approximately 80%-90% of patients will present with rapidly progressive glomerulonephritis, whereas 40%-60% of patients will experience alveolar hemorrhage (McAdoo & Pusey, 2017). Dyspnea and chest pain may also precede. Most affected individuals experience hematuria; however, glomerulonephritis and acute renal failure can develop weeks to years after the onset of pulmonary symptoms (Henderson & Salama, 2017). Approximately 60%-80% of patients with Goodpasture syndrome develop clinical manifestations of both pulmonary and renal disease, whereas 20%-40% experience disease limited to the kidneys. Fewer than 10% have pulmonary disease alone (Stojkovikj et al., 2016). The occurrence of clinical and pathological features observed in patients is summarized in Figure 3.
Figure 3. Summary of Major Clinical and Immunological Features of Patients with Goodpasture Syndrome

Note. A range of symptoms from patients was observed, as were the presence of anti-GBM autoantibodies and proteinuria as a sign of renal failure. From "Anti-Glomerular Basement Membrane Disease," by S. McAdoo and C. Pusey, 2017, Clinical Journal of the American Society of Nephrology, 12(7), 1162-1172 (https://doi.org/10.2215/CJN.01380217).

Diagnosis

Diagnosis of Goodpasture syndrome involves the accumulation of results from serology, renal biopsy, and chest X-ray to create an overall indication of the patient’s state. Other tests, such as respiratory function testing, are not routinely performed but may be used to indicate signs of hemorrhage in the lungs. The disease can become rapidly progressive and fatal if identification and subsequent treatment are delayed. Results must therefore be quickly obtained to ensure the best possible outcome for the affected individual.

Serology Test

Detection of anti-GBM autoantibodies is the most common method of diagnosing Goodpasture syndrome. Procedures including indirect immunofluorescence testing and direct enzyme-linked immunosorbent assay
ELISA are used to detect the autoantibodies and confirm diagnosis (Zhong et al., 2020).

Assays generally have high specificity and sensitivity, although 5%-10% of patients showing anti-GBM antibodies from renal biopsy had false-negative serology results when using the assay detection method (Gulati & McAdoo, 2018). According to a study by Gulati and McAdoo (2018), this can result from intrinsic sensitivity of the assay, antibody disappearance before clinical resolution, or high-affinity antibodies being rapidly removed from circulation. Similarly, a study by Henderson and Salama (2017) found that the ELISA method showed a high sensitivity of 95%-100% but a varying specificity of 90%-100%. Serologic testing is an urgent laboratory test and can produce tangible results within 24 hours. A patient affected by Goodpasture syndrome can deteriorate quickly, so rapid diagnosis is essential to early treatment. Approximately 10% of patients with Goodpasture syndrome do not always have identifiable antibodies with conventional assays, however, so serologic testing should not be the only diagnostic procedure performed (McAdoo & Pusey, 2017). If the patient presents with signs of glomerulonephritis but no anti-GBM antibodies are detected, a renal biopsy will be performed.

**Renal Biopsy**

Because of a higher risk of damage, lung biopsy samples are not regularly used to diagnose Goodpasture syndrome. Instead, renal biopsy, in combination with a serum assay for anti-GBM antibodies, is preferred and provides a higher yield of sample. The procedure involves collecting a sample of kidney tissue that is examined using both light and immunofluorescent microscopy to reveal structural or necrotic changes in the glomeruli and lines of antibodies attached to the basement membrane, as seen in Figure 4 (Gulati & McAdoo, 2018). Immunofluorescent staining of tissue reveals a linear deposition of IgG antibodies along the alveolar and glomerular capillaries. Such findings are an indication of Goodpasture syndrome. Crescent formation in the glomerular tissue is a histopathologic effect of anti-GBM disease. Findings from a study by McAdoo and Pusey (2017) suggest that 95% of affected individuals have crescent formation of the glomeruli revealed by a kidney biopsy. The number of observed crescents in the tissue correlates with the level of damage to the kidney.
Figure 4. Kidney Biopsy Sample Using Light Microscopy and Immunofluorescence Microscopy for IgG Antibodies


Chest X-ray

A crucial step in the diagnosis of Goodpasture syndrome is detecting the presence of hemorrhage within the lung. This is achieved by performing an X-ray or CT scan of the thorax. In about 80% of affected individuals, chest X-rays reveal bilateral frosted-glass opacities within alveolar regions, as shown in Figure 5 (Marques et al., 2020). The costophrenic angles and apices of the lungs usually remain unaffected. Where diagnosis of lung hemorrhage remains uncertain, a bronchoscopy is performed. Diagnosis is based on the absence of bronchial lesions which may have led to the alveoli filling with blood from erosion (Marques et al., 2020). Identification of pulmonary hemorrhage via X-ray is widely accepted, although appearance overlaps pulmonary edema (Li et al., 2020), so in the diagnosis of Goodpasture syndrome, this method must be included with a respiratory function test and antibody detection.
The main issue regarding diagnosis of this pathology is an extensive timeframe, resulting in delayed treatment. As each test cannot provide a definitive diagnosis individually, gathering multiple test results before determining the patient’s overall state is timely where disease can rapidly progress. Detection of anti-GBM antibodies using the serology technique is perhaps the most important step in diagnosing Goodpasture syndrome. Despite this technique’s rapid completion time, the occurrence of false-negative results reduces its overall reliability, and it cannot be used as a sole diagnostic procedure (McAdoo & Pusey, 2017). As laboratory antigen-detection techniques continue to advance, however, the quality of such findings is set to improve for future patients.

**Respiratory Function Testing**

Patients affected with Goodpasture syndrome will produce a quantitative change in breathing pattern when assessed by a clinical physiologist using a respiratory function test, although this test is not routinely performed. By incorporating a standard breathing technique, estimations of diffusing capacity after
the inhalation of carbon monoxide can be used to monitor the progression of intrapulmonary hemorrhage (Ewan et al., 2019). In the test, the patient inhales a gas mixture, which diffuses into the pulmonary capillaries from the alveoli. Carbon monoxide uptake depends on the availability of hemoglobin for binding and is represented by the rate constant for carbon monoxide transfer (kCO). When hemorrhaging has occurred, the alveoli fill with extravascular blood, which allows the uptake of additional carbon monoxide (McAdoo & Pusey, 2017). As the diffusion of gas from alveolar space to pulmonary capillary is the rate-determining step, the removal of this process increases the efficiency by which carbon monoxide binds to hemoglobin, overestimating the rate of transfer. The kCO will therefore appear abnormally high when corrected for the reduced hemoglobin levels (Martínez-Martínez et al., 2017).

A study published in the *British Medical Journal* performed the respiratory function test on 11 patients with Goodpasture syndrome to determine their kCO values in comparison to predicted normal values based on height, gender, and ethnicity. As shown in Table 1, kCO values corrected for hemoglobin levels were greater than predicted values for most patients (Ewan et al., 2019). This shows that, while patients are anemic, high levels of gas diffusion are taking place inside the lung due to pooling of blood from alveolar hemorrhage. Measuring carbon monoxide intake with the respiratory function test delivers a noninvasive and repeatable indication of lung hemorrhage as seen in Goodpasture syndrome (McAdoo & Pusey, 2017).
Table 1. kCO and Hemoglobin Measurements from Respiratory Function Testing for Patients With Goodpasture Syndrome

<table>
<thead>
<tr>
<th></th>
<th>Patient</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Hemoglobin (g/dl)</td>
<td>6.7</td>
</tr>
<tr>
<td>Predicted normal kCO (%sec⁻¹)</td>
<td>8.5</td>
</tr>
<tr>
<td>kCO hemoglobin corrected (%sec⁻¹)</td>
<td>15.0</td>
</tr>
<tr>
<td>% kCO change</td>
<td>176</td>
</tr>
</tbody>
</table>

Treatment

Multiple treatment options for Goodpasture syndrome target the anti-GBM antibodies that destroy the alveolar and glomerular basement membrane in an autoimmune response. Despite the rarity of this disorder, options to treat both aggressive and nonaggressive stages have become available because of recent advancements in biomedical technology. Management must immediately follow diagnosis, as the disease progresses rapidly and can become fatal. Rapid implementation of appropriate treatment initially involved discriminating Goodpasture syndrome from other similar pulmonary syndromes. Starting therapy despite a negative result for anti-GBM antibodies is still essential, as delay can lead to progression of the disease. Therapies currently used include plasmapheresis and immunosuppression involving corticosteroid medications. More-severe stages may require dialysis and organ transplantation. Treatment for Goodpasture syndrome involves three goals: rapid removal of autoantibodies from circulation, prevention of further autoantibody production, and removal of potentially harmful agents produced from the initial immune response (Nasser & Cottin, 2018).

Plasmapheresis

The rapid removal of circulating autoantibodies is achieved by plasmapheresis. Even if the patient has not yet been diagnosed with Goodpasture syndrome, plasmapheresis can commence following a case of severe pulmonary hemorrhage (Greco et al., 2015). The role of plasmapheresis is to remove unwanted toxins such as anti-GBM antibodies from the blood, thus eliminating the risk of an autoimmune response in the alveoli. In the procedure, blood is removed from the patient and the cellular contents and plasma are separated. Donated human-plasma replaces the patient’s-original plasma, and then the blood with the donated plasma is transfused back into the patient (Saladi et al., 2018). If a bleeding risk occurs after pulmonary hemorrhage or renal biopsy in patients, fresh frozen plasma is added to maintain normal coagulation. Plasma exchanges of about four liters are performed daily for two to three weeks or until anti-GBM antibodies are no longer detected (Pedchenko et al., 2018). Despite the usefulness of plasmapheresis in ridding the blood of harmful autoantibodies, research regarding side effects and overall efficacy is ongoing.

A study published in 2014 involving 28 patients diagnosed with Goodpasture syndrome compared the plasma-exchange methods of plasmapheresis
and immunoadsorption. Findings show that both methods produced similar levels of effectiveness, with 59.0% and 71.2% efficacy respectively, although only 62.7% of patients undergoing plasmapheresis had reduced IgG, compared to 83.5% of those undergoing immunoadsorption (Pedchenko et al., 2018). Despite these differences, the plasmapheresis group overall experienced fewer plasma-associated side effects, making plasmapheresis a safer method of plasma exchange.

Immuonoadsorption is another method of serum-antibody removal that is used for multiple autoimmune diseases involving antibody-mediated rejection (Saladi et al., 2018). Advocates have proposed immunoadsorption as a therapy in anti-GBM disease because of its high-affinity binding of IgG antibody subclasses and redundancy of fresh frozen plasma. With this technique, about 87% of IgG antibodies can be removed from 2.5 plasma volumes in one session. More than 98% of IgG can be cleared after multiple sessions (Henderson & Salama, 2017). Plasmapheresis is still the preferred therapy for Goodpasture syndrome, as the involvement of immunoadsorption in anti-GBM disease is yet to be properly evaluated. Minor case studies of the immunoadsorption method among 10 patients showed that antibody levels declined approximately 71%-84% after the initial therapy (Henderson & Salama, 2017). This is an improvement on report cases involving the plasmapheresis method, but it is not comparable to studies involving larger numbers of patients. More investigation into the large-scale effectiveness of immunoadsorption is therefore necessary before immunoadsorption can become the preferred plasma-exchange treatment for Goodpasture syndrome.

**Immunosuppression**

The prevention of ongoing anti-GBM antibody production is achieved via immunosuppression drugs such as corticosteroids and cyclophosphamide. It can also prevent rebound hypersynthesis of antibodies after the cessation of plasmapheresis (Greco et al., 2015). Reduction of the body’s immune response reduces activity of circulating antibodies, including the anti-GBM antibodies, long term. This helps lower the occurrence of autoantibody-induced destruction of the basement membranous structure in the alveoli and glomeruli. Reduction in the body’s ability to produce an immune response consequently leads to a higher risk of viral and bacterial infections within patients (Gulati & McAdoo, 2018). Corticosteroids and cyclophosphamide are initially administered for 6 and 3 months, respectively, to reduce immune activity and maintain leucocyte levels in the blood (Greco et al., 2015). Therapy durations are not well defined, so the presence of anti-GBM antibodies is regularly monitored. Individuals with
serologically active disease require an extended period of immunosuppression at around 6-9 months. Oral cyclophosphamide is used, as a 2012 study showed that intravenously administered cyclophosphamide was linked with increased mortality in patients with Goodpasture syndrome (Gulati & McAdoo, 2018). Both rituximab and mycophenolate are used as alternative options in treating anti-GBM antibody disease if cyclophosphamide use fails or yields severe side effects (Henderson & Salama, 2017). The anti-GBM antibodies become undetectable but produce variable renal outcomes.

Following immunosuppressive treatments, adverse side effects, including interstitial pneumonitis, erythrocyte aplasia, and thrombocytopenia, are common (Diehl et al., 2017). The consumption of such drugs therefore must be carefully planned and monitored by a healthcare professional throughout the course of treatment. Dosing of 375 milligrams/m² intravenously once weekly is recommended for rituximab, 0.5-3 grams/day for mycophenolate, and 500-1000 milligrams/m² intravenously monthly for six doses, plus corticosteroids for cyclophosphamide (Medscape, 2021a, 2021b, 2021c). Anti-GBM antibody titers must be monitored regularly. Once antibody becomes consecutively undetectable, plasmapheresis is terminated. Antibody titer procedures are performed monthly as a follow-up for at least six months. Corticosteroids are continued for a minimum of six months, and cyclophosphamide for three months to help eliminate any persisting anti-GBM antibody levels in patients (Segelmark & Hellmark, 2019).

Substantial evidence supports the effectiveness of plasmapheresis and immunosuppression as methods for treating Goodpasture syndrome. One five-year study demonstrated that immunosuppressive methods combined with plasmapheresis were more effective than was immunosuppression alone (Chung, 2019). This is because anti-GBM antibodies are more efficiently removed, allowing for faster patient recovery and less tissue damage. A research study conducted on the molecular basis of Goodpasture syndrome compared the effects of immunosuppression therapy with and without the use of plasmapheresis and found the mortality rate decreased 11% with the combined treatment (Betsy et al., 2019). This shows that although each treatment method is effective in alleviating the consequences of the autoimmune response, the best possible outcome for the patient is achieved when both therapies are employed simultaneously.

**Dialysis**

The final goal in treatment for Goodpasture syndrome is the removal of potentially harmful agents generated from the initial immune response. This
involves dialysis or the removal of excess toxins and solutes from the blood. Because of glomerular membrane damage, patients need regular checks to monitor renal function (Saladi et al., 2018). If the kidney can no longer function to an adequate level, dialysis will be performed indefinitely as a replacement therapy. Referral for organ transplantation may additionally be recommended, depending on suitability and availability. End-stage renal disease is still possible after renal dialysis, though only about 30% of surviving patients require this treatment method (Pedchenko et al., 2018).

**Organ Transplantation**

In the most severe cases of Goodpasture syndrome, affected individuals may receive lung or kidney transplants, depending on the level of original function lost. Patients who receive transplants can have a recurrence of anti-GBM antibody levels, causing relapse (Gulati & McAdoo, 2018). A patient should therefore wait about six months after plasmapheresis and immunosuppression before receiving a transplant, to ensure seronegativity with anti-GBM antibodies. This will reduce the risk of damage to the transplanted organ. Relapses are uncommon, however, and are associated with prolonged exposure to environmental factors such as irritants and smoking (Rohm et al., 2019).

**Future Directions**

There are several emerging directions for the management and treatment of Goodpasture syndrome. As research on this rare disease progresses, new mechanisms for understanding pathogenesis and treatment are under development. Collagen research and novel methods of plasma exchange are some of the areas currently under investigation.

Research regarding the type IV collagen family provides new information about the structural development of basement membranes in the alveoli and glomeruli. By understanding the biochemical nature of these collagen molecules acting as antigens in the autoimmune response, the molecular understanding of protomer assembly in the pathogenesis of Goodpasture syndrome can be better understood (Hudson et al., 2019). Certain developing insights suggesting membranous susceptibility to proteolysis can lead to the development of certain protease inhibitors and genetic therapy. Experimental models of this disease predict the potential use for costimulatory blockade of T-cell activation or prevention of
macrophage migration (Hudson et al., 2019). This work will provide a new understanding of collagen-related diseases such as Goodpasture syndrome.

More specific methods of plasmapheresis are being investigated for the treatment of Goodpasture syndrome. Some of these include immunoadsorption, cryofiltration, and enzymatic degradation. Current methods of plasma exchange are useful for removing anti-GBM antibodies that are in circulation but not those bound to tissue (Henderson & Salama, 2017). New approaches to this problem are under investigation, including exploration of the properties of proteins S. pyogenes (IdeS) and endoglycosidase (EndoS), which are secreted by the bacteria Streptococcus pyogenes. Human IgG antibodies are cleaved by IdeS, which produces monomeric Fc fragments (Henderson & Salama, 2017). This hinders the antibody’s ability to function. An experimental model of glomerulonephritis demonstrated the successful removal of glomerular Fc fragments of anti-GBM by IdeS, showing its effectiveness as an immunosuppressant. IdeS trials have progressed mainly in kidney transplantation but currently have directed clinical trials toward anti-GBM disease (Li et al., 2020). As previously mentioned, immunoadsorption removes circulating anti-GBM antibodies with the use of sepharose-coupled antihuman IgG but needs comprehensive validation before becoming an accepted technique.

Conclusion

Despite its rarity, Goodpasture syndrome has become increasingly better understood since its initial description in 1958. Given the discovery of antibodies and advancements in the field of molecular biology, affected individuals are able to be diagnosed and treated through various methods. Effective patient management is frequently achieved. The production of false-negative serology results reduces the overall reliability of the diagnostic procedure, however. As the disease can progress rapidly, an accurate diagnosis allowing for immediate treatment is essential to prevent any detrimental consequences in the patient’s health. Immunoadsorption and other alternate methods of treatment are being researched and undergoing trials, with the potential of improving therapy efficacy. Research into the type IV collagen family is also being conducted to improve overall molecular understanding of the pathogenesis and progression of the disease process. This may result in significant changes to the approach of disease management and may improve patient recovery for future practice.
References


glomerulonephritis. *Medicine, 97*(23), 10954–10957. https://doi.org/10.1097/MD.0000000000010954


COMPLEX REGIONAL PAIN SYNDROME: A SCHOLARLY REVIEW

JOSEPH O'BRIEN, UNIVERSITY OF RHODE ISLAND
MENTOR: GUY W. FRIED

Abstract

Complex regional pain syndrome (CRPS) is a mystifying, often disabling, neurogenic pain disorder affecting millions of individuals. Although this condition was described in medical literature as early as 1864, CRPS remains controversial, often misunderstood, and frequently incorrectly diagnosed and treated. Patients afflicted with CRPS often suffer unnecessarily because of these factors. This paper reviews the history of CRPS, the evolution of its diagnosis and treatment, and the leading theories on its pathogenesis. This paper also includes case studies of two female patients currently suffering from CRPS. Their stories demonstrate some of the unique challenges confronting CRPS patients, including medical, financial, and health insurance issues.

Complex regional pain syndrome (CRPS) is a life-altering, mystifying neurologic pain condition that arises from a traumatic insult to an extremity or peripheral nerve (2). This condition is not widely understood by medical professionals. No single diagnostic tool definitively confirms a CRPS diagnosis. There is also no universally effective treatment for this condition (1, 2, 6). Because it is not fully understood, this condition can cause confusion and uncertainty among patients and treating physicians. Patients complaining of CRPS pain can sometimes be dismissed as drug-seeking (15). Signs and symptoms of CRPS frequently overlap with those of other chronic pain conditions, which can lead to misdiagnosis and underdiagnosis of CRPS (9). The lack of a single, definitive treatment for CRPS can lead to a long, frustrating clinical path for both patients and clinicians. Even when CRPS is diagnosed timely and treatment is initiated appropriately, the prognosis remains uncertain (28). This little-understood condition profoundly affects the quality of life of patients of all ages and walks of life. This review examines the existing medical literature regarding CRPS and also includes case studies on two patients currently suffering from the condition.
History

Doctors have described signs and symptoms consistent with CRPS since the 16th century. During the American Civil War, physician Silas Weir Mitchell recorded what is now thought to be the earliest careful, thorough description of CRPS. He described patients with complaints of severe burning pain accompanied by red, glossy skin (2). In his 1864 book *Gunshot Wounds, and Other Injuries of Nerves*, Mitchell first used the term *causalgia* for this condition, from the Greek words *causos* (heat) and *algia* (pain; 2). Mitchell noted that all of the patients who exhibited these particular signs and symptoms had suffered peripheral nerve injuries (4).

The term *causalgia* was used as the diagnostic term for several decades for patients exhibiting these signs and symptoms (3). Then the term *reflex sympathetic dystrophy* (RSD) was coined by Massachusetts physician James A. Evans (4). Between 1946 and 1947, Evans described signs and symptoms exhibited by several of his patients that were very similar to those described by Mitchell (4). Evans chose the term *reflex sympathetic dystrophy* because he believed that reflexive abnormalities of the sympathetic nervous system, not necessarily a discrete injury to a peripheral nerve, caused the characteristic intense chronic pain as well as diaphoresis, atrophy, and red, inflamed skin (4). He noted that signs and symptoms characteristic of “causalgia” resulted primarily from fractures, sprains, and vascular complications rather than from direct insults to peripheral nerves, and he concluded that a different name for the condition was therefore appropriate. Thus, the term RSD came to largely replace the term *causalgia* (4).

This condition was destined to undergo yet another name change. In 1973, the International Association for the Study of Pain (IASP) was founded in order to standardize terminology used to classify chronic pain disorders, including RSD (3). In 1993, the IASP organized a consensus conference in Orlando, Florida, to devise diagnostic criteria specifically for RSD that would serve as the “gold standard” for clinical diagnosis (4). Before any criteria were drafted, however, it was agreed to change the name from *reflex sympathetic dystrophy* to *complex regional pain syndrome* (3, 4). The reason for the name change was that the term *reflex sympathetic dystrophy* implied a precipitating injury to a peripheral nerve, but clinicians recognized that many patients with RSD had not suffered a major nerve injury prior to the onset of pallor, sweating, edema, and other symptoms (4). The Orlando conference thus adopted the terms *CRPS-1* and *CRPS-2*. CRPS-1 replaced the term *RSD* for patients with no confirmed nerve injuries, and CRPS-2 replaced the term *causalgia* for patients with confirmed nerve injuries (4).
The Orlando consensus conference created the following criteria for the clinical diagnosis of CRPS:

1. A noxious event or immobilization able to start the process
2. Allodynia, hyperalgesia, or any pain out of proportion compared to the precipitating event
3. Presence of edema, changes in skin blood flow, or abnormal sudomotor activity of the affected region in any stage of the disease process

Notably, the diagnosis of CRPS can be excluded if the presence of this kind of pain and dysfunction can be related to other diseases.

The diagnostic criteria established at the Orlando conference are still used today. Those diagnostic criteria were purely subjective and based only on reported symptoms, however, which led to high rates of misdiagnosis and false positives (less than 50% specificity; 3). As a result of the vague nature of the original diagnostic criteria put into place by the Orlando conference, some patients diagnosed with CRPS were in fact suffering from other medical conditions. Another conference was thus held, in Budapest in 2003, with the goal of creating new and improved diagnostic criteria and classification systems for CRPS (4), and these have since been subscribed to by a majority of medical professionals:

1. Continuing pain which is disproportionate to any inciting event
2. Must report at least one symptom in three of the four following categories:
   - Sensory: reports of hyperesthesia and/or allodynia
   - Vasomotor: reports of temperature asymmetry and/or skin color changes and/or skin color asymmetry
   - Sudomotor/edema: reports of edema and/or sweating changes and/or sweating asymmetry
   - Motor/other: reports of decreased range of motion and/or motor dysfunction (weakness, tremor, dystonia) and/or trophic changes (hair, nail, skin)
3. Must display at least one sign at time of evaluation in two or more of the following categories:
   - Sensory: evidence of hyperesthesia (to pinprick) and/or allodynia (to light touch and/or deep somatic pressure and/or joint movement)
   - Vasomotor: evidence of temperature asymmetry and/or skin color changes and/or asymmetry
   - Sudomotor/edema: evidence of edema and/or sweating changes and/or sweating asymmetry
   - Motor/other: evidence of decreased range of motion and/or motor dysfunction (weakness, tremor, dystonia) and/or trophic changes (hair, nail, skin)
4. There is no other diagnosis that better explains the signs and symptoms

These diagnostic criteria are similar in nature to those drafted at the Orlando conference. As the name implies, however, the Orlando consensus conference
Criteria were entirely consensus-driven and resulted in low specificity rates (3). By contrast, the Budapest criteria were research-driven and provided significantly higher specificity rates (69%; 3). Since 2003, the revised Budapest diagnostic criteria for CRPS have been accepted by the IASP (3). These criteria have been increasingly utilized by medical professionals, as they allow for effective CRPS diagnosis and research (3). The result has been a reduction in the incidence of misdiagnosis and an increase in accurate diagnosis and initiation of timely treatment.

**Clinical Presentation**

CRPS can occur in any person at any age, although females are at the highest risk for developing the condition (6). The exact reasons for this phenomenon are still unknown, but studies suggest that hormonal factors may be to blame (6). CRPS occurs mainly in distal extremities and can be caused by a variety of precipitating events (1). The most common causes of CRPS-1 (no apparent inciting nerve injury) include minor fractures, sprains, and dislocations (1). The immobilization (i.e., casting) period following one of these injuries has also been shown to increase the likelihood of CRPS development (6). CRPS-2 results from trauma to a peripheral nerve. Elective surgeries, myocardial infarction, and stroke have also been shown to cause CRPS. CRPS occurs more frequently in upper extremities than in lower extremities, and it typically affects only one limb (2).

The most common tell-tale symptom of CRPS is severe, burning pain (1). This pain is constant, chronic in nature, and typically disproportionate to the inciting injury or event (6). According to the McGill pain index, patients with CRPS experience the highest-ranking, most painful form of chronic pain (worse than amputation of a digit or unprepared childbirth). This pain is often debilitating and may cause a number of secondary conditions (discussed below). According to the Budapest diagnostic criteria, other common signs and symptoms of CRPS include inflammation, allodynia, hyperalgesia, edema, hyperhidrosis, skin-temperature abnormalities, and trophic changes (hair, nails, or skin). Several images of these signs and symptoms are shown below.
CRPS may progress through three distinct phases: the acute/warm phase, intermediate phase, and chronic/cold phase (7). The acute/warm phase is characterized by burning pain and localized inflammation (7). Inflammation results in redness and swelling of the skin, which leads to an increase in skin temperature. Stiffness, decreased range of joint motion, and unusual hair and/or nail growth also accompany this phase (6). This acute phase lasts for roughly one to three months following the initial onset of CRPS (6). An accurate diagnosis and proper treatment during this crucial initial phase result in the most favorable patient outcomes (6). The longer the condition goes undiagnosed and untreated, the less favorable the prognosis.

The intermediate phase is defined by a reduction in inflammation and, in some cases, a return to normal skin temperature (7), although pain continues to
worsen and muscles begin to atrophy (6). This phase lasts from three to six months following the onset of CRPS (6). During this time, the adverse effects of CRPS can still be remedied or reversed by effective treatment.

Studies have shown that a decrease in norepinephrine levels accompanies the acute phase of CRPS (6). The decrease in norepinephrine leads to vasodilation and increased skin temperature in the affected region. If the CRPS is left untreated, the patient’s body eventually develops increased peripheral catecholamine sensitivity to compensate for this physiological change (2). This causes spasms in the affected peripheral musculature, as well as excessive vasoconstriction (6). Vasoconstriction causes skin in the surrounding area to become cyanotic and cold to the touch, which is a characteristic sign of the chronic, or cold, phase of CRPS. The chronic phase of CRPS is the most severe, clinically significant, phase of the condition (6). During this phase, pain spreads through the entire affected limb, and irreversible tissue damage may occur (6). Motor impairments (e.g., reduced hand function) begin to arise, causing a marked decrease in overall quality of life (6). It is important to note that not every patient will exhibit signs and symptoms indicative of each of the three major phases of CRPS (6), as the phases and rate of progression of CRPS are variable.

**Diagnosis and Diagnostic Tools**

As stated above, the currently accepted clinical diagnostic criteria for CRPS are the Budapest criteria, which are shown again below.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Continuing pain, which is disproportionate to any inciting event</td>
</tr>
<tr>
<td>2.</td>
<td>Must report at least one symptom in three of the four following categories:</td>
</tr>
<tr>
<td></td>
<td>• Sensory: reports of hyperesthesia and/or allodynia</td>
</tr>
<tr>
<td></td>
<td>• Fascolatary: reports of temperature asymmetry and/or skin color changes and/or skin color asymmetry</td>
</tr>
<tr>
<td></td>
<td>• Sudekum: reports of edema and/or swelling changes and/or swelling asymmetry</td>
</tr>
<tr>
<td></td>
<td>• Motor trophe: reports of decreased range of motion and/or motor dysfunction (weakness, tremor, dystonia) and/or trophe changes (hair, nail, skin)</td>
</tr>
<tr>
<td>3.</td>
<td>Must display at least one sign at time of evaluation in two or more of the following categories:</td>
</tr>
<tr>
<td></td>
<td>• Sensory: evidence of hyperalgesia (to pinprick) and/or allodynia (to light touch and/or deep somatic pressure and/or joint movement)</td>
</tr>
<tr>
<td></td>
<td>• Fascolatary: evidence of temperature asymmetry and/or skin color changes and/or asymmetry</td>
</tr>
<tr>
<td></td>
<td>• Sudekum: evidence of edema and/or swelling changes and/or swelling asymmetry</td>
</tr>
<tr>
<td></td>
<td>• Motor trophe: evidence of decreased range of motion and/or motor dysfunction (weakness, tremor, dystonia) and/or trophe changes (hair, nail, skin)</td>
</tr>
<tr>
<td>4.</td>
<td>There is no other diagnosis that better explains the signs and symptoms</td>
</tr>
</tbody>
</table>
For a confident diagnosis of CRPS, patients must display at least one symptom in three or more of the symptom categories and at least one sign in two or more of the sign categories. These diagnostic criteria are widely utilized by clinicians and serve as the current standard in CRPS diagnosis (2).

CRPS diagnosis can be difficult and is largely exclusionary (8). This means that conditions comprising the common differential diagnoses of CRPS must be ruled out before a CRPS diagnosis can be made definitively. Some of these conditions include neuropathy, erythromelalgia, cellulitis, lymphedema, deep vein thrombosis, vasculitis, and Raynaud’s syndrome (8). Because of the overlap of signs and symptoms among these conditions, CRPS has been cited as one of the most underdiagnosed and misdiagnosed conditions in medicine (9).

There is no single gold-standard test for definitive diagnosis of CRPS, as the diagnosis is largely clinical (1). Several objective tests can support the diagnosis, however, including triple-phase bone scan (TPBS), MRI, thermography, and sweat-production test (1). The TPBS is a generally accepted diagnostic tool (1); a TPBS performed on a patient suffering from CRPS will show certain characteristic abnormalities that clinicians use to confirm the clinical diagnosis of CRPS. The TPBS utilizes a radioactive tracer, so if a patient does have CRPS, this test provides clinicians with images showing increased perfusion and unequal blood pooling in the affected extremity (1).

MRIs are also used by clinicians as a diagnostic tool for CRPS, as they can reveal a multitude of tissue abnormalities common among patients with CRPS (1). Some of these abnormalities include soft-tissue edema, skin thinning/thickening, and, in severe cases, muscle atrophy (12). According to a study by M. E. Schweitzer and colleagues (12), MRIs can be used to distinguish between the three common phases of CRPS, particularly the acute and chronic phases. The authors found that in 35 patients with acute-phase CRPS, 31 exhibited skin thickening. No muscular atrophy was observed in the acute-phase cohort, but of five patients with chronic-phase CRPS, four exhibited muscular atrophy (12). This study by Schweitzer and colleagues provides evidence that MRIs can be useful in differentiating between distinct stages of CRPS, although conflicting evidence exists regarding the efficacy of MRIs in the initial clinical diagnosis of CRPS (1).

Thermography tests are highly accurate diagnostic assessments that easily detect the temperature asymmetry of the skin that is seen in patients with CRPS (11). The findings presented in a study by Gulevich et al. and reported by Sjoerd Niehof and colleagues (11) show that the sensitivity and specificity of infrared thermography used in the diagnosis of CRPS were 93% and 89%, respectively (11).
Furthermore, Niehof and colleagues sought to examine the effect of alterations to ambient temperature on the sensitivity and specificity of thermography tests used to aid in the diagnosis of 12 CRPS patients (11). They found that the patients’ affected and unaffected extremities did not exhibit optimal obtainable temperature differences at room temperature but exhibited much more pronounced temperature differences at warm and cold ambient temperatures (11). These findings suggest that alterations in ambient temperature act to further improve the sensitivity and specificity of thermography tests in the diagnosis of CRPS.

Along with skin-temperature abnormalities, abnormal diaphoresis commonly accompanies CRPS (12); sweat-production tests are thus commonly used to aid in the diagnosis of CRPS. These tests measure the levels of sweat on two limbs and may indicate a case of CRPS if the sweat levels are unequal (1).

Pathogenesis

The pathogenesis of CRPS is complicated and multifaceted. During the early phases of CRPS, exacerbation of neurogenic inflammatory mechanisms occurs (13). CRPS-related neurogenic inflammation occurs when Group C nerve fibers (essential in nociception) become continuously strained (13). Several neuropeptides—importantly, calcitonin gene-related peptide (CGRP)—are released in response to C-fiber stimulation (13). CGRP increases vasodilation while promoting sweat-gland activation and hair growth (13). Exacerbation of neuroinflammatory mechanisms with subsequent release of CGRP in patients with CRPS likely explains the characteristic increase in skin temperature, edema, and increased hair and nail growth seen in the early phases of CRPS.

Interestingly, the use of angiotensin-converting enzyme (ACE) inhibitors has been shown to increase the risk for CRPS development (14). ACE acts to inhibit CGRP; thus, the inhibition of ACE further enhances release of CGRP and causes neurogenic inflammatory effects (14). A study by de Mos et al. (14) revealed that a high dosage and prolonged use of ACE inhibitors further increases the incidence of CRPS.

Another key feature of the pathogenesis of CRPS is the onset of peripheral sensitization and, in some cases, central sensitization (15). In general, sensitization is defined as an increased responsiveness of neurons to normal input or recruitment of a response to subthreshold inputs (15). Sensitization arises when a continuous noxious stimulus, such as the chronic pain of CRPS, causes afferent neurons (and other nociceptive mechanisms) to exhibit heightened sensitivity over time (15).
This causes the brain to interpret previously nonpainful stimuli, such as light touch, as painful, which is indicative of primary hyperalgesia/allodynia (16). If left unchecked, peripheral sensitization itself can lead to central sensitization (15). According to a study by Latremoliere and Woolf, a noxious stimulus must be severe, recurrent, and sustained in order to induce central sensitization. This means that the abnormally high pain signals from CRPS cause peripheral sensitization, which in turn can lead to central nervous system (CNS) sensitization. Once the CNS becomes abnormally sensitized because of CRPS, a patient can experience severe pain in response not only to light touch but also to stimuli such as sound, light, and alterations in barometric pressure (15).

Glial cells of the CNS play a major role in the onset of central sensitization (15). Glial cells normally function in the CNS to form myelin and to provide support and protection for neurons (15). These cells are normally dormant in the CNS, but during central sensitization, they become active and produce inflammatory mediators (called cytokines), which in turn cause nerve inflammation (15). Two of these inflammatory mediators are tumor necrosis factor alpha (TNFα) and interleukin-1β (IL-1β). TNFα and interleukin-1β act on the dorsal horn of the spinal cord, which contains cell bodies of sensory neurons (16). The action of these inflammatory mediators on the dorsal-horn neurons may result in secondary hyperalgesia, which is defined as a spread of pain hypersensitivity past the original point of peripheral injury (16).

NMDA receptors—membrane-bound ion (importantly, calcium) channels that are numerous in the CNS and found on sensory neurons (15)—are also crucial in the development and maintenance of central sensitization (31). Peripheral sensitization and inflammation can alter the properties of these receptors and, in turn, the sensory neurons that they mediate (15). Typically, the calcium-channel gates of NMDA receptors are blocked by magnesium, but this magnesium block is released during the central sensitization process, resulting in a rapid influx of calcium ions and subsequent sensitization of neurons (31).

Evidence suggests that people with CRPS experience certain perceptual abnormalities. This may be caused by a phenomenon known as brain plasticity, or neuroplasticity, which is the ability of the brain to “rewire” itself and its connections (16). This is what allows the brain to develop from infancy to adulthood (16). In people with CRPS, this reorganization of the nervous system (particularly the brain cortex) is thought to occur in response to continuous painful stimuli (16). Several of the perceptual abnormalities resulting from brain plasticity in CRPS patients include an inability to identify objects by touch without visual
input, finger proprioception abnormalities, abnormal body scheme, and defects in the mental representation of the affected limb (16). According to Kuttikat et al. (16), one potential cause of these perceptual disturbances is reorganization of the cerebral cortex. This region of the brain functions in sight, hearing, and memory as well as sensory and motor function; thus, cortical reorganization may disrupt typical somatosensory functioning and lead to the sensory and motor impairments noted above (16). A study by Cohen et al. (17) examined 22 patients with chronic CRPS-1 to assess for parietal-lobe dysfunction. They found that 15 of the 22 patients had some form of parietal-lobe dysfunction and exhibited symptoms such as sensory extinction and dysynchria. They also found that increased parietal-lobe dysfunction was correlated with increased body-surface allodynia (17). Although studied extensively, the exact physiological underpinnings leading to nervous-system reorganization in CRPS patients are still unknown. This phenomenon greatly affects activities of daily living and can result in lifelong somatosensory dysfunction (16).

Treatment

Although clinicians employ a multitude of treatments for CRPS, the efficacy of some of these has yet to be established through peer-reviewed research. This section addresses only those treatment options that have been reported as efficacious in peer-reviewed literature. In terms of pharmacological treatment, evidence compiled by Perez et al. (21) suggests that subanesthetic ketamine, gabapentin, dimethylsulphoxide (DMSO), N-acetylcysteine (NAC), and others are at least partially efficacious at alleviating some of the pain experienced by patients with CRPS. Stellate ganglion blocks, spinal cord stimulation, and surgical sympathectomy have also been proven to be therapeutic in relieving CRPS pain (18). Additionally, physical therapy and occupational therapy are effective noninvasive, nonpharmacological treatments for CRPS (18).

As noted above, a number of evidence-based pharmacological treatment options for CRPS exist. It should be noted, however, that the scientific evidence for these treatments is somewhat limited (18). Ketamine infusions are commonly administered to people with CRPS (18). A study by Correll et al. (18) reviewed case notes for 33 CRPS patients who had undergone subanesthetic ketamine infusions for pain relief, in order to determine if ketamine was indeed effective in reducing pain. All 33 patients underwent ketamine infusions at least once; 12 underwent a second course of ketamine therapy, and 2 underwent a third (19). Correll et al. (18) found that after the initial administration of ketamine, 25 patients
reported complete pain relief, 6 reported partial relief, and only 2 reported no relief. The degree of pain relief also appeared to increase with multiple ketamine infusions, as the 12 patients who underwent second and/or third infusions all reported complete pain relief (19). The duration of reported pain relief ranged from three months to three years, and the patients who had undergone more than one infusion reported longer durations of pain relief (19). Although ketamine has been shown to effectively relieve CRPS pain, it may cause significant side effects, which can include hallucinations, dizziness, intoxication, and nausea (18, 19).

Gabapentin, an anticonvulsant, has also been shown to reduce neuropathic pain consistent with CRPS (18). An eight-week study by Serpell and the Neuropathic Pain Study Group (20) assessed average daily pain scores in 305 patients with various neuropathic pain syndromes. Of these 305 patients, 153 were given three separate, progressively increased doses of gabapentin while 152 were given a placebo (20). At the culmination of the eight-week study, it was found that the average daily pain scores of patients treated with gabapentin improved (i.e., decreased) by 21% (20). These findings are promising, but the study’s sample cohort consisted of patients with various neuropathic pain disorders, not only those with CRPS; thus, more evidence is needed to determine how effective gabapentin is in the treatment of CRPS pain specifically (18).

Some studies indicate that the free radical scavengers DMSO and NAC may be effective in treating CRPS patients (18). One study by Perez et al. (21) sampled 146 CRPS patients and sought to compare the effects of these two free radical scavengers on CRPS pain. The patients were divided randomly into two groups and were treated for 17 weeks. One group was treated with 50% DMSO cream five times per day, and the other was treated with 600-mg NAC tablets three times per day (21). Researchers concluded that both treatments were moderately effective in reducing CRPS pain. Interestingly, 50% DMSO cream was more therapeutic for patients with acute/warm-phase CRPS, and the 600-mg NAC tablets were more therapeutic for patients with cold-phase CRPS (21).

Treatment of CRPS with stellate ganglion blocks is widely utilized by medical professionals (15), and studies support the efficacy of stellate ganglion block therapy. One study, by Yucel et al. (22), included 22 patients with CRPS-1 who had suffered either distal radial fractures or soft-tissue hand traumas or had undergone surgery for carpal tunnel syndrome. All patients received three total stellate ganglion blocks, with one week in between each injection, and a 0–10 pain scale was used to evaluate each of the patients before treatment and two weeks after treatment (22). Wrist-joint range of movement (ROM) was also assessed before and
after treatment (22). Stellate blocks were found to significantly improve both CRPS pain and joint ROM in all patients (22). Overall wrist flexion improved from 50.2 ± 13.8 degrees to 69.4 ± 8.1 degrees; overall wrist extension improved from 38.9 ± 12.8 degrees to 58.7 ± 7.8 degrees; and overall supination improved from 41.1 ± 11.6 degrees to 63.1 ± 8.0 degrees (22). Blocks significantly improved pain in all patients as well (p < 0.05; 22). The study by Yucel et al. indicates that stellate ganglion blocks do indeed alleviate CRPS-related pain and improve joint ROM.

Spinal cord stimulation has been shown to reduce pain in individuals with chronic CRPS (18, 23). In a study by Kemler et al. (23), 36 patients with chronic CRPS were divided into two treatment groups, one receiving spinal cord stimulation with physical therapy and the other receiving physical therapy alone. Pain levels were assessed using a visual analog scale, with 0 cm being the lowest level of pain and 10 cm being the highest (23). The researchers found that after six months, patients who received spinal cord stimulation with physical therapy exhibited a mean reduction of 2.4 cm on the visual analog pain scale, whereas patients participating in physical therapy alone exhibited a mean reduction of only 0.2 cm (23). Although spinal cord stimulation significantly reduced pain in chronic CRPS patients, no improvement of functional capacity was noted (18, 23). There is also no evidence of pain reduction due to spinal cord stimulation for patients with non-chronic CRPS (18). It therefore appears that spinal cord stimulation is effective in reducing the subjective pain levels of patients with chronic CRPS, though it has not been shown to increase their level of function.

Surgical sympathectomy, in which a portion of the sympathetic nerve chain is severed in order to prevent nerve signals from passing into a specific area (i.e., the affected limb in CRPS), has been shown to relieve pain in patients with CRPS (18, 24). A study by Duarte et al. (25) sought to examine the effects of endoscopic thoracic sympathectomy (ETS) on pain levels of seven patients with CRPS. An ETS was performed on each of the seven patients, and subsequent pain levels were measured using the 0–10 visual analog scale (25). In all seven patients, resting pain completely disappeared following the ETS (25). Four of the patients reported pain upon movement, although the pain was less severe than it had been prior to surgery (25). All patients reported improvements in quality of life (25). Follow-up studies cited in the literature by Perez et al. (18) determined that pain relief due to surgical sympathectomy declined over time and that patients who received the surgery within three months of the initial inciting event exhibited the most favorable prognosis.
Both physical therapy (PT) and occupational therapy (OT) are increasingly common treatment interventions for patients with CRPS and are typically given in conjunction with medications (18). A study by Oerlemans et al. (26) compared the efficacy of both PT and OT in the treatment of CRPS, assessing a cohort of 135 patients with upper-limb CRPS of less than one year’s duration. These patients were assigned to a group receiving PT, a group receiving OT, or a control group (26). Pain was measured prior to the initiation of treatment using a visual analog scale and the McGill Pain Questionaire (MPQ) and was reassessed after six weeks, three months, six months, and twelve months (26). ROM was also assessed regularly (26). At the conclusion of twelve months, patients who had undergone PT—and, to a lesser extent, OT—reported greater improvements in visual analog pain scores and less pain on the MPQ than did patients in the control group (26). ROM of the upper distal extremities also improved with PT and, to a lesser extent, OT (26). It is evident that both pain and ROM can be significantly improved by PT and OT in patients with upper-limb CRPS. PT and OT have also increasingly become part of the standard course of treatment in patients with CRPS (18). Limited research exists on the benefits of these therapies in patients with chronic CRPS, however, and no studies exist regarding the efficacy of OT and PT in patients with lower-limb CRPS (18).

Prognosis

The prognosis for CRPS patients is variable and uncertain. The most reliable prognostic factor is timing of diagnosis and initiation of treatment. As previously stated, the earlier the condition is diagnosed and treated, the more favorable the prognosis. Studies indicate that some patients with CRPS experience spontaneous remission of symptoms, and in some cases, remission is permanent (28). More commonly, however, patients experience only partial remission, in which signs of CRPS linger (28). Even among patients who experience spontaneous complete remission in symptoms, relapse can occur (28).

Studies are mixed with regard to the prognosis for patients diagnosed with CRPS during childhood and adolescence. Some studies have found that patients with CRPS diagnosed during childhood have more favorable outcomes than do those diagnosed as adults. Other studies have found no significant difference in prognosis for childhood-onset versus adult-onset CRPS. A study by Wilder et al. (29) conducted primary and follow-up assessments of 70 adolescents (mean age of 12.5 years) with CRPS. It was found that younger patients experienced less-severe CRPS symptoms and more-favorable prognoses than did older patients (28, 29).
Additionally, upon follow-up, younger patients exhibited a shorter duration of deleterious symptoms related to CRPS (28, 29).

Several studies contradict the findings of Wilder and colleagues, however. In a study by Tan et al. (30), a quality-of-life survey was given to 42 adults who had been treated for childhood-onset CRPS-1. The average time between childhood diagnosis and survey response was 12 years (30). More than half of the patients reported pain at the time of follow-up, and many of the typical signs and symptoms of CRPS had remained unchanged through the course of the disease (30). Fifteen patients also reported relapse in symptoms (30). These findings suggest that children with CRPS may have a less-favorable prognosis than earlier studies have indicated.

The implications of these latter studies for young people diagnosed with CRPS are extremely dire. The direct and secondary effects of living with chronic pain, taking medications that carry serious side effects and risk of dependency, depression, and other secondary health conditions such as overuse injuries are likely compounded over the course of a lifetime; thus, early diagnosis and initiation of treatment in children with CRPS is paramount for the preservation of long-term quality of life and well-being.

Secondary Health Conditions

CRPS can lead to debilitating secondary conditions that profoundly affect quality of life. Several common secondary conditions related to CRPS include depression, anxiety, sleep disturbances, drug dependency, and overuse injuries in non-affected extremities (15). These secondary conditions are certainly not unique to CRPS, as they are commonly seen in a variety of chronic pain conditions.

Interestingly, secondary conditions were prominent in both patients profiled in the case studies in this paper. Patient 1, who was diagnosed with CRPS more than three years after the initial inciting trauma, has reported significant life-altering secondary conditions as a result of her chronic CRPS. Patient 1 is moderately depressed and unable to carry out previously simple activities of daily living (such as holding her children) and has reported severe pain in the non-affected extremity because of overuse. She also reports widespread allodynia. These physical and psychological symptoms, combined with her inability to perform customary household activities, have strained her marriage significantly.

Patient 2 does not have to juggle a career and raising young children like Patient 1 does; nevertheless, she reports significant secondary conditions related to
Case Studies

Patient 1

Patient 1 is a 30-year-old female. She experienced a household accident on Christmas morning of 2016 and was seen at the emergency department of a local rural community hospital for a laceration of the extensor tendon in the left ring finger. She also experienced numbness and tingling in the lacerated finger. She met with a hand surgeon the following day. Upon physical examination, there was no active flexion at the flexor digitorum profundus (FDP) joint or ability to flex the tip of the finger. The surgeon suspected a ruptured FDP tendon and promptly recommended surgery. An exploration of the left ring finger was performed on December 30, and it was found that the FDP tendon and the ulnar digital nerve were significantly damaged. The FDP tendon and ulnar digital nerve were then repaired, and an Axogen nerve guard was implanted. This patient sustained two distinct, significant insults to the ulnar digital nerve: the initial laceration, followed by the surgical repair; this patient’s CRPS would therefore be categorized as CRPS-2.

Several months after the initial surgery, Patient 1 continued to report significant pain in the left ring finger. The injury had already affected her employment status, and she needed regular PT. She also began to develop tendinitis in the right wrist from overuse. After careful examination of the finger by a second surgeon, Patient 1 was diagnosed with an ulnar digital nerve neuroma. Surgery was performed in order to remove the neuroma. Importantly, following this second surgery Patient 1 began to develop pain in the left shoulder and neck area, as well as skin temperature changes. Her doctors believed that this pain was mechanical in nature, perhaps due to an intraoperative traction injury. For the next three years, the patient’s symptoms persisted and gradually worsened. Over time, the left finger pain traveled up the entire left arm with a cold sensation. A “sunburn” type sensation in the left arm, accompanied by hyperalgesia, was also noted. In the fall of 2019, Patient 1 was diagnosed with CRPS-2 in the left upper extremity.

CRPS has affected every aspect of this patient’s life. She is a nurse and worked 38–42-hour weeks prior to the injury to her left finger. Following the initial injury, she was out of work for roughly five months. Since that time, she has been able to work only part time, as the left-hand and arm pain from CRPS have become almost unbearable. This has affected her family’s finances. Her doctors have
recommended regular ketamine infusion therapy, but this treatment is not covered under her health insurance, and her limited work schedule has made it impossible for her to pay for the ketamine infusions out of pocket.

Patient 1 is married and has two children. Since the initial injury and onset of CRPS, the simple action of holding her children has become impossible. This has greatly affected her marriage and her ability to care for her children. These life-altering events have led to a diagnosis of depression (a common secondary condition of CRPS) in this patient.

Below is a list of Patient 1’s current CRPS symptoms.

- Shoulder and neck pain following surgery
- Shooting pain from left ring finger to left side of neck
- Hypersensitivity of left ring finger (spread to other areas of the body)
- Cold skin on left hand and arm
- Temperature changes
- Color changes in left hand and arm
- Edema
- Numbness in left hand
- “Sunburn” type sensation throughout entire left upper extremity
- Abnormal sensations
- Plastic appearance of left hand compared to right hand
- Sleep disturbances due to pain in left upper extremity
- Trophic changes on left hand
- Cyanotic appearance of left upper extremity

Patient 2

Patient 2 is a 64-year-old female. She suffered a fall in February of 2017 while walking on a sidewalk with uneven pavement. She landed on her outstretched right arm and felt immediate severe pain in her right upper arm. She was taken to a local emergency department via ambulance. X-rays revealed a fracture of the right proximal humerus involving the surgical neck and greater tubercle. She was discharged with a sling and was told to follow up with the orthopedic doctor of her
choice. She did so, the fracture healing routinely, and she did well on the prescribed pain medications; however, in May of 2017, she began to develop pain and numbness in her right hand and fingers. She also had extreme grip difficulties. Her treating physician suspected CRPS and ordered a TPBS. The TPBS showed increased blood flow with soft-tissue inflammation and osseous uptake in the periarticular regions of the right upper extremity consistent with CRPS. This confirmed the treating orthopedist’s clinical diagnosis of CRPS. A second opinion also confirmed the CRPS diagnosis.

Since the time of her diagnosis, Patient 2 has reported referred pain in her neck and in her left (uninjured) arm. She has been prescribed gabapentin, tramadol, and ibuprofen and has been participating in regular PT. Unfortunately, these therapies have not resolved her severe pain. Her doctors have recommended that she undergo a stellate ganglion block, but Patient 2 has a severe phobia of needles because of complications that her mother suffered during a medical procedure years ago. In the spring of 2020, Patient 2 had decided to go forward with the injection therapy despite her phobia, because the pain had become unbearable, but the procedure was canceled because of the COVID-19 pandemic, and at the time of this writing, she was in the process of rescheduling her injection.

Below is a list of Patient 2’s current symptoms.

- Right-hand pain, swelling, and numbness
- Right-hand hyperhidrosis
- Tingling in right hand and arm
- Sensitivity to cold weather
- Feeling of intense cold in right hand
- Skin-color changes in right hand
- Trophic changes in right hand
- Sleep disturbances due to pain in right upper extremity
- Allodynia in right hand and arm
- Decreased motor function in right hand and arm

Patient 2 presents an interesting case, in that she was diagnosed with CRPS early, within several months of the inciting trauma. Because of her phobia of needles and resulting apprehension toward the evidence-based stellate ganglion block therapy, however, she has been unable to avail herself of the most promising
therapies for her condition. It remains to be seen if the injections that will be done roughly three years after diagnosis will be effective.

Both case studies featured above demonstrate the real-life barriers that individuals experience when attempting to receive the best, most effective treatment for the management of CRPS.

Conclusion

CRPS is a fascinating yet challenging condition for medical professionals. Accurate, timely diagnosis and effective treatment can be elusive with even the most experienced and conscientious physicians. Patients suffering from CRPS live with persistent severe pain, and this can affect every aspect of their lives, including career, family life, mental health, and overall quality of life. Additional research on CRPS is therefore vitally important and will, hopefully, lead to improvements in the diagnosis, treatments, and outcomes for patients with the condition.
References


15. RSDSA. *CRPS and RSD Research and Medical*. rsds.org/researchmedical/


PAINTING A PRETTY PICTURE: THE ROLE OF SOCIAL DESIRABILITY IN THE MEMORY SELF-EFFICACY OF YOUNG AND OLDER ADULTS

KEEGAN G. SAWIN, BUTLER UNIVERSITY
MENTOR: TARA LINEWEAVER

Abstract

This study examined the relationships between social desirability, depression, memory self-efficacy, and objective memory in both young- and older adult populations. The study was designed to replicate the findings of Lineweaver and Brolsma (2014) and to determine whether these findings would generalize to individuals in later adulthood. Participants were 45 young adults (88% female, 80% White) and 47 older adults (42% female, 100% White) and completed measures of depression, objective memory, memory self-efficacy, and social desirability. As predicted, older adults were higher in levels of social desirability than were young adults, but the memory self-efficacy of young adults was more closely related to social desirability than was that of the older age group. Although social desirability did not mediate the relationship between depression and memory self-efficacy, significant support was found for the mediation of social desirability on the relationship between objective memory and memory self-efficacy in both young- and older adult populations. Together, these results indicate that social desirability exerts influence on the memory self-perceptions of both young- and older adult populations and that taking social desirability into account may improve the accuracy of memory self-reports in healthcare settings.

Within the field of memory, many factors may affect an individual’s memory performance. One such factor is subjective beliefs about memory, or memory self-efficacy (Cook & Marsiske, 2007; Pearman & Trujillo, 2013). Self-efficacy is broadly defined as the opinions individuals hold about their own capabilities and potential for performing and completing a specific task or goal (West & Berry, 1994). In the context of memory, individuals’ personal beliefs about their memory abilities may inform their actual memory performance (Cook & Marsiske, 2007; Pearman & Trujillo, 2013). In addition to acting as a possible predictor of performance, memory self-efficacy may also influence the likelihood that older adults seek help for memory problems. In a 2011 study, Hurt and
colleagues examined a population of older adults with documented subjective memory complaints. Results showed that despite similar levels of memory performance, older adults with lower memory self-efficacy were more likely to report memory deficits and to seek help than were those with higher self-efficacy (Hurt et al., 2011).

Interestingly, measures of memory self-efficacy often do not accurately reflect objective memory performance. For example, a study by Mendes et al. (2008) assessed a large group of adults of varying ages on both subjective memory complaints and objective memory performance. This study documented no correlation between the two. These findings suggest that although memory self-efficacy may inform memory performance, additional variables beyond underlying memory abilities may, in turn, influence memory self-efficacy. For example, memory self-efficacy itself is vulnerable to the effects of aging. That is, memory performance generally declines with age (Lineweaver & Hertzog, 1998; Wells & Esopenko, 2008), and memory errors resulting from this decline, combined with negative stereotypes surrounding aging, reinforce negative self-beliefs about memory, which leads to decreased memory self-efficacy (West & Berry, 1994).

Several studies to date have also examined the effects of depression on memory self-efficacy, generally concluding that depression levels correspond with more negative memory self-perceptions (Cipolli et al., 1996; Tillema et al., 2001). For example, a study from Cipolli and colleagues (1996) investigated the memory self-efficacy of depressed older adults. Overall, results showed that highly depressed individuals were more likely to rate themselves as poor performers on memory tasks, demonstrating lowered self-efficacy. The relationship between depression and subjective memory may be rooted in self-beliefs about one’s potential to perform well. A variety of factors affect the origins of depression but create the same outcome: highly depressed individuals are more likely to have unrealistic standards and thus consistently rate themselves as inept in their performance (Tillema et al., 2001).

Another factor that has the potential to influence memory self-efficacy is social-desirability bias, which is the tendency to deny or underreport socially unacceptable actions while highlighting socially attractive behaviors in self-reports (Latkin et al., 2017). More importantly, social desirability has the power to distort individuals’ self-descriptions to the extent that they present only what they believe to be acceptable in the wider social sphere, in an effort to maintain their own self-concept or others’ positive opinions of them. This bias is critical to understand in health-related matters because social desirability can render health histories and
patients’ symptom reports inaccurate (Burke & Carman, 2017; Latkin et al., 2017). For example, a descriptive study by Latkin and colleagues (2017) demonstrated this effect within a drug rehabilitation clinic setting. Drug users high in social desirability rated themselves as less-frequent users than they actually were, in addition to rating themselves higher on a measure of subjective health status. Taken together, these results reflect the problematic nature of inaccurate self-reports as a result of the influence of social desirability.

Less is known about how social desirability may influence memory self-efficacy, self-reported memory complaints, and the help-seeking behaviors of older adults. Past research has documented an upward trend of social-desirability levels with increased age (Soubelet & Salthouse, 2011), but only one study to date has measured the specific effects of social desirability on memory self-efficacy. In this study, a sample of young adults completed measures of social desirability, memory self-efficacy, and current affect before taking a short memory test aimed at assessing their actual memory abilities (Lineweaver & Brolsma, 2014). The study documented a significant correlation between negative affect and memory self-efficacy, consistent with past research (Cipolli et al., 1996; Tillema et al., 2001). When social desirability was controlled for, however, the relationship between negative affect and memory self-efficacy weakened considerably, whereas the relationship between memory self-efficacy and participants’ performance on the memory test was strengthened (Lineweaver & Brolsma, 2014). These findings suggest that social desirability may strongly contribute to the correlation between measures of self-reported negative affect and memory self-efficacy, and that taking social desirability into account in future research may increase the accuracy of memory self-reports from participants.

The current study was designed to build on the research of Lineweaver and Brolsma (2014) by expanding the scope to older as well as young adults. The current study examined the relationships between social desirability, depression, memory self-efficacy, and objective memory in both young and older adult populations to determine whether the previous findings can be replicated and whether they generalize to individuals in later adulthood. The first aim of the current study was to determine whether levels of social desirability differ with age. Consistent with prior research (Soubelet & Salthouse, 2011), we predicted age differences in social desirability, such that older adults would evidence more social desirability than their younger peers. The second goal of this study was to explore how social desirability affects the memory self-efficacy of both young and older adults. We hypothesized that although social desirability would be higher in the
older adult population, social desirability would have a larger effect on the memory self-efficacy of young adult participants. Although this may seem counterintuitive, we suspected that the more normative nature of experiencing memory problems with advanced age would allow older adults to more readily admit to memory difficulties than it would young adults, even if they are high in social desirability. The third goal of this study was to explore the potential mediating effects of social desirability on the relationships between depression and memory self-efficacy and between memory self-efficacy and objective memory. We expected to find results paralleling those of Lineweaver and Brolsma (2014) in this broader age sample. Specifically, we hypothesized that depression scores and memory self-efficacy would become less related and that memory self-efficacy and actual memory performance would become more related in both young and older adults when social desirability was controlled for.

**Method**

This study is retrospective and thus centered on data previously collected.

**Participants**

Participants included 92 individuals: 45 young adults ($M = 20.02$ years of age, $SD = 1.19$) and 47 older adults ($M = 76.72$ years of age, $SD = 9.24$). The young-adult group included Butler University undergraduate psychology students, who were recruited via the online Sona research participant management system. Older adults were recruited through senior centers (Hendricks County Senior Center and the Social of Greenwood) and senior living communities (Robin Run, Marquette Manor, and Cambridge Square) within the Indianapolis area. Demographic characteristics of both age groups are summarized in Table 1.
Table 1. Demographic Characteristics of Participants in Younger Adult and Older Adult Age Groups

<table>
<thead>
<tr>
<th></th>
<th>Young Adults</th>
<th>Older Adults</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 45)</td>
<td>(n = 47)</td>
</tr>
<tr>
<td>Age*</td>
<td>20.02 (1.19)</td>
<td>76.72 (9.24)</td>
</tr>
<tr>
<td>Years of Education*</td>
<td>13.71 (1.14)</td>
<td>15.00 (3.41)</td>
</tr>
<tr>
<td>Gender (% female)*</td>
<td>88.24%</td>
<td>42.66%</td>
</tr>
<tr>
<td>Race (% White)*</td>
<td>80.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Depressive Affect</td>
<td>0.75 (0.44)</td>
<td>0.57 (0.44)</td>
</tr>
<tr>
<td>Social Desirability*</td>
<td>0.52 (0.22)</td>
<td>0.72 (0.20)</td>
</tr>
<tr>
<td>Memory Self-Efficacy*</td>
<td>3.65 (0.48)</td>
<td>3.42 (0.50)</td>
</tr>
<tr>
<td>Total Memory Score*</td>
<td>19.56 (5.10)</td>
<td>9.38 (4.23)</td>
</tr>
</tbody>
</table>

*A statistically significant difference existed between the two age groups.

Unsurprisingly, these two groups differed significantly in age, $F(1, 90) = 1665.33, p < .01, \eta^2_p = .95$. Groups also differed in their gender and ethnicity distributions, with the older adult group having more gender diversity than the young adult group, $\chi^2(n = 92) = 11.52, p < .01$. In contrast, the adults in the young adult group were more ethnically diverse than were their older adult counterparts, $\chi^2(n = 92) = 10.42, p = .03$. The two age groups also differed in their educational achievement [$F(1, 90) = 5.81, p = .02, \eta^2_p = .06$], with older adults having completed more years of education than the younger adults, although this difference
is likely because many of the young adults were still attending school at the time of data collection and had not yet finished their degrees. Younger adults also significantly outperformed older adults in a task of memory recall \( F(1, 90) = 108.89, p < .01, \eta_p^2 = .55 \) and perceived their memories more positively on a measure of memory self-efficacy \( F(1, 90) = 5.32, p = .023, \eta_p^2 = .06 \). The two groups were statistically equivalent in depressive affect \( F(1, 90) = 3.77, p = .06, \eta_p^2 = .04 \), although there was a trend toward younger adults endorsing more depressive affect than their older peers did.

**Materials**

**Center for Epidemiological Studies Depression Scale**

The Center for Epidemiological Studies Depression Scale (CES-D; Radloff, 1977) assessed participants’ recent depressive affect. Participants indicated the number of times during the past week that they had felt or behaved according to the questionnaire’s statements. Example items included “I was bothered by things that don’t usually bother me” and “I had trouble keeping my mind on what I was doing.” This 20-item questionnaire used a 4-point Likert-type scale ranging from 0 (“rarely or less than 1 day”) to 3 (“all of the time or 5–7 days of the week”). Higher scores indicated higher levels of depression.

**Memory Self-Efficacy Questionnaire**

Participants’ memory self-efficacy was measured using a version of the Memory Assessment Clinics Self-Rating Scale (MAC-S; Crook & Larabee, 1990) adapted by Lineweaver and Brolsma (2014). This adapted questionnaire consisted of 27 items from the original MAC-S, all of which evaluated participants’ beliefs about their memory abilities. Statements such as “I am bad at remembering who I was with at major events months ago” and “I never miss the point someone else is trying to make during a conversation” were rated by participants on a Likert-type scale from 1 (“Strongly Agree”) to 5 (“Strongly Disagree”). When applicable, responses to items were recoded such that higher scores on this measure represented better memory self-efficacy.

**Memory Task**

Participants’ memory abilities were assessed via an objective memory test (Lineweaver & Hertzog, 1998). Each participant spent three minutes studying a list
of 40 unrelated words, with the goal of remembering as many as possible. They then had two minutes to write down as many words as they could recall. Participants made predictions and postdictions concerning the number of words they believed they would be able to remember or had remembered.

**Marlowe-Crowne Social Desirability-Short Form C**

A shortened version of the original Marlowe-Crowne Social Desirability Scale (Reynolds, 1982) included 13 items measuring participants’ tendencies to answer questions in a socially desirable manner. Examples of statements included “It is sometimes hard for me to go on with my work if I am not encouraged” and “I sometimes feel resentful if I don’t get my way.” Participants rated each statement as either true or false. Higher scores reflected greater social desirability.

**Demographic Questionnaire**

The demographic questionnaire gathered general descriptive information from participants, including age, gender, years of education, highest degree earned, and ethnicity.

**Procedure**

All participants gave informed consent before completing the packet of questionnaires in a fixed order. Participants were tested in small groups and were offered extra credit or payment as an incentive for participation.

**Results**

**Age Differences in Social Desirability**

To address our first hypothesis—specifically, whether older adults differed from young adults in levels of social desirability—we ran a one-way between-subjects ANOVA in IBM SPSS Statistics. In alignment with the hypothesis and previous research, the two age groups differed significantly in their level of social desirability, $F(1, 90) = 21.46, p < .01, \eta^2_p = .19$. Older adults ($M = .72, SD = .19$) scored more highly in social desirability than did young adults ($M = .52, SD = .22$).
Relationship Between Social Desirability and Memory Self-Efficacy

For our second hypothesis, we investigated the individual relationships between social desirability and memory self-efficacy for young adults and older adults in two separate correlational analyses. Consistent with the second hypothesis, social desirability was more strongly related to memory self-efficacy in young adults ($r = .427, p = .003$) than in older adults ($r = .364, p = .012$), although a Fisher r-to-z transformation indicated that the strength of the two correlations did not differ from each other significantly ($z = 0.35, p = .36$).

Mediating Effects of Social Desirability

Before examining the mediating effects of social desirability on the relationships among depression, actual memory abilities, and memory self-efficacy, we calculated the univariate correlations between the study’s four primary outcome variables. Overall, depression was significantly related to both social desirability and memory self-efficacy, signifying a possible opportunity for mediation. Social desirability also possessed significant relationships with the variables of objective memory and memory self-efficacy. The relationships between objective memory and depression and between objective memory and memory self-efficacy did not reach significance (Table 2).

<table>
<thead>
<tr>
<th></th>
<th>Social Desirability</th>
<th>Memory Self-Efficacy</th>
<th>Memory Ability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depression</td>
<td>−.330*</td>
<td>−.417*</td>
<td>.121</td>
</tr>
<tr>
<td>Social Desirability</td>
<td></td>
<td>.241*</td>
<td>−.213*</td>
</tr>
<tr>
<td>Memory Self-</td>
<td></td>
<td></td>
<td>.179</td>
</tr>
<tr>
<td>Efficacy</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

To accurately measure any potential mediation effects, a bootstrap analysis was run in SPSS using the PROCESS macro, version 3.5 (Hayes, 2017). Five
thousand bootstrap samples were run in the analysis of Models 1 and 2, respectively.

Model 1 examined the possible mediating effects of social desirability on the relationship between depression and memory self-efficacy (Figure 1). The overall model with both predictors was significant, \( R^2 = .186, F(2, 89) = 10.15, p = .0001 \). The relationship between depression and social desirability was highly significant \([t(90) = -3.32, p = .001]\), demonstrating a strong relationship between social desirability and depression. In the mediation model, the direct effect of depression on memory self-efficacy was significant (95% CI \([-0.65, -0.20]\)), but the indirect effect through social desirability did not reach significance (95% CI \([-0.14, 0.04]\)). Thus, social desirability did not mediate the relationship between depression and memory self-efficacy.

Figure 1. Proposed Mediation Model for the Influence of Social Desirability on the Relationship Between Depression and Memory Self-Efficacy

Model 2 was assessed for mediation, specifically for the mediating influence of social desirability on the relationship between objective memory and memory self-efficacy (Figure 2). The second overall model also reached significance \([R^2 = .113, F(2, 89) = 5.70, p = .0047]\), and a significant relationship was identified between objective memory and social desirability \([t(90) = -2.06, p = .0418]\). The bootstrapping analysis identified significant mediation. Both the
direct effect of objective memory on memory self-efficacy (95% CI [0.0027, 0.0321]) and the indirect effect through social desirability (95% CI [-0.0113, –0.0002]) reached statistical significance. Taken together, these results suggest that social desirability serves as a mediator between actual memory and memory self-efficacy, strengthening the relationship between the two when social desirability is controlled for.

Discussion

The current study aimed to answer three primary questions. The first goal was to determine if age-related differences in social desirability exist between young adults and older adults. We predicted that in comparison to the younger age group, older adults would be higher in their levels of social desirability. The study’s second objective was to determine how levels of social desirability would affect the memory self-efficacy of both young and older adults; we predicted that social desirability would exert a larger influence on the memory self-efficacy of young adults, despite higher levels of social desirability in the older adult age group. Finally, the study aimed to replicate the findings of Lineweaver and Brolsma (2014)
with young adults and to expand on that work by examining similar relationships in older adults. More specifically, we hypothesized that social desirability would serve as a mediator in the relationship between depression and memory self-efficacy as well as in the relationship between objective memory and memory self-efficacy.

Considerable support was found for the first hypothesis. As predicted, young and older adults differed significantly in their levels of social desirability, with older adults exhibiting more social desirability than their younger peers. While this study observed social desirability as it relates to memory, this result demonstrates the need for social desirability to be taken into account in a clinical setting, especially for older adult patients. Past research has illustrated the ways in which social desirability can distort a patient’s reports of symptoms (Burke & Carman, 2017; Latkin et al., 2017). If not controlled for, social desirability may affect the validity of older adults’ self-reported concerns about their health and cognition, therefore affecting the overall efficacy of care and treatment planning they receive from their providers.

Similarly, support was also found for the second hypothesis, which stated that young adults’ memory self-efficacy would be more affected by social desirability than would that of older adults. Because memory-performance declines are considered a part of the normal aging process (Lineweaver & Hertzog, 1998; Wells & Esopenko, 2008), we hypothesized that social desirability would actually present a more significant impact on the memory self-efficacy of young adults because impaired memory performance for this age group is more atypical. As predicted, our correlational analyses provided support for this hypothesis. This particular finding expands on past literature because no study to date has examined the interplay of social desirability, memory self-efficacy, and age.

In an effort to replicate the findings of Lineweaver and Brolsma (2014), we predicted that social desirability would function as a mediator in the relationship between depression and memory self-efficacy. A goal of the current study was to expand upon the age demographic studied by Lineweaver and Brolsma (2014), so we hypothesized that this relationship would be observed in both young and older adults. Surprisingly, support was not found for this hypothesis in either age group. While significant relationships were identified between depression and social desirability, the indirect effect of social desirability on the relationship between depression and memory self-efficacy did not reach significance, signaling no mediating effects from social desirability.
A possible explanation for the lack of observed significance may be that the current study did not use a mood-state questionnaire but rather a measure of depression to measure current affect as it related to memory self-efficacy. Lineweaver and Brolsma (2014) included both the CES-D (administered a month prior to testing) and a mood-state questionnaire (administered during the testing session) as potential predictors of memory self-efficacy. They found that social desirability mediated the relationship between current mood state and memory self-efficacy but not the relationship between depression and memory self-efficacy. In designing this study, we utilized the standardized CES-D rather than a less-formal mood-state questionnaire, but we administered the CES-D concurrently with the other test measures. We observed a direct effect of depression on memory self-efficacy but no indirect effect of depression on memory self-efficacy via the influence of social desirability. Although this result is similar to Lineweaver and Brolsma’s (2014) results, if this study were to be improved upon for the future, it would be beneficial to add a current-affect or mood-state questionnaire to determine the true possibility of any mediating effects of social desirability on memory self-efficacy.

Finally, we found support for the mediation of social desirability on the relationship between actual memory and memory self-efficacy, replicating the findings of Lineweaver and Brolsma (2014) across multiple age groups. This suggests that when social desirability is controlled for in the case of both young and older adults, the relationship between actual memory abilities and perceived memory abilities is stronger. This signals that social desirability plays a role in modulating the personal perceptions of one’s own memory in both young and older adults. Support for this relationship remains practically relevant in a healthcare-related sense, such that social desirability may cause inconsistencies in an individual’s self-reports to caretakers and primary care providers. In turn, such inaccurate self-reports may prevent the application of necessary treatment for memory difficulties.

Although we found statistical support for two of our original hypotheses as well as for the second proposed mediation model, several aspects of the study may limit the generalizability of its results. For example, data collection involved a relatively small sample size with little diversity among participants. Additionally, the young-adult sample was made up exclusively of college students, which could introduce confounding factors such as those associated with higher levels of education or expectancy effects. Levels of social desirability may vary between individuals with differing levels of education, but future research would be
necessary to address this question directly. Finally, the current study did not include middle-aged adult participants, so results cannot be generalized across all age groups.

In the case of future replications of this particular study, a measure of framing effects could be added to understand its influence on memory self-efficacy. In their 2014 study, Lineweaver and Brolsma also investigated item-framing, utilizing a memory self-efficacy questionnaire with positively, neutrally, and negatively worded items. As they predicted, the wording of the items was important and, when combined with mood state, influenced the memory self-perceptions of participants. Additionally, future research would benefit from the addition of a humility index. On the opposite end of the spectrum from social desirability, individuals who are high in levels of humility may be prone to rating themselves lower on their memory abilities, negatively influencing their perceptions of their memory abilities. By studying the possible effects of both item-framing and humility, a more holistic picture of social desirability as it relates to memory self-efficacy may form.

Despite the existing limitations, the results of this study present valuable insights on the study of memory self-efficacy and the variables that influence it. Results document that young adults tend to be lower in social desirability than do older adults but that social desirability has a larger influence on their memory self-efficacy. Although the true mediating effects of social desirability on the relationship between depression and memory self-efficacy are not fully known, we did find support for social desirability’s role in reducing the accuracy of memory self-perceptions. This study adds to the body of literature on this topic by examining these relationships in older as well as younger adults and by focusing on the long-term relationship between depression and memory self-efficacy. Further research between these variables and their interactions with one another may serve as a valuable predictor of memory self-reports, especially in the healthcare sphere.
References


