2021/2022

Excellence in Upper-Level *Writing*

Excellence in Upper-Level Writing

2021/2022

Table of Contents

Excellence in Upper-Level Writing

Excellence in Upper-Level Writing 2021/2022

Sweetland Writing Prize Chair

Gina Brandolino

Sweetland Writing Prize Judges

Stacey Bishop

Sheira Cohen

April Conway

Caitlin Dyche

Elliot Greiner

Andrew Hersche

Brenna Larson

Michael Lempert

Kate O'Connor

Chao Ren

Catherine Schenck

Administrative Support

Laura Schulyer

Aaron Valdez

Winners List

Excellence in Upper-Level Writing (Sciences)

Ella Olesen

Growing Plants with Pee: It Works
Nominated by Emilia Askari, ENVIRON 320

Josephine Schmidt

The Grandmother Hypothesis
Nominated by Stacy Rosenbaum, ANTHRBIO 368

Excellence in Upper-Level Writing (Social Sciences)

Ethan VanValkenburg

Maintaining Peace and Biodiversity in Colombia
Nominated by Omolade Adunbi, AAS 322/ENVIRON 335

Grey Weinstein

Thesis Proposal: Abolition, Mutual Aid, and the Policing of
Transgender Communities
Nominated by Nancy Burns (Faculty) and Erin Cikanek (GSI), POLSCI 381

Excellence in Upper-Level Writing (Humanities)

Gabriel Consiglio

The Taste of Words Once Forgotten: Multi-Hyphenate Mediterraneanism and the
Universal Language of the Kitchen
Nominated by Jennifer Metsker, ARTDES 399

Anastasia Warner

Progress in Explainable AI for Deep Neural Nets
Nominated by Jimmy Brancho, WRITING 400

Upper-Level Writing Prize Nominees

Hannah Allbery

Niles Boothe

Hayyan Chaudry

Gabriel Consiglio

Renu Dabak-Wakankar,

Michael (Malamas) Daratzis

Sasha Fraser

Jonas Gomes

Stefania Gonzalez

Alexandra Jay

London Jones

Daniela Kabeth

Arya Kale

Namitha Kumar

Estelle Leibowitz

Hannah Lemos

Dominic Manzo

Brian McNair

Max Mittleman; Rachel Ellis

Connor Moore

Ella Olesen

Nina Elise Pacheco

Claire Pajka

Hunter Petz

Brianna Regan

Josephine Schmidt

Grace Shulman

Max Stoneman

Ethan VanValkenburg

Irene Wang

Anastasia Warner

Caitlyn Webster

Grey Weinstein

Kristie Wilcox

Matthew Yacoub

Vinay Yarlagadda; DJ Hong

Alp Yel

Hailey Zuverink

Upper-Level Writing Prize Nominating Instructors

Omolade Adunbi

Emilia Askari

Jimmy Brancho

Nancy Burns; Erin Cikanek (GSI)

Lynn Carpenter

Angela Coco

Rebecca Conway

Haley Dalian

Emmalon Davis

Jaimien Delp

Alora Fleenary

David Gold

Amanda Gregolynskyj

Julie Halpert

Ben Hansen

Ingrid Hendy

Katherine Hunter

Ashley McDermott

Jennifer Metsker

Katie Nissen

Anne Pitcher

Colleen Roberts

Stacy Rosenbaum

Fritz Swanson

Sophia Tonnessen

Basit Zafar

Introduction

The Gayle Morris Sweetland Center for Writing bestows six Upper-Level Writing Prizes annually in recognition of outstanding student work across the curriculum. Two prizes are awarded within each division of the College of LSA: humanities, social science, and science. All LSA undergraduates take at least one upper-level writing-intensive course. Those of us who teach these courses discover anew each term how a focus on writing improves students' thinking, learning, and rhetorical creativity. At the end of each term, we applaud our students' various achievements and the ways they have grown by applying themselves seriously to the hard work of designing a study, conducting research, addressing an audience effectively, crafting a multimedia essay, refining their prose style, reading and giving feedback on peers' drafts, or engaging with counter-arguments. These are rewarding courses to teach, as witnessed by the significant number of faculty from all disciplines who participate robustly in the upper-level writing program each year.

Each year, faculty and graduate student instructors encourage undergraduates to submit their very best work for the Upper-Level Writing Prizes. Fellows in the interdisciplinary Sweetland Seminar for Writing Pedagogy read the submissions and rank them according to their overall excellence. This is an intellectually interesting exercise and generates considerable discussion about what we value when we read students' work. The Fellows this year described for each other the qualities they admired in the submissions: the relationship between existing scholarship and a student's position is clearly detailed and persuasive; the argument is complex without loss of clarity or purpose; the piece presents a novel argument or assertion (e.g., coming to their own conclusions based on multiple texts); quotes are thoughtfully introduced and integrated into the argument; the evidence is compelling; and the writers use precise language, appropriate to their rhetorical purposes. All of the essays submitted are outstanding, and all of the students nominated for a prize should feel very proud of their achievements.

This volume showcases the prize-winning essays, which are truly impressive. They witness to the robust intellectual life of the university, and to the splendid courses and instructors who inspired and supported the writers' accomplishments. Each essay speaks to how much our students contribute to the creation of new knowledge. The prize-winning essays exemplify the best of their genres: a review of humanities scholarship, a narrative argument, a family history, a thesis proposal, a social science argument, a review of science scholarship, and a journalistic research essay. To read these essays is to encounter the students' rigorous thinking, careful organization of ideas, ability to integrate substantial research into their work, and resourcefulness in identifying scholarly and creative topics to which they could make substantial new contributions.

Thanks are due to the many people who made this volume possible. The Sweetland Fellows who thoughtfully judged the essays are April Conway, Sweetland Center for Writing; Andrew Herscher, Architecture and Urban Planning, History of Art, Slavic Languages and Literatures; Holly Hughes, Art and Design, Theatre and Drama, Women's and Gender Studies; Michael Lempert, Anthropology; Stacey Bishop, History; Sheira Cohen, Interdepartmental Program in Classical Art and Archaeology; Caitlin Dyche, Communication and Media; Elliot Greiner, Anthropology; Brenna Larson, History of Art; Kate O'Connor, American Culture; Chao Ren, History; and Catherine Schenck, Interdisciplinary Program Ancient History. Much gratitude is due to Aaron Valdez, who designed this volume; Laura Schuyler, who coordinated the submission and judging process; and Gina Brandolino, who chaired the Sweetland Prize Committee and edited the volume. Finally, thank you to the students and instructors who strive for—and achieve— excellence in writing and writing pedagogy.

Theresa Tinkle
Sweetland Center for Writing
Arthur F. Thurnau Professor and Professor of English

Excellence in Upper-Level Writing (Sciences)

Growing Plants with Pee: It Works
by Ella Olesen
From ENVIRON 320
Nominated by Emilia Askari

The moment that Ella Olesen suggested writing her news feature story about growing plants with pee, my co-teacher, Julie Halpert, and I were intrigued. We have taught environmental journalism for more than two decades. It is rare that students propose completely novel story ideas, develing into topics that we have never read about before. Ella did just that. Throughout the winter 2021 term, our interest grew as Ella prepared an annotated news roundup on the subject of human pee as a fertilizer, finding very little general-interest coverage of the topic; then found and interviewed leading experts in the practice, including some faculty at our own university; and also interviewed a gardener with a large Youtube following who promotes pee-cycling. As we learned from Ella, that's the common name for the practice of sprinkling pee on plants to help them grow. We realized that Ella had the ingredients of a successful news story that might appeal to millions of potential readers who embraced home gardening during the pandemic. When Ella began to write sections of her story, we were even more impressed. Ella's writing about this potentially icky topic is at once lively, respectful, and full of interesting scientific detail. Her first paragraph – known in journalistic writing as the lede – beautifully draws a scene for the reader, managing the usually private act of peeing in a way that encourages the reader to pay attention rather than look away. Throughout the rest of the story, Ella smartly writes in a clear,

easily understood way about complex scientific issues including the contributions of fertilizer production to climate change; challenges at sewage treatment plants; and the decomposition of medical compounds in soil. Although Ella is an avid gardener herself, she avoids over-enthusiasm for her subject, writing about potential downsides of pee-cycling. Ella concludes her story by returning to the opening scene in her story, with an engaging quote. It's a nice rhetorical device that wraps up the story for the reader. I am honored to nominate Ella's article for a Sweetland Upper Level Writing Award.

-- Emilia Askari

Growing Plants with Pee: It Works

Sara Bäckmo with her Swedish-designed watering-can-potty.

With rosy cheeks and radishes on her mind, Sweden's top garden blogger, Sara Bäckmo, 42, perches over her greenhouse watering-can-potty and pees. She adds water, then pours the pale yellow cocktail near the base of a hungry vine.

Though unconventional, others are following her lead. The pandemic turned 16 million bored, quarantined Americans into backyard gardeners in 2020, according to the Garden Media Group's latest Garden Trends Report, and some are turning to urine to nourish their gardens. The Rich Earth Institute, a Vermont-based nonprofit that researches urine-derived fertilizers, saw a tenfold increase in their annual "pee-cycling" conference audience — 200 people swarmed the virtual event.

Urine contains the same nutrients found in synthetic fertilizers: nitrogen, phosphorus, and potassium. It's a nutritional goldmine for plants.

But synthetic fertilizers continue to dominate the fertilizer market. Making them requires fossil fuels and a lot of energy, contributing 1% to global carbon dioxide emissions, according to a 2020 UK study. While certain barriers prevent global-scale urine fertilizing, Rich Earth's team believes in the liquid as a great alternative at the local level. It's organic, sustainable, and readily available — everyone pees.

Aside from feeding plants, fertilizing with urine also helps reduce nutrient pollution. Urine makes up most of wastewater's nutrients. Currently, wastewater treatment plants don't remove all these nutrients before discharging effluent into water bodies. Too many nutrients in a water body lead to harmful algal blooms. Gardeners repurposing rather than flushing urine are saving fish and aquatic life.

Bäckmo was fertilizing with urine two decades before the COVID-19 pandemic. She's an experienced gardener, with almost 80,000 subscribers between her two gardening YouTube accounts. Besides the environmental benefits, there is another reason she prefers it.

"It's absolutely free," she said. "You get pretty much the same thing as you buy in a bottle [of fertilizer] at the garden store."

Bäckmo's method produces enough vegetables to feed her six-person family year-round, despite Sweden's long, dark winters.

In 2013, one of Rich Earth's earliest studies confirmed what Bäckmo sees in her home garden: fertilizing a crop with urine resulted in the same yield as a crop grown with synthetic fertilizers.

"It works amazingly well," said Bradley Kennedy, a Research Associate at Rich Earth.

Kennedy receives lots of questions about what happens to pharmaceuticals in urine. Rich Earth is wrapping up a six-year research project in collaboration with the University of Michigan addressing this concern from many angles. The results haven't been published yet, but Kennedy provided a preview:

"When you add these [pharmaceuticals] — they're just like organic molecules — to the soil, they mostly get broken down by all the microbes in the soil. There's a few that get taken up in the plant but they're in extremely tiny, almost unfathomably tiny, amounts," she said. She pointed out that the current wastewater system doesn't treat or filter pharmaceuticals before discharge. Pulling the pharmaceuticals out via urine separation "gives you the potential to manage them in a healthier way."

Rich Earth researchers studying the fate of pharmaceuticals in soil and a lettuce crop.

Weighing urine risks against benefits was a major focus of Dr. Heather Goetsch, Research Engineer at National Renewable Energy Laboratory and former Rich Earth researcher. Her research found that potential pathogens in urine are generally inactivated after storage, making the liquid safe to use. But she remains careful about recommending urine use to all.

"There's no such thing as zero risk. It's a calculation of how much risk is acceptable, either to you personally or to the collective society," Goetsch said. "The risks [of fertilizing with urine] are, in my opinion, a lot lower than the benefits, as long as it's stored."

In 2006, the World Health Organization recommended that large systems — meaning gardens growing to feed more than one household — should store urine in a sealed container for six months before use as a fertilizer so it becomes sanitary. However, the WHO does not recommend any storage or treatment for single households eating their own crops. Fresh urine happens to be relatively sterile (unlike human feces).

Bäckmo tried storing her urine, but ran into a problem when she poured it on the lawn.

"It smells disgusting, absolutely disgusting." She closed herself in her house for two days, afraid to face her neighbors, and changed her game-plan. She fertilizes with fresh urine now, preferring to dilute it with water, and finds no smell.

The 'ick' factor is an important part of Rich Earth's research. Researchers are considering how to teach consumers about urine fertilizers while mitigating potential disgust. Julia Cavicchi, Rich Earth's Education Director, frames fertilizing with urine as an ancient practice, one that other cultures around the world have done throughout history and continue to do today.

"It's a pretty powerful way of reconnecting yourself to the local ecosystem, in that you're doing this act with your own bodily nutrients in your own backyard that has this powerful, broad, ecological impact," Cavicchi said. "The strange and bizarre thing is mixing and flushing our bodily nutrients with water. *That's* the weird thing."

Spinach grown by Rich Earth staff in 2019. The leftward crop was fertilized with urine; the rightward crop received no fertilizers.

But it's not as simple as stopping peeing in toilets altogether. Most big wastewater treatment plants need some of urine's nutrients to treat and break down waste into clean, dischargeable water.

"If we were to take all the nitrogen out, we would have to change how we treat wastewater," said Earl Kenzie, Ann Arbor's Wastewater Treatment Manager. U-M researchers are relying on his facility's data to explore how much urine can be removed before treatment plants would require renovations. But upgrades are not easy.

"I have to be the pragmatist," Kenzie said. "We don't have an endless supply of money." Updating a wastewater treatment plant can cost hundreds of millions of dollars.

Despite the challenges, nutrient recovery and sustainability are the focus of the wastewater treatment industry as well. It's a matter of deciding which upgrades get funding first.

Michael Nemeth, Senior Advisor of Agricultural and Environmental Sustainability at Nutrien (one of the world's largest fertilizer companies), has looked into harnessing urine as a nutrient resource at the business scale at Nutrien.

"It's technically possible," Nemeth said, "but until things are proven at scale, it's really hard to invest in them." Logistics, existing infrastructure, and costs stand in the way of large-scale urine recycling. In North America, urine disposal would need to change (think urine-separating toilets). Then treatment, transportation, and application methods would need to be optimized up to scale. "In North America, we're not even close."

However, Rich Earth is working on scale. The Institute's spin-off LLC program is developing a building-scale system capable of producing ready-to-use fertilizer inside. The Institute has also just received funding to launch its first community-scale program. "It will be our first time roping in a whole new network of urine donors and teaching people about this in a new location," said Cavicchi. "That's pretty exciting."

At the smallest scale, Cavicchi encourages home gardeners to "start now!" She threw her hands in the air for emphasis. The Rich Earth website includes a written guide with tips for newbies. Bäckmo's blog features advice too, and Bäckmo continues championing fertilizing with urine on her social media pages.

Though Bäckmo grows all year, she is especially looking forward to summertime gardening with her family. "Before we go to bed, we go through the garden down to the greenhouse — like a duck's family — and everyone pees on the potty. We might pick some berries, and just have a chat in the sunset. It's really, really nice."

Bäckmo in her garden.

Excellence in Upper-Level Writing (Sciences)

The Grandmother Hypothesis
by Josephine Schmidt
From Anthrbio 368
Nominated by Stacey Rosenbaum

Josephine Schmidt's essay on the Grandmother Hypothesis is an outstanding introduction to a difficult and intriguing puzzle in the study of human evolution: why do women stop reproducing many years before their death? Evolutionary theory predicts that ceasing reproduction early should be bad for evolutionary fitness, and therefore selected against. It is therefore unsurprising that in most animals, reproduction does not cease until very shortly before death. In humans, however, women routinely live 30+ years after the onset of menopause. This essay provides a concise, clear, and engaging explanation of competing (and in some cases, complementary) scientific hypotheses for this phenomenon.

-- *Stacy Rosenbaum*

The Grandmother Hypothesis

Overview:

Most primates continue reproducing until they die, which is advantageous because the longer an organism reproduces, the more offspring they will have and the larger their genetic contribution to subsequent generations. Humans, on the other hand, exhibit the unique life history of reproductive senescence in the absence of somatic senescence, also known as menopause. Decades go by where human females are alive but do not produce offspring. For this to be selected for, the direct fitness costs of stopping reproduction early must be outweighed by some indirect fitness benefit. Many theories have been proposed to explain the existence of this long post-reproductive lifespan and what the indirect fitness benefit could be. In this essay, I will describe one of the most popular theories, the Grandmother Hypothesis, as well as some of the alternatives to show different perspectives on the difficult-to-answer question of why menopause evolved in humans.

Background Theory:

To answer why reproductive senescence occurs prior to somatic senescence in humans, it is important to understand why organisms senesce in the first place. Natural selection should favor increasing life spans as this would allow organisms to reproduce longer and have higher fitness. It is obvious that organisms do not live forever, so there are clearly constraints as to how long a lifespan can be. Life history theory explains that organisms have a limited amount of energy and must allocate it to growth, reproduction, and/or maintenance. Energy is not infinite and is allocated in a way that enables an organism to have the highest fitness possible. Therefore, as energy is allocated to growth and reproduction, it is taken away from maintenance which eventually leads to senescence.

Pleiotropy is also thought to play an important role in the existence of senescence (Williams, 2001). Because a single gene can affect multiple expressed traits, a gene that produces a favorable trait early in life may produce an unfavorable trait later in life. Favorable traits at younger ages are prioritized over favorable traits

at older ages because the strength of natural selection decreases with age (Williams, 2001). The likelihood of making it to age 70 is less than the likelihood of making it to age 60, and the likelihood of reproducing decreases as the age increases because organisms must be alive to reproduce. Furthermore, the proportion of fitness that is affected by a favorable trait is greater at a young age because many more offspring would be impacted by this trait. Therefore, selection favors positive traits at younger ages over negative traits at older ages (Williams, 2001).

The combination of limited energy and pleiotropy explain why senescence occurs, but it does not explain why, in humans, reproductive senescence occurs much before somatic senescence. When looking at energy allocated specifically to reproduction, there is a tradeoff between putting energy into the production of new offspring versus putting energy into already existing offspring with the latter being more favorable at an older age where the chance of living longer is low. However, as mentioned previously, other primates continue to reproduce late into their lifespans despite also having long periods of childhood dependency like humans. Humans also share the life history trait of menopause with whales, suggesting that this is an evolved trait and is not just a byproduct of extended lifespans in humans (Johnstone & Cant, 2010). Furthermore, modern hunter-gatherer groups, who do not have access to modern medicine, also exhibit menopause and extended post-reproductive lifespans. Those who reach 45 years old are expected to live an additional 20 years (Blurton Jones et al., 2002). Lastly, humans and other great apes also have reproductive periods that are similar in length. Therefore, Hawkes et. al (1998) argues that the derived trait in humans is a long post-reproductive lifespan, not stopping reproduction early. They proposed the Grandmother Hypothesis which attributes the evolution of long-post reproductive life spans to mother-child food-sharing.

A lengthened post-reproductive lifespan could be selected for if grandmothers provide resources such as food to their children and grandchildren and improve their fitness. The genes that lengthened the grandmother's life would already be passed onto her children and grandchildren, and because they have higher fitness due to food-sharing, they would have a larger genetic contribution to subsequent generations,

perpetuating the advantageous trait of a long post-reproductive lifespan (Hawkes et al., 1998).

Charnov's Model, which quantifies life history trends, can be used to test the Grandmother Hypothesis. This model is based on the idea that production is allocated to either growing oneself or growing one's offspring. This is split into two time periods: conception to independence and independence to maturity (α). The adult mortality rate (M) is inverse of the average adult lifespan. As lifespan increases (α), the adult mortality rate (M) decreases so α and M vary inversely and determines the ideal time at maturity by comparing the probability of making it to a certain age to the benefits of growing longer. Furthermore, the size and age of weaning relates to adult body size. The size at weaning divided by adult body size (δ) is approximately constant, and because growth allometry increases faster than production, the size of weanlings increases faster with maternal size than the energy mothers can put into them. Therefore, the number of daughters produced each year (b) decreases as the age of maturity (α) increases (Hawkes et al., 2000).

Charnov's Model in Action:

If the Grandmother Hypothesis is true, three life history characteristics should be present in humans: 1) age at maturity should be adjusted to entire adult lifespan, 2) the size and age of weaning should be significantly lower than expected for the mother's size, and 3) birth rates in humans should be significantly higher than those of other primates (Alvarez, 2000).

Humans reach age at maturity much later than other primates, but αM is like that of other primates. Hawkes et. al (1998) found the αM in humans to be 0.44 and in orangutans, gorillas, and chimpanzees to be 0.46, 0.45, and 0.46, respectively. Furthermore, Alvarez (2000) who extended the study to 14 different primates including humans found the αM mean for all the primates to be 0.50 (0.44-0.58).

Next, δ is expected to be low in humans if the Grandmother Hypothesis is true. Both Hawkes et al. (1998) and Alvarez (2000) found the actual weaning weight in humans to be lower than the predicted weight. However, Alvarez (2000) found the weaning weight of Ache infants, a human hunter-gatherer population used in the

study, to be within the 95% confidence interval of the regression equation despite being lower than predicted. Both studies also found that gorillas wean their infants earlier than expected.

Although gorillas and humans both share lower than expected weaning weights, only humans have higher than expected birth rates for our age at maturity (αb). Hawkes et. al (1998) found αb in humans to be 2.05 and in gorillas, chimpanzees, and orangutans to be 0.79, 0.52, 0.70 respectively. Alvarez (2000), found the actual birth rates in 4 different human populations (Ache, !Kung, Turkana, and Amele) to be 2 to 5 times higher than expected based on primate invariance.

When compared to great apes and many other primate species, humans show the expected trends if the Grandmother Hypothesis were true. Age at maturity, size and time at weaning, and birth rate are all affected. The similarity in αM across primates indicates that humans, despite stopping reproduction earlier in our lifespans, are contributing to our own fitness for around the same portion of our lifespan as other primates. Consequently, α is adjusted to the whole lifespan because humans still contribute to their fitness after menopause through grand-mothering rather than directly producing offspring. The size and time at weaning should be lower than expected because food-sharing provides weaned infants with food that they could not get on their own allowing mothers to wean earlier. With less time spent nursing due to the contribution of grandmothers and less energy spent feeding their own children, women can allocate their energy toward reproduction sooner, shortening the interbirth interval and increasing the birth rate.

The Benefits of Close Geographic Proximity:

Geographic distance can be used to supplement the evidence found using Charnov's Model and further support the Grandmother Hypothesis. Data was collected from thousands of French settlers living in Canada between 1608 to 1799 to analyze how geographic distance of grandmothers correlated with grandchildren's fitness (Engelhardt et al., 2019). To control for genetic factors that can play a role in grandchildren's fitness, within family analyses were first conducted. The first within family analysis was done on families where the grandmother passed away after the

first sister reproduced but before the second sister reproduced. Environmental and age-related factors were also taken into account. When grandmothers were alive for first reproduction, daughters had on average 2.08 more offspring born and 1.14 more offspring that made it to 15 years of age than daughters who reproduced after their mother (grandmother) died (Engelhardt et al., 2019). The second within family analysis was done on families where the grandmother was alive when both of her children reproduced, and the geographic distance between the grandmother and her children were analyzed. The daughter living farther away from her mother (grandmother) had 1.75 fewer offspring born and 1.45 fewer offspring that made it to 15 years of age than daughters living closer (Engelhardt et al., 2019). They were also 1.37 years older when they first reproduced than their sisters who lived closer (Engelhardt et al., 2019). Therefore, grandmothers being alive and close by are correlated with more offspring being born, more offspring making it to age 15, and younger age of first reproduction. The same results are seen when analyzed overall and not just within families (Engelhardt et al., 2019).

The Grandmother Hypothesis is supported by these results because grandmothers seem to influence their children's fecundity (b) and their grandchildren's fitness which was measured as the grandchildren reaching 15 years of age. This fits with the points discussed previously that birth rates in humans should be significantly higher than other primates. It should be noted that there is a difference in the strength of the effect that grandmothers have on the offspring born versus offspring making it to 15 years of age (2.08 versus 1.14) and could be accounted for by the contribution grandmothers make at various stages in the grandchildren's life. Grandmothers could contribute more at earlier stages in their grandchildren's life due to the risks imposed by high childhood mortality rates and better personal health. Regardless, there is still a positive correlation between grand-mothering and children's and grandchildren's fitness. Age at first reproduction was also affected by geographic distance of grandmothers. This was not an explicit prediction of the Grandmother Hypothesis, but the age of first reproduction relates to fecundity. It can be assumed that earlier first reproduction leads to a longer reproductive period and more children. Therefore, age

of first reproduction also fits in with the prediction that grand-mothering should lead to higher birth rates.

Reproductive Conflict Hypothesis:

The Reproductive Conflict Hypothesis supplements the Grandmother Hypothesis. It explains how fitness costs are related to the evolution of menopause and suggests that a post-reproductive lifespan prevents competition between mothers and daughters by avoiding overlapping reproductive periods (Cant & Johnstone, 2008). The lack of overlap is not seen in other primates indicating that this life history trait may be a derived feature in humans. Follicular attrition accelerates in humans around age 38 leading to the cessation of menstruation around age 50 (Cant & Johnstone, 2008). As a result, menopause occurs 20 years earlier than expected if this acceleration did not occur (Cant & Johnstone, 2008). The acceleration also occurs around the same age that mother's daughters would begin to reproduce themselves. Furthermore, evidence suggests that the early dispersal pattern for *Homo* was female-biased meaning females leave their natal group at maturity and join another one (Cant & Johnstone, 2008). Consequently, females entering a new group have no relatedness to the group and rearing new offspring would always be more beneficial than helping older females raise their own. Older females already living in the group, on the other hand, are at less of a disadvantage when choosing to help raise their own grandchildren instead of rearing new offspring as long as there is some probability that her son is the father of the grandchildren. Furthermore, continued reproduction of older females can lead to fitness costs by competing with their own children's reproduction.

Cant and Johnstone (2007) argue that the Reproductive Conflict Hypothesis can be used to supplement the Grandmother Hypothesis by showing that the relatively lower fitness cost of grand-mothering in patrilineal societies in addition to the fitness benefits of grand-mothering in matrilineal societies can help explain the evolution of menopause. To add on to this idea, the combination of fitness costs and fitness benefits is important to look at because natural selection works both ways. It can negatively select for traits that are disadvantageous and positively select for traits that are advantageous. The Reproductive Conflict Hypothesis claims that continued

reproduction would be disadvantageous because younger females have nothing to lose with producing offspring in a new group while older females face costs that the younger females do not. The Grandmother Hypothesis claims that menopause is advantageous because it allows grandmothers to increase their fitness without having children themselves by contributing resources to their children and grandchildren. Therefore, both hypotheses can be used to explain why menopause may have evolved.

Local Resource Competition Hypothesis:

When looking at modern-day patrilineal groups, grandparents living with children are less positively correlated with their grandchildren's survival than grandparents who do not live with their grandchildren (Strassmann & Garrard, 2011). This contradicts the Engelhardt et al. (2019) findings that grandmothers living in close proximity to grandchildren led to higher fecundity and lower age of first reproduction as well as the idea that menopause in patrilineal societies would be selected for. Strassman and Garrard (2011) proposed the Local Resource Competition Hypothesis as an alternative to explain the contradiction.

The Grandmother Hypothesis, along with the Confidence of Paternity Certainty Hypothesis, predict that maternal grandmothers should invest the most in grandchildren followed by fraternal grandmothers and with maternal and fraternal grandfathers investing the least (MM > FM > MF & FF) (Strassmann & Garrard, 2011). Grandfathers are expected to invest the least because they are assumed to be more interested in mating opportunities. Paternity uncertainty leads fraternal grandmothers to be less inclined to invest in offspring because they cannot be certain that they are investing in a child that contains their own genetic material while maternal grandmothers can be certain. Maternal grandmothers are also associated with being geographically closer to their offspring, but this assumption is not true in patrilineal populations.

Strassman and Garrard (2011) conducted a meta-analysis to analyze the relationships between fraternal and maternal grandparents and grandchild survival in patrilineal populations. Fraternal grandparents (FM & FF) were not associated with any change in grandchild mortality risk, but the maternal grandparents (MM & MF)

were associated with a decrease in grandchild mortality risk. The probability that each grandparent had a positive association with grandchild survival was also assessed with FM = 0.58, FF = 0.53, MM = 0.83, and MF =0.74 further evidencing that maternal grandparents were more positively associated with grandchild survival and, therefore, with decreased mortality. The results show that MM & MF > FM & FF as opposed to the predicted MM > FM > MF & FF.

The results from the MM fit the predictions of the Grandmother Hypothesis, but the results for the MF and the FM are unexpected. Maternal grandparents favor 'investment' over fraternal grandparents even in patrilineal populations where fraternal grandparents live with the grandchildren. The Local Resource Competition Hypothesis claims that grandparents and grandchildren compete for the same resources, so having higher fecundity with more children to compete with in a population is not beneficial. The competition increases as grandparents age and become less self-sufficient. This is further evidenced by the Dogon where the presence of the FM is associated with a 200% increase in childhood mortality risk and a 54% increase with the presence of the FF (Strassmann, 2011). The FM must also leave the group if the FF dies and return to her natal patrilineage which suggests that the Dogon recognize the liability grandparents are to the family (Strassmann, 2011). Therefore, the Local Resource Competition Hypothesis suggests that menopause could be selected for to avoid additional kin competition – as explained by the Reproductive Conflict Hypothesis -- and consequently more resource competition.

I think that both the Grandmother Hypothesis and the Local Resource Competition Hypothesis can be combined to explain the trends seen in the Strassman and Garrard (2011) as well as the Engelhardt et al. (2019) studies. It appears that grandparents can either contribute resources to their grandchildren or compete with their grandchildren for resources. Maternal grandparents contribute resources in both matrilineal and patrilineal groups. In matrilineal groups, maternal grandparents can directly provide resources for their grandchildren increasing both their children's and their grandchildren's fitness (Engelhardt et al., 2019). In patrilineal groups, maternal grandparents are not directly providing resources, but they are indirectly

giving the grandchildren more resources by not competing with them. Only the Strassman and Garrard (2011) meta-analysis looked at the correlation of fraternal grandparents and grandchild survival, and they only looked at patrilineal groups. Therefore, the correlation of fraternal grandparents and grandchild survival in matrilineal groups cannot be determined, but in patrilineal groups they have either no correlation or a negative correlation with grandchild survival. This suggests that fraternal grandparents are not contributing to their grandchildren and/or competing with their grandchildren for resources. One explanation for this is that contribution from fraternal grandparents is not selected for due to paternity uncertainty where the chance of contributing to offspring that share the same genetic material is lower than with maternal grandchildren. Selection for contribution would be stronger on the maternal side because there is certainty around the genetic relationship between grandchildren and grandparents. Therefore, it would be advantageous to contribute to offspring from daughters and less advantageous to contribute to offspring from sons combining the Grandmother Hypothesis and the Local Resource Competition Hypothesis to explain the trends seen in the Strassman and Garrard (2011) as well as the Engelhardt et al. (2019).

Conclusion:

Grandparents in patrilineal versus matrilineal societies appear to have different effects on their grandchildren. Grandfathers tend to have either neutral or negative effects on their grandchildren except for maternal grandfathers in patrilineal populations where they have a positive correlation with grandchild survival. In this case, the maternal grandfathers are not directly contributing to their grandchildren's survival but are avoiding competing with them which positively affects their grandchildren's survival. Grandmothers are especially beneficial to their grandchildren's survival in matrilineal populations and have varying effects in patrilineal populations.

The first factor that I believe plays a major role in the selection for these varying levels of contribution and competition is paternity uncertainty. The benefits of contributing to grandchildren will outweigh the costs of stopping reproduction early when the genetic relatedness of grandchildren is certain. In matrilineal societies,

where the genetic relationship is certain, grandmothers seem to be more positively correlated with grandchild survival than grandmothers in patrilineal societies.

The second factor is differing sexual selection pressures between men and women. Men do not experience reproductive senescence much before somatic senescence like women do which may be a result of women prioritizing resources and men prioritizing finding mates in response to sexual selection pressures. Older women are less likely to successfully compete against younger women when competing for mates, and because it is more favorable for older women to invest in their grandchildren whom they are related to than for a younger female to invest in an older female's offspring whom she is not related to, stopping reproduction to invest in their grandchildren may be more beneficial for older women. Women must make trade-offs between raising new offspring or investing more in their current offspring due to limited resources and energy. Therefore, menopause may allow for the highest fitness where females can still contribute to their fitness throughout their entire lifespan even if they are not directly producing offspring. Men do not face the same trade-offs, so continued reproduction late into life is not as costly as it is to women. I propose that looking at the evolution of menopause as a consequence of sexual selection may be an interesting direction to go in.

Lastly, I would like to acknowledge that correlation does not mean causation. There is a lot of evidence that shows positive, negative, or neutral correlations between grandparents and grandchild survival, but none of these correlations can prove that the evolution of menopause was caused by any of these factors. Grand-mothering appears to be beneficial to grandchild survival in matrilineal populations, but this does not mean that menopause evolved to allow for grand-mothering. All these correlations could be aftereffects of a different cause which makes the question of "Why did menopause evolve?" so difficult to answer.

Bibliography

Alvarez, H. P. (2000). Grandmother hypothesis and primate life histories. *American Journal of Physical Anthropology, 113*(3), 435–450. https://doi.org/10.1002/1096-8644(200011)113:3<435::AID-AJPA11>3.0.CO;2-O

Blurton Jones, N. G., Hawkes, K., & O'Connell, J. F. (2002). Antiquity of postreproductive life: Are there modern impacts on hunter-gatherer postreproductive life spans? *American Journal of Human Biology, 14*(2), 184–205. https://doi.org/10.1002/ajhb.10038

Cant, M. A., & Johnstone, R. A. (2008). Reproductive conflict and the separation of reproductive generations in humans. *Proceedings of the National Academy of Sciences, 105*(14), 5332–5336. https://doi.org/10.1073/pnas.0711911105

Engelhardt, S. C., Bergeron, P., Gagnon, A., Dillon, L., & Pelletier, F. (2019). Using Geographic Distance as a Potential Proxy for Help in the Assessment of the Grandmother Hypothesis. *Current Biology, 29*(4), 651-656. e3. https://doi.org/10.1016/j.cub.2019.01.027

Hawkes, K., O'Connell, J. F., Jones, N. G. B., Alvarez, H., & Charnov, E. L. (1998). Grandmothering, menopause, and the evolution of human life histories. *Proceedings of the National Academy of Sciences, 95*(3), 1336. https://doi.org/10.1073/pnas.95.3.1336

Hawkes, K., O'Connell, J. F., Jones, N. G. B., Alvarez, H., & Charnov, E. L. (2000). The Grandmother Hypothesis and Human Evolution. *In Adaptation and Human Behavior*. Routledge.

Johnstone, R. A., & Cant, M. A. (2010). The evolution of menopause in cetaceans
and humans: The role of demography. *Proceedings of the Royal
Society B: Biological Sciences, 277*(1701), 3765–3771. https://doi.
org/10.1098/rspb.2010.0988

Strassmann, B. I. (2011). Cooperation and competition in a cliff-dwelling people.
Proceedings of the National Academy of Sciences, 108(Supplement_2),
10894–10901. https://doi.org/10.1073/pnas.1100306108

Strassmann, B. I., & Garrard, W. M. (2011). Alternatives to the Grandmother
Hypothesis. *Human Nature, 22*(1), 201. https://doi.org/10.1007/s12110-
011-9114-8

Williams, G. C. (2001). Pleiotropy, Natural Selection, and the Evolution of
Senescence. *Science of Aging Knowledge Environment, 2001*(1), cp13–
cp13. https://doi.org/10.1126/sageke.2001.1.cp13

Excellence in Upper-Level Writing (Social Sciences)

Maintaining peace and biodiversity in Colombia
by Ethan VanValkenburg
From AAS 322/ENVIRON 335: Introduction to Environmental Politics: Race, Class and Gender
Nominated by Adunbi Omolade

In this paper, Ethan grapples with the aftermath of the 2016 Colombian peace agreement and its impact on people and the environment. With incredible nuance, he examines how the treaty has given the government access to indigenous peoples in former rebel territories. It is the depth of Ethan's theoretical analysis of the government's intentions and actions that makes this paper a true gem. He uses theories such as "slow violence," "nature commodification," and "population legibility" to explore the concealed aspects of government policy. He concludes by arguing that a truly post-conflict Colombia needs to prioritize environmental justice by leveraging indigenous knowledge and expertise. Ethan's paper is thorough, deep, and detailed and a must-read for anyone interested in Colombia's efforts at making peace.

-- *Adunbi Omolade*

Maintaining peace and biodiversity in Colombia

I. Introduction

Colombia harbors the second highest species diversity of any country in the world and a tumultuous history of conflict between the government, paramilitary groups, cartels, and indigenous peoples. While conflicts among these varied interests have simmered perennially, Colombia has sought to address them directly in the past decade through peace agreements, development plans, conservation projects, and economic growth. The stated intention prioritizes rural communities and just transitions to better lives, but the realized results have had unintended consequences. This paper will explore those consequences, attempting to unravel how Colombia's management of biodiversity, peace, and development in Colombia has partially failed the people and their environment.

Colombia represents many Afro-Colombian, Indigenous ethnic groups, and a mestizo and white majority, many of which are descendants of Spanish colonizers ("Colombia - The World Factbook," n.d.). Conflict and violence have accompanied these diverse groups, ranging from European colonization, independence wars, civil wars, and *La Violencia*, the latest of which involves a half-century long armed conflict between Colombia and the FARC-EP (Revolutionary Armed Forces of Colombia—People's Army) (Unda & Etter, 2019). The paramilitary group sought to control territory, continue illicit drug production, and gain power in the country, sustaining a conflict of over 50 years and 220,000 deaths (Turkewitz, 2021). In 2016, the Colombian Government signed a peace agreement with the FARC, a compromise involving rural reform, crop replacement, and an end to the conflict (Final Agreement, 2016). Consequently, these reforms have invoked larger social, political, economic, and environmental challenges.

These complicated cultural and political dynamics have been occurring among the Andean-Amazonian foothills, the rainforests of the Pacific region, and the tropical savannas of the Orinoquía. This ecological corridor at the confluence of Central and South America maintains a diverse array of plants and animals (Hoffmann et al., 2018).

Since the start of the 21st century, more than a third of the Colombian forests have shifted to cattle production, urbanization, and colonization in the lowlands (Clerici et al., 2020). Concurrent with these land use changes, warming and water insecurity are projected to increase, placing further stress on the environments and the people they support. According to the Intergovernmental Panel on Climate Change, projected warming scenarios predict warming by as much as 3-4° C with precipitation severely declining by midcentury (Salazar et al., 2018). These dramatic environmental stressors emphasize the need for effective conservation and an acknowledgement of the stress they will place on rural communities and economies.

This paper attempts to explore the ways these diverse cultural, political, and natural histories intersect with the current efforts to maintain peace in Colombia after their resolution to end the decades-long conflict with the FARC. The accompanying goals of promoting social and environmental sustainability have been largely ineffective. The past five years of new development programs, natural resource extraction projects, green-grabbing, and legibility promoted by the state threaten continued peace and conservation. These efforts hinder a truly post-conflict Colombia that promotes environmental justice and the well-being of communities and biodiversity.

II. Development programs and the commodification of nature

Main components of the Colombia-FARC peace agreement include the reduction of jail time for participants in the conflict, political participation of the FARC party, and amnesty, but the most ambitious and important components include programs for the substitution of illicit crops and land redistribution efforts (*FAEACBSLP*, 2016). The prospect of crop substitutions and redistribution were also the most well-supported by the public as of 2018 (García-Sánchez & Carlin, n.d.). Recent reporting after 5 years of supposed implementation reveals that support has dwindled for the largely ineffective programs (Turkewitz, 2021). In practice, much of the new development program simply seeks to allocate more funds, resources, and political pressures to existing development strategies. While many of the current projects have not revealed long-term consequences, examining past implementation of development could reveal patterns of environmental injustice in Colombia.

Escobar (2004) describes the history of displacement from development in Colombia in the context of its extremely skewed income distribution, land distribution, and the abundance of displaced indigenous and Afro-Colombian communities. He describes the consistent convergence economy and armed violence having resulted in conflict over the access to territory and resources. These conflicts have left large populations in terrible material conditions under constant threat of further displacement and death. As a response, Colombia has stepped-up military repression and surveillance. This is paired, argues Escobar, with the global constant of development and modernity as displacement-creating processes. Fundamentally, the socio-economic projects undertaken by the Colombian state are at odds with those of peasant ethnic communities (Escobar, 2004).

The state's apparent lack of care for informed consent and consultation with local communities before implementing these development and economic policies contributes to the issue. Examples from the country's past support this narrative. In the mid-1980's the Salvajina Hydroelectric Dam was proposed to promote the capitalist agricultural development in Western Colombia. It had devastating social and ecological consequences for the surrounding communities by altering their traditional farming practices, access to fishing, cultural practices, important transportation routes, and displaced many from their homes (Machado et al., 2017). This illustrates the fundamental difference between the government and community's relationship to the environment. For the state, the environment was merely a variable attached to their development project. By commodifying nature—damming the river for energy, supporting commercial agriculture, and seeing it as a place for resource extraction—they neglected the holistic relationship and the peasant communities' view of environment as territory and home.

Another important motivation of the peace agreement and subsequent development goals in Colombia has been the pursuit of legibility. Legibility for Colombia, as with other governments, involves simplifying and manipulating rural settlements to facilitate state control (Scott, 1998). In the agreement, Colombia promised to incorporate long-neglected communities, pave roads, promote access,

build schools, and offer a better quality of life (*FAEACBSLP*, 2016). While ostensibly beneficial, these targets, when implemented, do not prioritize the actual needs of the peasant communities they are fit to serve. In some ways, increasing legibility can erase important aspects of the culture and traditional relationships with the environment despite good intentions. At its worst, the pursuit of legibility can be seen as a thinly veiled attempt to stimulate the national economy, enable surveillance and access, and allow for the general cataloging and control of the state's people and places. Part of Colombia's agenda involves carefully documenting all the agricultural land to organize and catalog smallholder forms throughout rural Colombia (*FAEACBSLP*, 2016). This is an effort to reduce coca production—the raw ingredient of cocaine—in favor of other farming practices. To do so, Colombia has managed to distribute thousands of land titles and initiated hundreds of regional development plans. Nonetheless, progress has been extremely slow, potentially due to the fact of current President Duque's party's allegiance with many of the larger landholders in the country (Turkewitz, 2021). In practice, these efforts have not benefitted peasant communities, resulting in ongoing violence and the growth of dissident groups in the region (Grattan, 2020).

III. Conservation and displacement

Alongside the social and political goals of the Colombia-FARC peace agreement (*FAEACBSLP*, 2016), Colombia has sought to prioritize the conservation of its natural resources and natural areas. This interest has been echoed by many scientists and global conservation organizations given Colombia's prominence as one of the most biodiverse and species rich countries in the world (Hoffmann et al., 2018). Global concern about its biodiversity and deforestation has only increased as natural resource extraction, colonization, and development in demilitarized regions have increased (Reardon, 2018). This growing concern presents threats of green-grabbing and displacement, demanding a close examination of the history of conservation in Colombia and policy measures since the agreement.

The peace agreement was only a small part of a much larger national initiative of environmentalism. Other policy instruments include low carbon emission strategies,

UN Convention of Biodiversity policies, and several National Natural Parks and National Natural Preserves (Salazar et al., 2018). These parks, which have only been buttressed by the peace agreement, uphold a long history of green-grabbing and a conservation of displacement. The conservation of Tayrona National Park in Colombia has displaced rural communities and restricted their access to livelihood strategies of fishing and farming (Ojeda, 2012). Even liberal officials seem to share the perspective that natural resources are *objects* to be exploited, overcome, or protected by restriction. Post-agreement Colombia has continued this pattern by promoting a discourse that converts nature into natural capital which it intends to use to promote a national green economy model that incorporates ecotourism, carbon sequestration, and biodiversity conservation (Forero, 2016). While these are not necessarily bad, the Colombian model of corporate governance and conservation at the national level without the involvement of local peasants results in damaging outcomes for the rural communities.

Despite these strengthened, colonial, and exclusive conservation efforts in Colombia, conservation of natural resources in the region has regressed since the agreement in 2016. In the subsequent 5 years, deforestation increased 177% (Clerici et al., 2020). In just the year of the peace accord, deforestation increased by 44% (Reardon, 2018). This is due in large part to expansions of development into rural areas, including the country's own development projects—establishment of roads, rangelands, and co-opting land for agriculture. Key actors include drug cartels, large landowners, dissidents, and some peasant farmers who have capitalized on the safer rural areas of untapped natural resources after the peace agreement (Murillo-Sandoval et al., 2020a).

A primary reason for the increase in illegal deforestation likely lies with Colombia's pattern of imposed administrative centrality, focusing all the control of protected areas at the level of its national government, removing the autonomy of local decision-making and enforcement. The current conservation plan is stratified in time and location, attempting to assign tasks to different large institutions, while the government maintains its own contradictions in its objectives (Clerici et al., 2020). In juggling attempts to protect biodiversity, promote sustainable development for peasants, and maintain peace, it has fallen short of each of the individual goals. Now,

much of the illegal deforestation, grazing, and production of illicit drugs that results are being carried out by large, commercial agricultural operations and cartels, while smallholder farmers are excluded and remain unsupported (Walker et al., 2013).

The theory of gunpoint conservation represents another proposed explanation for the increase of deforestation in Colombia. The warfare ecology literature explains that conflict has varied effects on natural resources and conservation. In Colombia's case, the FARC maintained longstanding rules of land use governance allowing for limited harvesting of forests and other resource extraction within their own communities (Murillo-Sandoval et al., 2020a). This interacts with other passive conservation mechanisms where conflict discourages development (Clerici et al., 2020). In other instances, the 'refugee effect' forces people to flee, incurring less of a pressure on natural resources. Extraction is further stifled by the focus of governments on controlling the conflict and keeping people safe rather than risk development in strained environments (Murillo-Sandoval et al., 2021).

In other cases, conflict can reduce the institutional capacity to enforce conservation laws, especially in rural areas. As a result, deforestation hotspots are often linked to narcotrafficking corridors. Protected areas can overlap with centers of drug production, prompting the expansion of roads, airstrips, coca production, and drive deforestation (Murillo-Sandoval et al., 2020a). However, many of these practices persist in the post-conflict Colombia, and are even expanding, as alternative development programs have failed to meet the needs of some rural populations.

Inevitably, a narrative could develop wherein conflict preserves nature. This misses the point and neglects the dysfunctional nature of the *de facto* conservation and the more systemic and deep-rooted problems. The inverse is also true; many international biodiversity conservation organizations demand fortress-protected areas and promote green-grabbing practices (Negret et al., 2017). Many argue that protected conservation areas and biodiversity reserves are essential for the conservation of biodiversity (Clerici et al., 2020). These policies can lead to environmental injustice and the slow violence of displacement from conservation, observed globally and within Colombia's own history.

One common point advocates for the potential of ecotourism to stimulate the

economy while promoting conservation (Clerici et al., 2016). Ostensibly, ecotourism provides an opportunity where the economy benefits from the conservation of nature, allowing for fortress biodiversity reserves that offer employment to local community members. Inevitably, however, these strategies have severe consequences for local peasants and indigenous communities. In the case study of Tayrona National Park in Colombia, Ojeda (2012) describes the "touristification" of the local community that led to a complete restructuring of the people's local resource access and control. The neoliberal conservation initiatives were pitched as justifiable sources of capital that would attract "conscious tourists" as a solution to environmental depletion. They even cast rural communities as beneficiaries, but local people were inevitably seen as the invaders and illegal occupants who were not welcome in their own land. As a result, traditional livelihood practices, agriculture, fishing, and general access were completely transformed. Such a pattern has been repeated elsewhere in colonial fortress conservation and ecotourism practices all over the world.

Some organizations have recognized the tendency for biodiversity conservation to sacrifice vulnerable communities. Colombia presents a unique opportunity to implement alternative conservation that prioritizes local actors and involves them in the decision-making and enforcement practices, ensuring that their needs are met alongside that of the environment. Because Colombia is in the process of changing many aspects of their political, social, and economic structures, especially in rural areas, they should develop stable governance structures that attend to the interests of local actors (Murillo-Sandoval et al., 2020b). Jaime Góngora, a wildlife geneticist from the University of Sydney, has been a vocal advocate for involving local community members. He has been involved with research institutions in Colombia, surveying territory previously held by the FARC for biodiversity research. While most research projects tend to involve international elites, Góngora has been training ex-FARC combatants in the study of their native plants, animals, and conservation science. These workshops and involvement were initiated after a survey of ex-combatants revealed that 84% of them were interested in jobs related to terrestrial and river environmental restoration (Pérez Ortega, 2020). Presumably, this could empower local communities to be better

advocates for their local environments. So long as these trainings are not imposed, they offer opportunities for peasant communities to join conservation, research, and decision-making—activities that are historically reserved for international or national groups. This implementation could potentially enable conservation without making local peasants unwelcome and vilified occupants. Instead, it could grant more agency in local conservation and encourage participation in global research efforts.

Even still, the green economy narrative may not suffice to benefit rural communities and their environment. Many of the proposed conservation frameworks—ecotourism, carbon offsets, protected areas—uphold the scenario in which the environment is at the service of the corporation and state (Forero, 2016). This prioritization is reminiscent of the original causes of distrust between rural communities and the national government and a primary cause of the Colombia-FARC conflict. In order to maintain peace and benefit peasant communities, conservation practices must address the underlying causes of tension and conflict in the region.

IV. Their knowledge, their nature

Conservation, environmental justice, rural economies, and peace inevitably intersect, especially in a place with as much biological and cultural diversity as Colombia. Therefore, no strategy should treat these linked aspects as separate, but instead develop plans that directly support justice and agency for rural communities, empowering them to be stewards of their own land and advocates for best conservation practices.

Colombia's tendency towards a peace agreement and development plan that treats these components as separate is a symptom of a tendency to divide the well-being of nature and humans, neglecting the varied ways that they depend on each other. Renewing and supporting Indigenous conceptions of human relationships and responsibilities to nature could contribute important frameworks for maintaining peace, justice, and biodiversity and Colombia. Traditionally, however, Colombia's national government has devalued Black and Indigenous people and their livelihoods. Despite efforts to consult with communities and arrive at consent, the underlying power asymmetries and capitalist incentives sacrifice vulnerable communities. Nowhere

is this more evident than the history of the Salvajina Dam in Western Colombia. The dam sacrificed the livelihoods of poor, Black communities for capitalist development. In an interview with Machado et al. (2017), community members expressed, "[t]he government says that the common good is more important than the individual good. Are we not a community, part of that common good?" (Machado et al., 2017) While the project was ostensibly beneficial for the *people*—or more accurately the country's GDP, companies' ability to develop land, and powerful stakeholders—this project and others too often sacrifice and devalue the Indigenous and Black communities. The common good has become an excuse for consistently taking advantage of vulnerable people.

A contrast between social movements in Afro-Colombian communities of Western Colombia and the traditional biodiversity discourse reveals the distinct ways grassroots, local, and traditional movements differ from national ones in their outcomes and empowerment of communities. Escobar (1998) describes the economic, technological, and managerial applications of dominant biodiversity discourse. Advocates from national and international groups *prescribe* resource use limitations and *impose* restrictions on development and communities from the outside. Fundamentally, they emphasize loss of habitats rather than underlying *social* causes of biodiversity loss. Alternatively, the theory of "biodemocracy" advocates from progressive southern NGOs in Colombia shifts the liability to the Global North, colonialism, and imperialism as sources of the biodiversity crisis. Requirements of the holistic alternative include "local control of natural resources, suspension of megadevelopment project, halting diversity-destroying capital activities, and supporting practices based on the logic of diversity and a redefinition of productivity (Escobar, 1998, pp. 59–61)." Cultural autonomy and movements of Black communities in the Pacific rainforest of Colombia have demonstrated the effectiveness of this political ecology. Starting in the early 1990s, activists in Black communities sought to reaffirm their cultural identities and seek "cultural, social, economic, political, and territorial rights" (Escobar, 1998, p. 65). This mobilization united the community and accomplished many of their stated goals. Despite *biodiversity*'s absence from the overriding concerns of the movement, their

apparent definition of biodiversity as "territory plus culture" has created a framework that both conserves biodiversity and maintains cultural traditions (Escobar, 1998).

Colombia's national plans should therefore prioritize the support of these local, grassroots movements and facilitate and empower others. Doing so would integrate social justice, development, peace, and biodiversity into movements designed by peasant communities *for* peasant communities. Furthermore, Eisenbarth et al. (2021) argues that community monitoring is more effective at reducing deforestation in monitored areas in Colombia than centralized management from external parties. The programs involve statutory recognition of local community's right and ability to monitor their own forests and offer guidance and support for their needs. As such, conservation and development should center communities; Escobar states, "[a]ny development alternative must articulate a vision of both the present and a possible future based on *collective aspirations*. It must go well beyond the creation of infrastructure and the improvement of material conditions to strengthen local cultures and languages." This holistic and kinship approach could be more effective at biodiversity conservation and better support communities.

V. Conclusion

Colombia still sits at a crossroad, attempting to transition from a past of conflict between paramilitary groups like the FARC, cartels, diverse ethnic groups, and its national government to a future of peace, development, and biodiversity. While the Colombia-FARC peace agreement was a requisite step towards that future, implementation of the agreement and various development and conservation programs jeopardize progress. New development, green-grabbing, ecotourism, and the imposition of legibility have failed to benefit and gain the trust of the rural people. In failing to do so, they are losing their grip on peace as new cycles of violence reemerge (Turkewitz, 2021). The FARC has maintained a foothold in some rural areas surrounding the protected reserves, building economic and social networks to rebuild their standing and recruit new members (Clerici et al., 2020). Failures of the state could accelerate this process, ending the peace agreement without sustained benefit to the environment or people. To

redeem the agreement and bolster trust from these communities, Colombia must center the well-being and agency of peasant communities and provide equal consideration for their needs. The Colombian Government should seize this unique opportunity of transformation and seek solutions that will benefit all people and environments, resulting in more sustainable resolutions to conflict, societies, and environment.

VI. References

Clerici, N., Armenteras, D., Kareiva, P., Botero, R., Ramírez-Delgado, J. P., Forero-Medina, G., Ochoa, J., Pedraza, C., Schneider, L., Lora, C., Gómez, C., Linares, M., Hirashiki, C., & Biggs, D. (2020). Deforestation in Colombian protected areas increased during post-conflict periods. *Scientific Reports, 10*(1). https://doi.org/10.1038/s41598-020-61861-y

Clerici, N., Richardson, J. E., Escobedo, F. J., Posada, J. M., Linares, M., Sanchez, A., & Vargas, J. F. (2016). Colombia: Dealing in conservation. *Science, 354*(6309), 190. https://doi.org/10.1126/SCIENCE.AAJ1459/ASSET/CD62CFA9-068C-4390-8DA9-9256CC1E5F30/ASSETS/GRAPHIC/354_190B_F1.JPEG

Colombia - The World Factbook. (n.d.). In *CIA*. Retrieved December 13, 2021, from https://www.cia.gov/the-world-factbook/countries/colombia/#people-and-society

Eisenbarth, S., Graham, L., & Rigterink, A. S. (2021). Can community monitoring save the commons? Evidence on forest use and displacement. *Proceedings of the National Academy of Sciences of the United States of America*, 118(29). https://doi.org/10.1073/PNAS.2015172118/-/DCSUPPLEMENTAL

Escobar, A. (1998). Whose Knowledge, Whose nature? Biodiversity, Conservation, and the Political Ecology of Social Movements. In *Journal of Political Ecology* (Vol. 5).

Escobar, A. (2004). Development, Violence and the New Imperial Order. *Development 2004* 47:1, 47(1), 15–21. https://doi.org/10.1057/PALGRAVE.DEVELOPMENT.1100014

Final Agreement to End the Armed Conflict and Build a Stable and Lasting Peace,
(2016). https://www.peaceagreements.org/wview/1845/Final%20
Agreement%20to%20End%20the%20Armed%20Conflict%20and%20
Build%20a%20Stable%20and%20Lasting%20Peace

Forero, L. F. (2016). Territorial Peace and Land Grabbing in Colombia.
Transnational Institute. https://www.tni.org/en/article/territorial-peace-and-
land-grabbing-in-colombia

García-Sánchez, M., & Carlin, R. E. (n.d.). *The FARC in the Public Eye: Negotiation,
Integration, and Political Participation.*

Grattan, S. (2020, October 24). Four years after FARC peace deal, Colombia
grapples with violence | Conflict News | Al Jazeera. *Al Jazerra*. https://www.
aljazeera.com/features/2020/11/24/four-years-after-peace-
deal-colombia-grapples-with-violence

Hoffmann, C., García Márquez, J. R., & Krueger, T. (2018). A local perspective
on drivers and measures to slow deforestation in the Andean-Amazonian
foothills of Colombia. *Land Use Policy*, 77, 379–391. https://doi.
org/10.1016/J.LANDUSEPOL.2018.04.043

Machado, M., López Matta, D., Campo, M., Escobar, A., & Weitzner, V. (2017).
Third World Quarterly Weaving hope in ancestral black territories in
Colombia: the reach and limitations of free, prior, and informed
consultation and consent. *Third World Quarterly, 38*(5), 1075–1091.
https://doi.org/10.1080/01436597.2017.1278686

Murillo-Sandoval, P. J., Gjerdseth, E., Correa-Ayram, C., Wrathall, D., van den
Hoek, J., Dávalos, L. M., & Kennedy, R. (2021). No peace for the
forest: Rapid, widespread land changes in the Andes-Amazon region
following the Colombian civil war. *Global Environmental Change, 69*,
102283. https://doi.org/10.1016/J.GLOENVCHA.2021.102283

Murillo-Sandoval, P. J., van Dexter, K., van den Hoek, J., Wrathall, D., & Kennedy, R. (2020a). The end of gunpoint conservation: forest disturbance after the Colombian peace agreement. *Environmental Research Letters, 15*(3). https://doi.org/10.1088/1748-9326/ab6ae3

Negret, P. J., Allan, J., Braczkowski, A., Maron, M., & Watson, J. E. M. (2017). Need for conservation planning in postconflict Colombia. *Conservation Biology, 31*(3), 499–500. https://doi.org/10.1111/COBI.12935

Ojeda, D. (2012). Green pretexts: Ecotourism, neoliberal conservation and land grabbing in Tayrona National Natural Park, Colombia. *https://doi.org/10.1080/03066150.2012.658777, 39*(2), 357–375. https://doi.org/10.1080/03066150.2012.658777

Pérez Ortega, R. (2020). A conservation scientist enlists Colombia's ex-guerrillas in a new cause: preserving their country's biodiversity. *Science*. https://doi.org/10.1126/SCIENCE.ABD7316

Reardon, S. (2018). FARC and the forest: Peace is destroying Colombia's jungle - and opening it to science. *Nature, 558*(7709), 169–170. https://doi.org/10.1038/D41586-018-05397-2

Salazar, A., Sanchez, A., Villegas, J. C., Salazar, J. F., Ruiz Carrascal, D., Sitch, S., Restrepo, J. D., Poveda, G., Feeley, K. J., Mercado, L. M., Arias, P. A., Sierra, C. A., Uribe, M. del R., Rendón, A. M., Pérez, J.C., Murray Tortarolo, G., Mercado-Bettin, D., Posada, J. A., Zhuang, Q., & Dukes, J. S. (2018). The ecology of peace: preparing Colombia for new political and planetary climates. *Frontiers in Ecology and the Environment, 16*(9), 525–531. https://doi.org/10.1002/FEE.1950

Scott, J. C. (1998). Nature and Space. In *Seeing Like a State* (pp. 11–22). Yale University Press.

Turkewitz, J. (2021). Five Years After Peace Deal, Colombia Is Running Out of
Time. New York Times. https://www.nytimes.com/2021/09/26/
world/americas/colombia-peace-deal-farc.html

Unda, M., & Etter, A. (2019). Conservation opportunities of the land restitution
program areas in the colombian post-conflict period. *Sustainability
(Switzerland), 11*(7). https://doi.org/10.3390/su11072048

Walker, N. F., Patel, S. A., & Kalif, K. A. B. (2013). From Amazon pasture to the
high street: Deforestation and the brazilian cattle product supply chain.
Tropical Conservation Science, 6(3). https://doi.org/10.1177/
194008291300600309

Excellence in Upper-Level Writing (Social Sciences)

Thesis Proposal: Abolition, Mutual Aid, and the Policing of Transgender Communities
by Grey Weinstein
From POLSCI 381: Political Science Research Design
Nominated by Nancy Burns and Erin Cikanek (GSI)

Weinstein develops a powerful proposal, bringing together theoretical work on mutual aid and carceral feminism to develop hypotheses about the conditions under which transgender individuals would have lower levels of interaction with the police. Weinstein offers a sophisticated portrait of the two primary theoretical literatures. Throughout the paper, Weinstein is authoritative, whether Weinstein is writing about theory, hypothesis, or method. The paper is an engaging and deeply practical guide to the project Weinstein will carry out.

-- Nancy Burns

Thesis Proposal: Abolition, Mutual Aid, and the Policing of Transgender Communities

Introduction

Police are widely understood as one arm of the prison-industrial complex within the United States. From an abolitionist perspective, police serve to surveil and control populations, especially marginalized or systematically oppressed groups. Transgender people are one such group. Transgender individuals, especially trans women of color, are notoriously overpoliced; fifty-nine percent of transgender individuals are stopped by police, compared to just twenty-eight percent of their non-LGBTQ peers (Make the Road New York 2012, 15). Yet there is a dearth of research on transgender communities in general, and their interactions with carceral systems in particular. My proposed thesis aims to fill this gap in scholarly research by examining the policing of transgender communities through an abolitionist framework.

How can transgender communities combat the epidemic of overpolicing and incarceration their members are facing? In the absence of meaningful policy reform or abolition from those in power (such as the US government or police unions), many transgender communities establish active mutual aid networks wherein individuals both give and receive resources to one another. Mutual aid within transgender communities is best understood in the context of the social marginalization these communities experience. A significant segment of the transgender community in the United States is widely understood as living on the margins of American society. This can be credited in large part to institutionalized discrimination; transgender people, particularly low-income trans people or trans people of color, face hiring discrimination, housing discrimination, and discrimination by banks and other financial institutions, among other forms of marginalization. At the same time, transgender individuals are also the subject of the intense social stigma surrounding gender transition and gender nonconformity. Disgust and malice are often directed at transgender people by their cisgender peers. This stigma causes many trans youth to be kicked out of their homes, cuts trans people off from familial support networks, and makes it difficult for trans

people to obtain public-facing jobs. Of course, this is not the case for all transgender people; many middle-class and wealthy trans individuals assimilate into mainstream society, a process made all the easier for white trans people due to their racial privilege. Nonetheless, many others, especially trans people of color and those from low-income backgrounds, find themselves confined to the margins of society by these institutional barriers and stigmatization. In this situation, trans people often turn to one another for support, creating networks of mutual aid. Mutual aid is a communal method of addressing poverty and lack of resources, but it also builds an alternative system for meeting the needs of its members. The alternative structure created by mutual aid exists outside of the established systems that often discriminate against or exclude transgender individuals.

My thesis seeks to answer the research question, "Does non-governmental aid provided to transgender populations effectively prevent transgender individuals from unsolicited interaction with the police?" I hypothesize that mutual aid networks that provide housing, gender-affirming healthcare, and assistance legally changing one's name and gender marker have the effect of reducing transgender individuals' unsolicited interaction with the police. I have several reasons to believe this to be the case. Housing allows transgender individuals to remain off of the streets and safely away from police forces who are likely to harass or arrest them, especially in areas of the United States where homelessness is essentially criminalized by laws such as those prohibiting loitering or sleeping in public spaces. Gender-affirming healthcare could potentially decrease transgender individuals' contact with the police by helping them to avoid being profiled as transgender and consequently harassed or arrested because of their gender expression. More specifically, it is a common experience for trans people to be harassed or stopped by the police simply because they are transgender, especially for trans women who are often accused of engaging in sex work. (This will be discussed further in my literature review.) By aligning their appearance with their gender identity, gender-affirming healthcare such as hormone replacement therapy or surgery can help transgender people "pass" as cisgender. This in turn has the potential to decrease the likelihood that they will be stopped by police simply for appearing

transgender. Finally, having legal identification with one's name and gender assigned at birth can be dangerous for transgender individuals. A transgender woman who has medically transitioned and who "passes" as a cisgender woman, for example, will likely have the legitimacy of her ID questioned if it contains a male name or gender marker. In many situations, suspicion of possession of a false ID will lead to police involvement, potentially risking the safety of the transgender individual.

By posing this research question, I hope to unite scholarly debate about mutual aid as an anti-hierarchical (and potentially anti-capitalist) tool with research on the targeting and overpolicing of transgender communities. By directly engaging with transgender individuals through survey research, I aim to center the lived experiences of trans people in this thesis. I also want to bring an intersectional approach to my data collection and analysis by scrutinizing the racial and class disparities in policing and access to mutual aid. By centering this research within an abolitionist framework, my thesis will endeavor to discover where, if at all, mutual aid fits into an abolitionist vision of transgender liberation from mass incarceration.

Literature Review

My research question and hypothesis are built on the assumption that transgender individuals will be safer and have a higher quality of life if their interaction with police is kept to a minimum; to justify this assumption, my literature review will examine the policing and incarceration of transgender populations through an abolitionist framework. I will next discuss mutual aid as an anti-capitalist and anti-authoritarian system, and as a tool used by incarcerated transgender individuals for survival and resistance. Finally, I engage with the critiques of carceral feminism, which posits that carceral systems such as prisons and police should be employed to protect vulnerable populations (mainly women). I examine carceral feminism as an alternative framework to abolitionism which presents a vision of a reformed police force that defends transgender women from abuse. I will conclude with research that argues against the effectiveness of carceral feminism when applied to transgender issues.

Abolitionist Theory

This research will study the concept of police surveillance within the framework

of abolitionist theory. Abolitionist theory advocates for complete dismantlement of the prison-industrial complex; the abolitionist argument is rooted in history and a racial analysis of the prison-industrial complex.

Angela Davis (2003) defines the prison-industrial complex as "an array of relationships linking corporations, government, correctional communities, and media" (Davis 2003, 84). Davis uses qualitative historical analysis to draw conclusions about the prison-industrial complex. She traces the history of the prisons to the development of penitentiaries which, through poor living conditions and forced labor, closely mirrored the conditions of slavery. After the Civil War, she argues, Black Codes were created to criminalize "vagrancy, absence from work, breach of job contracts, the possession of firearms, and insulting gestures or acts [...] only when the person charged was black" (Davis 2003, 28). This intentionally incarcerated Black people at extremely high rates. As a result, Black communities were continually forced to work in penal servitude, since slavery and involuntary servitude remain legal in prisons under the Thirteeth Amendment (Davis 29). Penitentiaries also had the effect of assigning criminality to Black people, making the concept of criminality highly racialized (Davis 2003, 30).

Shifting her focus to the modern day prison-industrial complex, Davis (2003) finds that prisons are deeply entwined with systemic racism, capitalism, and militarism. Her research brings her to the conclusion that complete abolition is the most effective and pragmatic way to address the abuses of the prison-industrial complex. Davis first contends that prisons are increasingly linked to the military industrial complex. She explains that police and correctional officers arm themselves with military grade weapons; this in turn funnels a huge amount of money into the defense industry (Davis 2003, 86-87). She also argues that prisons are firmly rooted in an exploitative capitalist economic system in several ways. Prisons are increasingly tied to corporations which treat prisoners as "a potential source of profit" (Davis 2003, 89). Companies like private construction firms, healthcare corporations, food providers, etc. make enormous profits from their contracts with prisons. As a result, "public prisons have become so thoroughly saturated with the profit-producing products and services of

private corporations" (Davis 2003, 99-100) that they have an economic incentive to perpetuate the prison-industrial complex. These vast imprisoned populations, mainly consisting of Black men, are exploited for their underpaid or unpaid labor. This likewise generates a huge profit at the expense of human life, which in turn creates an incentive to continue to incarcerate people so that the exploitative prison labor system may remain intact. Because Black Americans are incarcerated at disproportionate rates in comparison to their hite counterparts, it also continues to maintain a racialized system of labor exploitation (Davis 2003, 94-95). Because the prison-industrial complex is so deeply and inherently racist, and because it is so firmly entrenched in existing capitalist and military industrial institutions, Davis argues that we ought to abolish prisons completely rather than attempt some sort of reform (Davis 2003, 103).

Police Surveillance as a Means of Control

Systems of policing are a key component of the prison-industrial complex as defined by Davis. Within an abolitionist framework, policing is conceptualized as a means to surveil and discipline populations. Foucault (1975) defines discipline as control over the body; by controlling how bodies move through space, institutions can wield discipline to control the actions and movements of entire groups (Foucault 1975, 138). He found that the carceral system assigned an "intrinsically criminal character" to those it deemed "delinquent" in order to control and discipline these groups (Foucault 1975, 100). This phenomenon can be observed in the modern prison-industrial complex through the behavior of the police towards marginalized groups. Dorothy E. Roberts (2017) studies loitering laws, which are used to disproportionately target Black Americans (Roberts 2017, 1601). Roberts further argues that these loitering laws promote racist notions of criminality by labeling Black so-called loiterers "visibly lawless" (Roberts 2017, 1600), essentially conflating Blackness with criminality in order to legitimate police harassment and regulate how Black people are allowed to show up in public spaces. This increased harassment leads to the detainment of Black people and serves to keep them in poverty as a result.

Anna Akbar (2020) argues from an abolitionist perspective, finding that policing and incarceration are mutually reinforcing structures that seek to perpetuate

violent control over marginalized groups. Criminalization places blame on individuals and their actions for what society deems social ills, such as poverty (Akbar 2020, 1822). Rather than blame systems of inequality or poor policy, individuals who steal, loiter, sleep on public property, etc. are labeled "criminal" for existing as poor. By shifting blame away from inequitable systems, policing strengthens these systems by allowing them to continue their existence (Akbar 2020, 1823). Because law enforcement targets low-income people, people of color, and transgender people, guilt and criminalization are assigned in a manner that is racialized, gendered, and divided by class (Akbar 2020, 1824).

Police violence is a tool through which to discipline and subjugate criminalized populations. Within an abolitionist framework, violence is argued to be inherent to policing because of the pervasive lack of consequences within the criminal justice system for police use of force. Akbar finds that police violence is authorized by law; "judges allow police to 'do mostly anything they want,' including harassment, brutalization, killing, SWAT raids, body cavity searches, and the use of chemical weapons" (Akbar 2020, 1790-1791). Because the police are incentivized to stop people in order to generate profit from traffic stops and other fines or fees, and because use of force is legally permissible by police, violence becomes "enmeshed in routine policing" (Akbar 2020, 1795). Violence is thus deeply institutionalized through the police and it therefore becomes "routine and commonplace" (Akbar 2020, 1795). This violence is used to target mainly low-income people and people of color as well as transgender and gender nonconforming people, thus creating a society stratified by race, class, and to a lesser extent, gender (Akbar 2020, 1978).

By arguing that police surveillance is, at its core, inherently violent, many researchers come to the conclusion that abolition of the police is the only viable solution to police brutality. Akbar, Davis, and Roberts all argue for alternative structures for achieving community safety, structures that intentionally avoid the criminalization of individuals or groups (Akbar 2020; Davis 2003; Roberts 2017). Allegra M. McLeod (2019) examines the case study of Chicago in the late 2010s as abolitionist groups like the #LetUsBreathe Collective and the Movement for Black Lives called for alternate,

abolitionist structures to the criminal justice system. McLeod concludes that abolition was called on by contemporary civil rights activists a necessary measure in order to solve the problem of police torture in Chicago. Use of torture was widespread from the 1970s to 1990s under Commander Jon Burges (McLeod 2019, 1624), but punishing him as an individual would fail to dismantle the structures that allowed him to wield so much power in such a harmful way in the first place. Solely punitive solutions would thus leave the opportunity for someone else to repeat his abuses in the future. Thus, activists turned to abolition of the police as a framework for accountability but also harm prevention (McLeod 2019, 1626-1628).

Transgender Populations within the Prison-Industrial Complex

Transgender individuals have unique (and usually uniquely harmful) experiences within the prison-industrial complex. Similar to the way in which Roberts finds that police systemically target Black people with loitering laws (Roberts 2017, 1600), data from several studies shows that anti-solicitation laws are used by law enforcement to target transgender people, specifically trans women of color. Nadal (2014) conducted a qualitative study which was mainly based on interviews with transgender individuals. Nadal found that being stopped, searched, and arrested by police was a frequent experience for trans women on the (often false) basis that they might be sex workers (Nadal 2014, 180). Similarly, Courtney Daum (2015) finds that transgender people are disproportionately targeted by the police in the enforcement of anti-solicitation laws, not only in the frequency of their detainment and arrest but also in the intensity of force used upon them (Daum 2015, 566-567).

Multiple quantitative studies confirm that transgender people are targeted by police at higher rates than their cisgender counterparts. One 2012 survey conducted by the organization Make the Road New York included both transgender individuals and non-LGBTQ+ individuals, asking both groups whether they had been stopped by the police that year. This survey found that fifty-nine percent of transgender individuals surveyed had been stopped by police, compared to just twenty-eight percent of cisgender heterosexual individuals (Make the Road New York 2012, 15). Similarly, a 2011 survey of transgender individuals by Jaime M. Grant et al. divided transgender

respondents by gender and found that twenty-one percent of trans women and ten percent of trans men had been sent to jail in their lifetime (Grant et al. 2011, 163). Perhaps more significantly, twenty-two percent of respondents had been harassed by police due to gender-based bias (Grant et al. 2011, 158). Likewise, the Human Rights Watch conducted a 2012 study that included interviews with individuals about their interactions with police; although not quantitative, this provided qualitative data about the treatment of transgender and cisgender womens' interactions with police, and led to the conclusion that trans women were widely verbally and physically harassed (Human Rights Watch 2012).

Police surveillance of trans women is not only more frequent but also often more forceful and violent than the treatment received by their cisgender counterparts. By analyzing a number of interviews with transgender people conducted by various researchers, Daum (2015) found that trans women suspected of solicitation by the police were often subjected to indiscriminate use of force. Furthermore, this use of force was often intended to debase and humiliate transgender women on the basis of their gender identity. Daum writes, "Transgender individuals profiled by law enforcement often are verbally and physically harassed and "defaced" by police officers who forcibly remove their breasts and wigs in public spaces" (Daum 2015, 566). A 2012 study by Zamani et al. conducted a series of interviews and found that trans women were "routinely presumed to be sex workers, simply based on their gender or gender expression" and frequently were "inappropriately touched by police or searched by the NYPD in order to 'determine their gender'" (Zamani et. al 2012, 12) The interviews also revealed that "these 'searches' can be aggressive and are often experienced as sexual assaults" (Zamani et al.2012, 12).

Daum argues that the unequal enforcement of anti-solicitation laws is a tool to control trans bodies and behavior, erasing them from public space. She makes the case that this targeting of trans people both criminalizes gender nonconforming bodies and upholds the privilege of gender conforming bodies, thus reinforcing gender normativity. Daum's conclusion is similar to the research of both Davis and Akbar, who both describe police as assigning traits of "criminality" to the Black community, often through the

enforcement of loitering laws, in order to surveil and control Black populations (Akbar 2020; Davis 2003). Likewise, Daum argues that by "profiling transgender individuals engaging in mundane activities" police assign criminality to transgender individuals, allowing them to "elicit fear and exercise control over this population" (Daum 2015, 566). Thus, police use anti-solicitation laws to criminalize gender nonconformity, and as "a tool of social control that governs the behavior of transgender individuals" (Daum 2015, 567) much in the manner described by Foucault.

It is important to note that discrepancies in police harassment, detainment, and use of force between transgender and cisgender populations are amplified among populations of color. While Grant et al. found in their survey of transgender individuals that twenty-two percent had been harassed by the police due to gender based violence, the percentage among transgender respondents of color was between twenty-nine and thirty-eight percent (Grant et al. 2011, 158). Furthermore, although only seven percent of the respondents had been detained in a cell due solely to their gender expression, this number was as high as forty-one percent for Black respondents and twenty-one percent for Latin respondents (Grant et al. 2011, 158). Just as cisgender people of color face more police discrimination and violence than their white cisgender counterparts, so to do transgender people of color experience worse profiling and treatment by police than white transgender people.

Mutual Aid as an Abolitionist Solution

Faced with the problem of police discrimination and police brutality, abolitionists seek the complete dismantlement of policing systems and the creation of compassionate, effective, and safe alternatives. Within our present system, where police presence remains, abolitionists continue to build and advocate for "alternative means of preventing harm" (McLeod 2019, 1628). In her 2019 article, McLeod studies the alternative forms of justice that activists often advocate for within an abolitionist framework. In addition to transformative justice (wherein one addresses interpersonal harm while attempting to make broader political, social, or economic change), McLeod finds that many abolitionists try to reduce harm by offering alternatives to contact with the police (McLeod 2019, 1628). These alternative first responders, who

are often community members, intervene to de-escalate violence or mediate dialogue. McLeod refers to these alternate first responders as a form of mutual aid (McLeod 2019, 1628). She holds up two case studies as prime examples: the Harm Free Zone project in Durham, North Carolina, and the Audre Lorde Project's Safe OUTside the System Safe Neighborhood Campaign in Brooklyn, New York, which "both educate interested community residents and train them to take action to prevent harm without police intervention" (McLeod 2019, 1629).

Spade (2020) defines mutual aid as a project that both changes political institutions and cares for community members by building new, sustainable social relations (Spade 2020, 136). Spade puts forth a three-pronged approach for movements that seek to change material conditions. This approach includes dismantling existing harmful systems (like prisons), caring for those targeted by these systems, and building alternative structures that can meet people's needs. Mutual aid falls under this third pillar (Spade 2020, 134). Arani (2020) asserts that "mutual aid works to build lasting alternatives to state-sponsored systems of care by organizing grassroots networks of support that are reciprocal, transparent, and guided by participants' resources, skills, and knowledge" (Arani 2020, 654). As such, Spade (2020) contends that within a capitalist society where resources must be acquired through relationships based on exploitation, community care through mutual aid is a radical act. Mutual aid is anti-authoritarian in that it builds networks for services and resources outside the existing system without coercion (Spade 2020, 138). In contrast to resources given by charities or governmental organizations, which "frequently blame social problems on individuals' moral failings," mutual aid instead "recognizes that capitalist, white supremacist institutions are responsible for producing poverty, inequality, and violence" (Arani 2020, 654). It brings people together on the basis of shared material needs, building solidarity as a result; it also often has the effect of raising political consciousness as people are invited to consider the ways in which the current system has failed to provide for their needs (Spade 2020, 137). Spade cites the Black Panthers' Free Breakfast for Children initiative as an example; this program not only helped children access the food they needed to survive, but also brought many people who

wanted social change into the Black Panther organization (Spade 2020, 136).

Mutual Aid in Trans Communities

Mutual aid networks play a central role in many transgender communities both within and outside of prisons. Hwang (2019) observes and analyzes the living conditions of incarcerated transgender individuals in the California Substance Abuse Treatment Facility (CSATF) and Prison, Corcoran. They specifically study the Transgender Alternative Lifestyles group, a support group run by transgender inmates within the prison. The group was "led for and by mostly black and Latinx transgender women" and was a space "to air grievances, share resources, address interpersonal conflict, build community trust, and offer gender-affirming somatic exercises to counter past and present traumatic experiences of harm" (Hwang 2019, 567). Hwang argues that this group utilized "practices of mutuality and collective care, advocacy, peer support, dialogue, and strategies toward naming connections between structural, interpersonal, and material violence" (Hang 2019, 567). Thus, the Transgender Alternative Lifestyles Group organized transgender inmates to mutually support one another and help each other access services and resources, all in the face of the prison's institutionalized violence. Hwang therefore identifies the group as a type of grassroots mutual aid network operating within the prison in order to improve transgender inmates' quality of life. Drawing on Spade's research on mutual aid, Hwang argues that mutual aid networks among transgender inmates are anti-hierarchical and thus capable of radical opposition to the carceral, capitalist systems of the prison-industrial complex (Hwang 2019, 563-564).

Furthermore, Hwang's observations of life within the CSATF provide evidence for their claim that queer and trans mutual collective support is effective and empowering. They record how the Transgender Alternative Lifestyle Group organized in reaction to receiving transparent shower curtains that left trans inmates feeling vulnerable to assault and as if their privacy had been violated. By constructing a buddy system to use while showering, and by sharing towels, the trans women managed to keep each other hidden from view and safe during showers (Hwang 2019, 569). Hwang uses this as an example of how "deviant care" provides communal safety and

resistance to the carceral system. They also cite the group's success in obtaining gender-affirming items like women's underwear and makeup for trans inmates as a victory of trans mutual aid. These items provided trans inmates a sense of dignity within a system built to deny them their humanity (Hwang 2019, 568). Thus, although this example of mutual aid is confined within the prison-industrial complex, Hwang defines it as radically anti-carceral. The Transgender Alternative Lifestyle Group creates an egalitarian community based on mutual respect and the desire to meet one another's needs, in opposition to a top-down power system that focuses on degrading trans women and denying their material needs (Hwang 2019, 568). As a result, Hwang writes that "the success of receiving 'female designated' personal items was considered a huge win" because it showed that "organizing among women inside is in and of itself a mode of survival and resistance against the consolidation of hegemonic state power" (Hwang 2019, 568).

In her 2019 study of abolitionist alternatives to policing, McLeod notes that several programs which "provide alternative first responders, mediation support, or other forms of mutual aid to those who would otherwise likely be subject to victimization, arrest, possible police violence, or incarceration" (McLeod 2019, 1628) have a focus on transgender communities. She notes that the Safe Neighborhood Campaign focus on queer and gender nonconforming people of color who are subjected to greater police harassment. This mutual aid network effectively prevents violence by providing mediation skills that enable trans individuals to resolve conflict without police involvement (McLeod 2019, 1628-1629).

Complementing research on mutual aid networks, Bouris and Hereth (2020) argue that it is possible to "operate from an abolitionist framework while also engaging with state systems and structures" (Bouris and Hereth 2020, 368). Bouris and Hereth analyze three different LGBTQ+-focused organizations which do work around criminal justice. They examine Black and Pink, an organization which runs a pen pal program that sends letters to LGBTQ+ inmates. They also study the Transformative Justice Law Project's name change clinic, which connects transgender individuals with lawyers who can help them navigate the legal name change process. Their third case study is

BreakOUT!, which used legislative lobbying to advocate for policies that would ban police from discriminating on the basis of gender identity or expression. Hereth and Bouris conclude that "direct action, service delivery, and policy advocacy" intervene within carceral systems, which "is a crucial harm reduction strategy for preventing LGBTQ+ young people from ever becoming entangled within the criminal legal system" (Bouris and Hereth 2020, 370). However, these harm reduction strategies can be the first step towards the ultimate goal "to dismantle the criminal legal system, or 'starve the beast,' by preventing arrest and incarceration" (Bouris and Hereth 2020, 367). Thus, their analysis places these case studies in a framework of prison abolition and finds that the anti-carceral work these organizations do is rooted in the values of prison abolition. They further conclude that such work can create a meaningful foundation for abolition in the future.

Carceral Feminism as an Alternative to Abolition

In contrast to abolition, some activists seek to use the power of the state and the prison-industrial complex as tools to combat institutional misogyny and patriarchy. One central example is the campaign for the Violence Against Women Act of 1994. The National Organization for Women advocated for the passage of the Violence Against Women Act as a way to protect women from domestic violence. They argue that "violence against women is perpetrated by a culture that tacitly tolerates misogyny" and that as a result of this misogyny there is "harmful gender bias in law enforcement and judicial systems, leading to inadequate investigations and prosecution in violent crimes against women" (National Organization for Women 2021). In this view, more thorough investigation and harsh prosecution of domestic violence crimes would better protect women. By locking up dangerous, violent, and misogynistic abusers, the carceral system can serve women by preventing domestic abusers from reoffending (and thus protecting their potential victims) and by serving justice to those who have survived abuse. In support of the Violence Against Women Act, the women's legal defense and educational fund Legal Momentum called for an increased number of "police officers, prosecutors, judges, probation and corrections officials" among others to respond to domestic violence incidents (Legal Momentum).

Indeed, the Violence Against Women Act was passed as part of a larger crime bill, the Violent Crime Control and Law Enforcement Act of 1994; it allocated $9.7 billion of funding to prisons along with additional funds for 100,000 new police officers (US Department of Justice 2021).

Similarly, Zorza and Woods (1994) argue in favor of mandatory arrest laws, which require police to make an arrest after receiving a domestic violence call. They argue that such laws "increase arrest and shorten the time involved in putting victims in contact with helping agencies. Mandatory arrest also communicates to the entire community that domestic assault is a crime" (Zorza and Woods 1994). Zorza and Woods find that although it is not sufficient to solve the problem of domestic violence alone, "as part of a carefully planned community system with sufficient resources and monitoring, mandatory arrest can improve protection for battered women and their children" (Zorza and Woods 1994). As of 2008, 22 states (and the District of Columbia) all had adopted mandatory arrest laws (Hirschel 2008, 19). These laws, like the Violence Against Women Act, strengthen carceral systems by implementing increasingly harsh penalties against perpetrators of domestic abuse.

The Violence Against Women Act, mandatory arrest laws, and other initiatives like them have been widely criticized by abolitionists. "Carceral feminism" was coined by Bernstein (2007) to describe "within the context of the U.S. antirape and battered women's movements, a drift from the welfare state to the carceral state as the enforcement apparatus for feminist goals" (Bernstein 2007, 143). While Bernstein's focus was on anti-trafficking legislation, Law (2014) specifically targeted VAWA and mandatory arrest laws as examples of carceral feminism because they operate on the belief that "policing and prisons [are] the solution to domestic violence" (Law 2014).

Whalley and Hackett (2017) argue against the effectiveness of carceral feminism. The researchers offer a comprehensive review of feminist political theory as applied to various feminist organizations, including both carceral feminist groups (such as specific crisis rape centers) and abolitionist groups. Whalley and Hackett synthesize past research on these organizations to draw conclusions about the efficacy of these programs through qualitative analysis. They conclude that activism driven

by an ideology of carceral feminism is ineffective at protecting women from gender based violence. This is because by advocating for punitive solutions, carceral feminist organizations serve to legitimize and strengthen state carceral systems such as the prison-industrial complexes. In doing so, they strengthen the very institutions that oppress and harm many women of color, transgender women, low-income women, and female immigrants (Whalley and Hackett 2017, 464-466).

The phenomenon of vulnerable groups turning to the very carceral systems that target them for protection can be observed within the LGBTQ+ community as well. Russell (2018) observes that "many LGBTI individuals and organizations have precipitated and embraced efforts to expand criminalization in the name of LGBTI protection, such as hate crime sentencing laws and policing strategies" (Russell 2018, 334).

If the premise of my hypothesis is incorrect, and policing does not materially harm transgender individuals in a significant manner, then carceral solutions to transphobia offer an alternative solution. However, preliminary research suggests that this is not the case. One of the prime examples of carceral solutions to transphobia is K6G, a unit within the Los Angeles County Men's Central Jail exclusively housing homosexual male inmates and transgender female inmates who had been incarcerated within the men's jail. Hwang (2016) observes the conditions within K6G as part of a case study. They report poor access to trans healthcare and healthcare in general, mental health resources, and educational resources, as well as wide abuse of power by correctional officers who regularly use force unnecessarily against inmates and punish them arbitrarily (Hwang 2016, 90-93). Furthermore, they find few avenues for inmates to hold these correctional officers accountable, partly because few legal resources were available to trans inmates. However, the lack of accountability was perpetuated by positive press that praised the prison for recognizing the needs of gay and trans inmates. This press made it difficult for organizers to gain the public support needed to effectively oppose the prison's abuse of power (Hwang 2016, 90). Thus, Hwang found that carceral solutions only strengthened the very institutions that target trans women of color in the first place, further perpetuating transphobia.

Similarly, Sumner and Sexton (2016) administered in-depth interviews with ten transgender prisoners, a random sample of 27 cisgender prisoners, and a sample of 20 staff who were said to have the most interactions with transgender prisoners, across four state prisons for men. Summer and Sexton conclude that across all four prisons, cisgender inmates, transgender inmates, and staff members were hyperaware of the fact that the gender identity of transgender inmates transgressed the masculine norm of male prisons (Summer and Sexton 2016, 627-629). They further conclude that this acknowledgment of trans women's marginalized status leads them to be victimized by violence or ostracized and met with disgust and scorn in the hypermasculine setting of a male prison (Summer and Sexton 2016, 629-63).

The data provided by Hwang's case study of the K6G unit and Summer and Sexton's interviews support Russell's conclusion that "as certain queer subjects seem to transition from one side of 'the criminalization/protection line' to the other (Lamble 2013, 232), the unevenly distributed violence of the law itself remains largely unchallenged" (Russel 2018, 334). In other words, just as Law argues, along with Whalley and Hackett, that strengthening carceral systems in the name of feminism only serves to harm marginalized women further, so too does the strengthening of carceral systems in the name of queer rights only further target vulnerable transgender populations. I believe that these findings justify my research's focus on mutual aid as a means to decrease contact with the police, as part of an abolitionist strategy of harm reduction.

Hypotheses

My primary hypothesis is that access to non-governmental aid decreases the amount of unsolicited contact transgender individuals experience with police. However, if I am incorrect and an individual's access to non-governmental aid does not affect their level of police interaction, then race and income level provide an alternate explanation as to why some transgender individuals experience higher rates of police contact than others. Thus, I have two alternative hypotheses. The first is that transgender individuals who belong to a marginalized racial group are more likely to interact with police, regardless of the amount of aid they have access to. The second

is that transgender individuals who receive a lower income are more likely to interact with police, regardless of the amount of aid they have access to.

The foundations of all three of my hypotheses rest upon the underlying assumption that unsolicited contact with the police is generally a negative experience which causes harm to transgender individuals, and is thus an experience that ought to be minimized. Although there is ample scholarly debate that supports this conclusion (as presented in the abolitionist research discussed in the above literature review), that the police cause net harm to transgender communities is not a universally accepted fact. If in fact the police do not overwhelmingly negatively impact transgender communities, carceral feminism may perhaps provide a viable alternative to abolitionism as a means of addressing the needs of the transgender community. Thus, my research will account for the fact that there is a possibility that transgender people tend to have positive opinions of the police, and/or that they tend to have positive experiences with the police. I will thus include four additional hypotheses. The first is that transgender individuals have negative experiences with and low opinions of the police. The second is that transgender individuals have positive experiences with and high opinions of the police. The third is that transgender individuals have negative experiences with the police but high opinions of the police. The fourth and final additional hypothesis is that transgender individuals have positive experiences with the police but low opinions of the police.

For the sake of clarity, my seven hypotheses (including my primary hypothesis, two alternative hypotheses, and four additional hypotheses) are listed in the table below.

Primary Hypothesis (H1)	Access to non-governmental aid decreases the amount of unsolicited contact transgender individuals experience with police.
First Alternative Hypothesis (HA1)	Transgender individuals who belong to a marginalized racial group are more likely to interact with police, regardless of the amount of aid they have access to.
Second Alternative Hypothesis (HA2)	Transgender individuals who receive a lower income are more likely to interact with police, regardless of the amount of aid they have access to.
First Additional Hypothesis (Ha1)	Transgender individuals have negative experiences with and low opinions of the police.
First Additional Hypothesis (Ha2)	Transgender individuals have positive experiences with and high opinions of the police.
First Additional Hypothesis (Ha3)	Transgender individuals have negative experiences with the police but high opinions of the police.
First Additional Hypothesis (Ha4)	Transgender individuals have positive experiences with the police but low opinions of the police.

Variables

The primary dependent variable for this experiment is the amount of unsolicited contact that transgender people have with the police. "Unsolicited contact with police" will be defined as all interpersonal interactions between the individual and one or more police officers in which the police (and not the transgender individual)

initiated the interaction. I will count interactions in which the subject is both alone or in a group. This definition includes being stopped by police either on foot or in a vehicle, being searched by police in any manner, verbal or physical police harassment, and detention (including in a police vehicle, jail, sheriff's office, or other type of holding cell). It does not include instances in which the transgender individual called the police or asked for police assistance.

The secondary dependent variable is the quality of the experiences of transgender individuals who have interacted with the police. This will be measured in two ways; first, by whether transgender people have positive, neutral, or negative interactions with the police, and second, by what opinions trans individuals have towards police as an institution.

The primary independent variable is the type of non-governmental aid provided to the transgender community. For the purposes of this study, I will examine the effects of three main types of aid: access to gender-affirming healthcare, housing, and legal assistance with the name change and gender marker change process. I will compare these variations of the independent variable to each other as well as to a fourth variation, those who receive no resources.

Constraints around data collection have led me to conclude that in order to gather enough data, I must partner with nonprofit organizations that serve transgender communities. Thus while my research is largely rooted in a scholarly conversation around mutual aid and its implications, I will also include data from transgender respondents who have received gender-affirming healthcare, housing, and legal assistance with the name change and gender marker change process from nonprofit organizations. I plan to ask these nonprofit organizations to circulate my survey to their clients in order to obtain a greater sample size. Admittedly, nonprofits are often criticized for being hierarchical, top-down organizations, in contrast to the equitable, anti-hierarchical organization of mutual aid. However, for the purpose of this study I believe that since nonprofit organizations accomplish the task of distributing nongovernmental aid and resources to individuals free of charge and with few requirements, including their clients in the data collection process will enrich my research. For this reason, I

have defined my independent variable as "non-governmental aid" to include resources provided by mutual aid groups and nonprofit organizations.

The secondary independent variables are the race and class status of transgender individuals. Race will be measured by the subjects' self identification with a set of predetermined racial classifications. Class will be measured by income level.

Methodology

To test these hypotheses, I will be conducting a survey among transgender populations. This survey will have four components. The first will ask respondents for demographic information including race, age, and income level. The second component will ask what types of community resources the individual has received, focusing on housing, gender-affirming healthcare, and legal assistance with the name change and gender marker change process. The third component will focus on the respondent's interactions with the police. It will ask for the number of police interactions they have had within the past five years, and how many of those interactions were positive, negative, or neutral. It will also ask for a description of these interactions, in which the respondent will be given the ability to write in-depth responses where they can retell and narrate their personal experiences. Lastly, the survey will ask questions about the respondent's attitudes and emotions toward the police using a Likert scale. This scale will pose statements such as, "The police are here to protect me," "I call the police when I am in danger," and, "I feel unsafe around police," and will ask respondents whether they strongly disagree, disagree, feel neutral towards, agree, or strongly agree with the statement. An equal number of these questions will be phrased with a positive connotation (such as, "The police are here to protect me") as with a negative connotation (such as, "I feel unsafe around police") in order to minimize any unintentional bias the survey may convey to the subjects.

My biggest concern about my methodology is my ability to distribute this survey to a representative sample of transgender respondents that is both large and inclusive of individuals from a variety of racial backgrounds and income levels. This may present a challenge due to the fact that, due to redlining, many geographical communities are racially segregated. Additionally, most geographical communities

are made up of members who can afford to live in a particular location, and thus tend to be of similar income levels. As a result, obtaining a racially and economically varied sample population might be difficult. (For example, if I were to focus on a transgender community in Michigan, the city I choose to focus on will likely consist of one dominant racial and class group.) Thus, I will likely need to find a way to distribute my survey within several different geographic locations in order to obtain a varied sample that represents members of many different income levels and racial groups.

Existing literature offers solutions to mitigate this challenge; Grant et al. (2011) sent an online survey to "more than 800 transgender-led or transgender-serving community based organizations in the U.S." (Grant et al. 12). They also sent 2,000 paper copies of the survey to "organizations serving hard-to-reach populations--including rural, homeless, and low-income transgender and gender non-conforming people" (Grant et al. 12). By distributing the survey widely to a large variety of geographical locations and community organizations, researchers obtained an extremely diverse group of subjects. Their respondents were fairly equally distributed between annual income levels ranging from "under $10K" to "$100K+" (Grant et al. 3). Furthermore, their sample also represented individuals from "American Indian, Asian, Black, Latino/a , White, [and] Multiracial" racial groups in essentially equal numbers (Grant et al. 4). Likewise sending my survey to a large number of transgender community groups across the country can potentially help me obtain a large and representative sample. The first step to this is establishing relationships with organizations that serve transgender populations.

I plan to use snowball sampling to increase my sample size. Snowball samples are described by Luker (1984) as the practice of recruiting additional research subjects by asking existing subjects to identify potential new participants among people they know (Luker 1984, 251). In my study, I plan to ask respondents to my survey to circulate the survey among their transgender friends, acquaintances, and community members. In this way, my survey will ideally accrue more respondents in a relatively short period of time. Because transgender individuals marginalized by society tend to turn to one another for community and support (Arani 2020, 654), it is my hope

that snowball sampling will be particularly effective in my research as respondents will likely know other transgender community members who might participate. Luker acknowledges that one risk of snowball sampling is that "there is always the question as to how representative that selection is" (Luker 1984, 253). However, she felt confident that her own snowball sample was representative of her population of study, by nature of the fact that she took "different paths into the pool of [subjects]" (Luker 1984, 253). By sending my own survey to a wide variety of geographic locations with diverse demographics (discussed in greater detail above), I too hope to take "different paths" into my subject pool and thus gather a representative snowball sample.

Analysis

I will use data about the respondents' race, income level, types of non-governmental aid received, and number of unsolicited interactions with police to perform three tests of statistical significance. The first will test whether there is a statistically significant relationship between the respondents' race and their level of interaction with the police. The second will test whether there is a statistically significant relationship between the respondents' income level and their level of interaction with the police; the third will test for a relationship between non-governmental aid received and level of police interaction.

Next, I will examine respondents' in-depth descriptions of their interactions with police to analyze the intensity of these interactions and the use of force experienced by transgender individuals in their encounters with law enforcement. I will code survey responses in order to identify and compare rates of police harassment and brutality; I will also code responses for instances of positive interactions with police. Specifically, when coding for police harassment and brutality I will be looking for mention of the following experiences within survey responses: misgendering of transgender respondents by the police; use of force such as slapping, kicking, punching, or hitting; destruction of gender-affirming items such as breast forms and wigs; groping or other unwanted sexual touch; solicitation of sex in exchange for release; sexual assault or rape; and incarceration in a cell intended for the sex that the individual was assigned at birth. When coding for positive interactions with police I will be looking for

mention of the following experiences within survey responses: correct gendering of transgender respondents by the police; respectful demeanor or tone of voice from the police; action taken by the police on behalf of transgender respondents (for instance, helping a transgender victim of crime fill out a police report); assistance from police obtaining resources (such as police recommending a homeless shelter or food bank to transgender respondents); and incarceration in a cell intended for the sex that matches the individual's gender identity. (Admittedly, incarceration is rarely if ever a "positive experience." Nonetheless, comparing data between those who were incarcerated with the gender they identify as and those who were not remains an important distinction.) After coding the data from these descriptive responses, I will be able to determine whether a transgender individual's race, income level, or access to non-governmental aid correlates to higher levels of physical, verbal, or sexual assault by police, as well as whether they correlate to higher levels of respectful treatment by police.

I will then compare the number of self-identified positive, negative, and neutral interactions with the police, to determine whether transgender respondents have an overall positive, negative, or neutral *experience* with the police. This will be accompanied by an analysis of the Likert scale questions to determine whether transgender respondents have an overall positive, negative, or neutral *opinion* about the police. This part of the analysis will provide evidence for or against hypotheses H_{a1}, H_{a2}, H_{a3}, and H_{a4}. The foundation of my primary hypothesis rests upon the assumption that interaction with the police generally has a negative impact on transgender populations and that reducing those interactions is thus a worthwhile goal. However, if this part of the data analysis finds that H_{a2}, H_{a3}, or H_{a4} are supported, then that premise will be challenged. If that is the case, then carceral feminism may provide a more effective framework to problems faced by the transgender community.

Conclusion

My next steps for this project include reaching out to mutual aid and nonprofit organizations that serve transgender individuals, ideally prioritizing those which focus on providing the three main types of nongovernmental aid my project is focusing on. I plan to contact staff members at these organizations, introduce myself

and my research project, and ask them to offer the survey to their transgender clients. I also plan to format the survey so that it can easily be sent to others in order to aid my snowball sampling. In order to embark upon this first step I will need to compile a list of relevant organizations and their contact information.

I will then need to construct the survey. I plan to work with my thesis advisor to word the questions in a manner that will convey the least bias possible in order to avoid priming my respondents to answer in a particular way. I will also try to format the survey so that it can be distributed both online and via physical print copies, so that I can reach potential subjects who lack internet access. To accomplish this, I will need to figure out how I will collect paper copies of the survey once they have been completed by the subjects.

While much more work remains to be done on this research project, I am convinced that there is a need for research that focuses on how to most effectively limit transgender individuals' contact with the police, as a means of protecting the safety and quality of life of trans communities. This conviction stems from the convincing arguments of existing abolitionist research which demonstrates that policing enacts systemic harm towards marginalized groups, as well as the statistical research conducted by trans rights groups which shows that police target, harass, and use force against transgender populations as a means of exercising control and domination over them. Because there is a strong theoretical basis for the radical, anti-hierarchical, and anti-carceral nature of mutual aid, I am excited to focus on networks of community-provided resources within transgender communities. Existing research on the strength of mutual aid as a means of decreasing interactions between individuals and the police is mainly theoretical; therefore, I aim to fill a gap in the existing scholarly conversation by providing numerical data to support this claim. I believe that this research will lead to important conclusions about the interaction between transgender communities and police, and how transgender individuals can best be protected.

Works Cited

1. "ACTION ALERT: Tell Senate Republicans to Stop Delaying VAWA – Pass it NOW!" *now.org*, National Association for Women, 2021, now.org/ action-alert-tell-senate-republicans-to-stop-delaying-vawa-pass-it-now/.

2. Akbar, Anna A. "An Abolitionist Horizon for (Police) Reform." *California Law Review*, vol. 108, no. 6, Dec. 2020, pp. 1781-846.

3. Arani, Alexia. "Mutual Aid and Its Ambivalences: Lessons from Sick and Disabled Trans and Queer People of Color." *Feminist Studies*, vol. 46 no. 3, 2020, pp. 653-662, DOI: https://doi.org/10.15767/feministstudies.46.3.0653.

4. Bernstein, Elizabeth. "The sexual politics of the "new abolitionism"." *differences*, vol. 18 no.3, 2007, pp. 128-151, DOI: https://glc.yale.edu/sites/ default/files/pdf/sexual_politics_of_new_abolitionism_.pdf

5. Daum, Courtenay W. "The war on solicitation and intersectional subjection: quality-of-life policing as a tool to control transgender populations." *New Political Science*, vol. 37 no. 4, 2015, pp. 562-581, DOI: 10.1080/07393148.2015.1089030.

6. Foucault, Michel. *Discipline and Punish: The Birth of the Prison*. New York City, Vintage Books, 1995.

7. Davis, Angela Y. *Are Prisons Obsolete?* New York City, Seven Stories Press, 2003.

8. Grant, James M., Mottet, Lisa A., Tanis, Justin, Harrison, Jack, Herman, Jody L., and Keisling, Mara. "Injustice at Every Turn: A Report of the National Transgender Discrimination Survey." *National Center for Transgender Equality and National Gay and Lesbian Task Force*, 2011, pp. 1-222, DOI: http://www.thetaskforce.org/static_html/downloads/reports/reports/ ntds_full.pdf.

9. Hereth, Jane, and Alida Bouris. "Queering Smart Decarceration: Centering the Experiences of LGBTQ+ Young People to Imagine a World Without Prisons." *Affilia*, vol. 35, no. 3, Aug. 2020, pp. 358–375, doi:10.1177/0886109919871268.

10. Hirschel, J. David. *Domestic violence cases: What research shows about arrest and dual arrest rates*. US Department of Justice, Office of Justice Programs, National Institute of Justice, 2008.

11. "History of VAWA." *legalmomentum.org*, Legal Momentum, www.legalmomentum.org/history-vawa.

12. Hwang, Ren-yo. "Accounting for Carceral Reformations: Gay and Transgender Jailing in Los Angeles as Justice Impossible." *Critical Ethnic Studies*, vol. 2, no. 2, University of Minnesota Press, 2016, pp. 82–103, https://doi.org/10.5749/jcritethnstud.2.2.0082.

13. Hwang, Ren-yo, "Deviant Care for Deviant Futures: QTBIPoC Radical Relationalism as Mutual Aid against Carceral Care." *Transgender Studies Quarterly* 1, vol. 6 no. 4, November 2019, pp. 559–578. doi: https://doi-org.proxy.lib.umich.edu/10.1215/23289252-7771723.

14. Law, Victoria. "Against Carceral Feminism." *Jacobin Magazine*, 17 Oct. 2014, www.jacobinmag.com/2014/10/against-carceral-feminism/.

15. Luker, Kristin. *Abortion and the Politics of Motherhood*. Berkeley, University of California Press, 1985. Print.

16. McLemore, Megan, et al. "Sex Workers at Risk: Condoms as Evidence of Prostitution in Four US Cities." *Human Rights Watch*, 2012, pp. 1-112, DOI: https://www.hrw.org/report/2012/07/19/sex-workers-risk/condoms-evidence-prostitution-four-us-cities#.

17. McLeod, Allegra M. "ENVISIONING ABOLITION DEMOCRACY."
    Harvard Law Review, vol. 132, no. 6, Apr. 2019, pp. 1613-1649,
    DOI: link.gale.com/apps/doc/A581621266/AONE?u=umuser&sid=
    bookmark-AONE&xid=3c6d1725. Accessed 13 Oct. 2021.

18. Nadal, Kevin L., Davidoff, Kristin C., Fujii-Doe, Whitney."Transgender Women
    and the Sex Work Industry: Roots in Systemic, Institutional, and
    Interpersonal Discrimination." *Journal of Trauma & Dissociation*, vol. 15
    no. 2, 2014, pp. 169-183, DOI: 10.1080/15299732.2014.867572.

19. "1994 Violent Crime Control and Law Enforcement Act ." *ojp.gov*, US
    Department of Justice, 14 Feb. 2020, www.ojp.gov/ojp50/1994-violent-
    crime-control-and-law-enforcement-act.

20. Roberts, Dorothy E. "Democratizing Criminal Law as an Abolitionist Project."
    Northwestern University Law Review, vol. 11 no. 6, August 2017,
    pp. 1597-1608, DOI: https://scholarlycommons.law.northwestern.
    edu/nulr/vol111/iss6/11

21. Russell, Emma K. "Carceral Pride: The Fusion of Police Imagery with LGBTI
    Rights." *Feminist Legal Studies*, vol. 26, 22 August 2018, pp. 331-350,
    DOI: https://doi.org/10.1007/s10691-018-9383-2.

22. Spade, Dean. "Solidarity Not Charity: Mutual Aid for Mobilization and
    Survival." *Social Text*, vol. 38 no. 1, 1 March 2020, pp. 131–151,
    doi: https://doi.org/10.1215/01642472-7971139.

23. Sumner, Jennifer, and Sexton, Lori. "Same Difference: The 'Dilemma of
    Difference' and the Incarceration of Transgender Prisoners." *Law &
    Social Inquiry*, vol. 41, no. 3, 2016, pp. 616–642., doi:10.1111/lsi.12193.

24. "Transgressive Policing: Police Abuse of LGBTQ Communities of Color in Jackson Heights." *Make the Road New York*, October 2012, pp. 1-33, DOI: http://www.maketheroad.org/pix_reports/MRNY_ Transgressive_Policing_Full_Report_10.23.12B.pdf

25. Whalley, Elizabeth and Hackett, Colleen. "Carceral feminisms: the abolitionist project and undoing dominant feminisms." *Contemporary Justice Review*, vol. 20 no. 4, 2017, pp. 456-473, DOI: 10.1080/10282580.2017.1383762.

26. Zamani, Nahal, Bastedo, Kip; Krumholz, Meghan; Platt, Ashley; Tasaki, Aya; Williams, An-Tuan; Bollag, Sascha; Flanagan, Whitney; Jordan, Elizabeth; Kondo, Alex; Wohlschlegel, Kari; Ritchie, Andrea; and Hogarth, Sarah. "Stop and Frisk: The Human Impact." *Center for Constitutional Rights*, July 2012, pp. 1-31, DOI: https://ccrjustice. org/sites/default/files/attach/2015/08/the-human-impact-report.pdf.

27. Zorza, Joan, and Laurie Woods. *Mandatory arrest: Problems and possibilities*. National Battered Women's Law Project, National Center on Women and Family Law, 1994.

Excellence in Upper-Level Writing (Humanities)

The Taste of Words Once Forgotten: Multi-Hyphenate Mediterraneanism and the Universal Language of the Kitchen

by Gabriel Consiglio

From ARTDES 399: Writing in Art and Design
Nominated by Jennifer Metsker

Writing about creative practice is not an easy task. To capture the act of making in such a way that a reader can not only understand its relevance but also experience it for themselves requires an inventive use of language and the ability to connect a non-verbal act to a complex historical context. Gabriel's essay, The Taste of Words Once Forgotten: Multi-Hyphenate Mediterraneanism and the Universal Language of the Kitchen, does this and more as it reflects upon his ceramic work through the lens of his family and culture, taking us on a journey from the food that inspires him to the ceramics studio and beyond, diving deeply into research on the ways identity has been explored through art. The questions he asks are genuine and uncover the challenges of attempting to better understand one's identity through art, revealing possible ways that he can move forward in his practice as well as ways we might think differently about art. And he does this in exquisitely written prose that allows the experience of art viewing to be made palpable on the page.

-- *Jennifer Metsker*

The Taste of Words Once Forgotten: Multi-Hyphenate Mediterraneanism and the Universal Language of the Kitchen

When my grandfather was coming of age in the suburbs of Syracuse, New York, he had already worked hard to take the Arab out of himself. His name is Richard, his father's Ed and his uncle's Ned, but it wasn't until my early adulthood that I learned Ed was short for Edim and Ned for Nedim. Their last name, George, which my mother has kept in her multi-hyphenate married name, was changed at Ellis Island from Massoud; it was a stroke of luck that they would only have to hide their first names and not their last. My *jidoo*'s parents had, in order to make life easier for their children, stopped speaking Arabic in their home, hoping that the traces of exoticism still left on their tongues would be distilled away as the children grew up like other American children. This was especially hard for my *jidoo*, by no fault of his own; he was dark-skinned even for an Arab, and in the predominately white landscape of upstate New York he stuck out like a sore thumb. Even so, he forgot the words that were never passed down to him, and by the time he married my grandmother, an English-Irish-French-and-Dutch New Yorker, at the age of 19, he was a bona fide American. My *jidoo* lived out an incarnation of the American Dream: he opened his own screen-printing business, had four children in short succession, and went to night school simultaneously to earn a degree from Syracuse University after the better part of a decade. He worked tirelessly, with the same aim that his parents had—to make his children's lives as easy as possible—and though he had succeeded in a fiscal manner, he had unknowingly been relieved of the sociocultural burden of his Arabness by his parents' efforts to assimilate. My grandmother, however, inherited this identity somewhere along the way. Today, her seven grandchildren call her *sitoo*, the Arabic word for grandmother, even though she has not a drop of Middle Eastern blood in her ancestry. She learned the recipes of her husband's family as though they were her own, and her hands that had previously made tomato soup pork chops began to wrap *dolma* and hollow out *cousa*. In a way, my *sitoo* caused the generations-long assimilation effort

to combust furiously, years after its originators had been laid to rest. She filled her home with the self-Orientalist decoration that had come to represent my family's culture in the forced absence of any actual artifacts, with camels and drums and ornate vases in every nook and cranny. She passed down the recipes that she had acquired to each of her four children, and she held annual Thanksgivings and family reunions where she would make *kibbeh nayeh* alongside Butterball turkeys. I've grown up watching my mother and her mother cook food that was forgotten for generations, and I can't remember a time in my life where I was not certain that I was an Arab. I cook the same food on my own now, but even as the memories of recipes and cooking methods become written into the muscles of my hands, only one word, outside the names of our dishes, has come back to me: *khara*—shit! A dull exclamation of disappointment, of remorse at the language and the history lost at the hands of my ancestors, who had no other choice but to forget them.

My father is a purebred Sicilian and a first-generation American, with a quaint apartment on the top floor of the tallest building in a seaside town just outside of Palermo. The building, decades ago, was erected with the intent of housing each member of his father's family, though he is now the only descendant that still owns their piece. His parents both passed through Ellis Island separately, and they were allowed to keep their names—all five of them; in my grandmother's case, a multi-hyphenate like my mother. They worked long and hard hours, automotive factory jobs having lured them to Detroit, so my father grew up a bona fide American, forgetting his language without ever intending to. My father is a purebred Sicilian, and my mother is a third-generation English-Irish-French-and-Dutch-and-Lebanese New Yorker. She has no land in Beirut and has never been to Lebanon—neither has her father nor, I suspect, her grandfather. Yet Lebanon feels closer to me than Sicily, despite the Terrasini summers of my childhood and the thick immigrant's accent that clouded the conversations I had with my grandmother. I've long been puzzled by my relationship with my family's history; I have had coffee and *gelato* in the same *piazza* that my grandmother once did almost a century ago, yet I feel closer to a culture that my family has not known firsthand for many years. If I am more of a Sicilian, both by proximity and genealogy, then what does it mean to be an Arab?

I search for the answer to this question, and for inspiration, in the kitchen. It is here where different cultures intersect—even ones outside of my own—and it is here where not just foods and ingredients but objects and vessels from all over the world sit alongside each other. I think a lot about the new conception of Arabness that I create for myself here, one that my mother created for her children, her mother for hers and so on. I've drawn inextricable associations between the Sicilian majolica dishes that act as decoration and utility simultaneously and the Arabic cuisine that is served on them. I've brought sumac and allspice into the dishes I acquire from college roommates and in return I leave them with the taste for *baba ghanoush* and *kafta*; they will doctor up these dishes in their own ways, just as I bend the food that almost skipped a generation to the constraints of new diets and grocery store inventories. I am intrigued by the cultural crossroads of the kitchen, whether it is in my mother's home or my college home, and I think that it is because of this that I am so attracted to the medium of ceramics. A meal and its vessel work together to spin a sentence that traces along a winding culinary lineage; a bowl and its contents can say more about a person than an essay. It is for this reason that I find myself drawn to create utilitarian objects, and that the ceramic works of my childhood home still appear so vividly in my mind.

In my work, I find both food and language creeping their way into my imagery—one idea that I know well and one that has continued to elude me. In the packaging that my mother sends me pictures of from her pantry, I know intuitively the contents of each box. I can't read most of the words on them; some are written in a gibberish half-English aimed at customers who might not have a full grasp of the language. Others are completely in Arabic, imported goods that make their way into the Iraqi grocers of Dearborn and then into my mother's kitchen. I have her send me packaging indiscriminately—my original intention is to transfer the labels of her kitchen staples onto the surface of my ceramic vessel, but most of these staples happen to be Arabic ingredients, and a theme begins to arise without me knowing it. I find myself taking on an interesting quality of line and brushwork when depicting these labels that I don't find myself doing for any other subject. I'm writing words that I don't know, letters that my fingers have never traced before, and because of this I paint with a painstaking slowness. I often find myself making marks quickly, as I like the swift, unwavering quality of a line that is drawn with a succinct and confident hand. However, on this vessel, I embrace how my hand quivers as I move along its curved surface, trying to capture Arabic words as geometric forms rather than something meant to be read. In this process, I realize that I am in a way working backwards—the hand that designed these food labels wrote the words from right to left, while I trace along them from left to right. There is a strange connection with the language that I begin to feel here, one not based on comprehension but on pure beauty of form and line. The deep blues of the underglaze that I paint in sit in stark contrast against the white glaze on the surface of my vessel, and when they are fired this will only grow deeper. Though I don't feel that this cobalt hue is sullen or depressive, it serves to mute and refine the loudness of the original food labels that I use as my source material, which scream in the colors of the flags of Palestine and Lebanon and Yemen so that they can be located from afar by a person who needs them. These labels are taken from different corners of the world and shelves of the pantry and put into the same color palette, all contained within a horizontal band across the belly of my vessel. For me, the text on these labels no longer has any literal meaning; it is purely compositional, a matter of mark-making and patterning against the white sheen of the glaze.

My grandfather unwittingly lost the language of his ancestors, and I think that, in working through my family's history through food, this fact influenced the way that I looked at text in this piece. It feels strange to look at a script that is entirely incomprehensible to me and know that, just a few generations ago, my family was reading and writing almost exclusively in it. I think that immersing myself in this script, moving slowly along the strokes of each letter and following them with my own hand, is my own idea of how it might have felt as my grandfather, being reintroduced to aspects of his culture much later in his life, after growing up somewhat removed from them. I think that this cultural loss, a result of attempted assimilation that has effectively cut most members of my family off from a language that their forefathers once knew fluently, is the reason that I often fixate on Arabic script rather than other recognizably Arabic pieces of imagery. My family and I have been able to reclaim a lot of other aspects of Arab-American culture, our food being the most significant but also our music, dancing, and holidays. I know these things well, and it is because of this that I move toward language and text, as I situate myself in the one aspect of my culture that I might never be able to reclaim. I dramatize here, knowing that I could simply enroll in an Arabic language class and attempt to remedy this myself. However, I think that there is certainly a subconscious influence that this idea has on me that pushes me to embed it into my artwork.

The process of cooking itself has also been informative to my art-making practice, though this was a realization I made in the inverse of my revelations about language. I observed my growing fascination with Arabic script while I was actively decorating the piece shown above, but I began to see the contrasts between my creative and culinary processes only when I had left the studio. It is certainly no revelation that food, from preparation to plating to flavor profiles, is an art form, but I have found that the way I operate in the kitchen is starkly different from how I do in the studio. To me, food is an art-making space where I become impulsive and intuitive, making split-second decisions and working within strict constraints; cooking requires sensitivity, as too much heat or too much time spent can easily ruin a dish. However, I never abandon a dish, even though failure is a regular occurrence. Working with

clay demands attention to a lot of the same elements that are important in cooking: a ceramicist must have an intuitive understanding of a clay body's moisture, the time they spend working with it, and even the pressure that they use to manipulate it. Given these similarities, the differences in my creative practices become more evident the more I cook. I relish the challenges that I am presented with in a culinary mode of working, yet when my hands are in clay, I often find myself paralyzed, unable to experiment and driven toward a single, unwavering vision of what I aim to create. I become frustrated with myself when a vessel starts to take on an imperfect shape, even though I eat every dish I cook, no matter much I over-season it. I wonder how my outlook on my art-making practice might change if I think of my creative projects in the ways that I think about my meals: everyday rituals that carry the thread of my personal narrative from one day to the next, ever-changing and evolving. What stories might be waiting to be unearthed if I look at the studio in the same way I do the kitchen?

Food, to me, plays a similar role to that of a language. As I strip the meaning from each of the characters that I paint onto my vessel, I look to other means to tell its story. Not only in this piece but in ceramic and culinary practice alike, I've found that there are ways other than words with which a narrative can be woven. The form of a ceramic object instructs a person how to hold and use it, and the texture of its surface, even when devoid of ornamentation, serves to remind them of its maker's touch. The label on a can of chickpeas evokes cultural and historical flashbacks, and the decorative smear into the surface of a freshly made batch of *hummus*—the grooves that will carry olive oil and paprika—brings a family's forgotten history a little bit closer to home. In my family, food is the resting place of cultural memory. When we cook and when we eat, we are reminded of a past that was nearly lost to us, as well as the new futures that we write every day. And though the alphabet of our ancestors has become nearly extinct in our English-Irish-French-and-Dutch-and-Lebanese-Sicilian-American vernacular, we can come to understand them more as we eat.

Power-Building in the Void of Stolen Histories

bell hooks writes that, a half-century after the abolition of American slavery, "elderly black people developed a cultural passion for the camera, for the images it produced, because it offered a way to contain memories, to overcome loss, to keep history" (60). Black Americans have had their cultural histories forcefully erased for hundreds of years, beginning with the enslavement of Africans—and the obfuscation of their origins and cultural identities as they were shipped across the Atlantic—and continuing through segregation and apartheid to the present day, where the hegemonic institutions and modes of visual representation that hooks describes in *Art on My Mind* exclude black images that "would subvert the status quo" (hooks 58). When black people were denied access to museums and gallery spaces, whether explicitly or not, "the walls of images in Southern black homes [became] sites of resistance" (hooks 59). Though hooks in her writing focuses on the role of photography as a "critical intervention, a disruption of white control over black images" (59), I take a greater interest in the way that these images functioned in the privacy of the home: photograph-covered walls serving as "private, black-owned and -operated gallery [spaces]" (59), curated by grandmothers and grandfathers. It is on these walls that hooks shows history can be preserved, and simultaneously rewritten.

hooks outlines a dichotomy here, a fundamental difference between a photo album and a photo wall. A photo album is closed away, accessible "only by request" (hooks 61), while the walls, in this case the walls of her grandmother, "were a public announcement of the primacy of the image, the joy of image making" (61); placement and sequence mattered here, and the movement through a space dotted with images imparts a narrative to the domestic gallery-goer. Though the contrast that hooks calls attention to is along a private-public axis, with photo books hidden away and walls on constant exhibition, both these modes of documentation exist within the private domain of the home. Sarah Oldham's photo walls retain not only a cultural history but a specific familial one due to their place in the home, and even though her images shout out to anybody who passes through their hallways, this publicity exists only within the intimacy of their setting.

I hesitate to draw comparisons between hooks' black American experiences and those of my Arab immigrant ancestors just yet; I feel the need to articulate the stark differences in the way that these two groups were able to navigate the American landscape and the privileges that were afforded to my family's ethnic group, categorizations that they intentionally pursued. However, as to not let my reader grow too hungry, I will dig into one parallel before I do so. As I paged through bell hooks' *Art on My Mind* for a second time—the first before I had started this creative-cultural-culinary exploration—I found the nature of the photo wall of the black American home to be strikingly similar to the way in which I view food. The private foodways of a household, like the galleries that unfurl across the walls of Sarah Oldman and many other black Americans, are a public exhibition within a private, intimate context. Home cooking shouts loudly, giving guests an implicit lesson in cultural history in the same way that gazing upon a curated arrangement of photographs would; outside of the home, however, it becomes difficult for ethnic food to exist uncompromised by the dominant white structures that govern restaurants and grocery stores. While Arabs from diverse and vibrant cultural backgrounds often fall into patterns of palatable Levantine menus and self-Orientalizing imagery when they open their eateries, the meals that they eat at home are often quite different. White galleries and museums make it difficult for black Americans to produce and exhibit work that falls outside of a hegemonic imagination of what blackness might look like, so we see the "construction of an oppositional black aesthetic" (hooks 57) take place in the home. Similarly, Arabness—along with virtually all other ethnic expression—is diluted for consumption by a largely white public; it is in the home, in the kitchen, where culture is preserved and stereotypes are challenged.

"In 1952, Malcolm Little recognized the history of anti-Black violence embedded in his surname…[and] dropped the name of his ancestors' slave owner in exchange for an 'X.' The letter represented his ignorance of his family's 'Original name,'" a memory that had been stolen from him generations before his own birth (Onaci). Agency in one's own naming is an idea that Malcolm X engaged with both before and after his time in prison, when he took on the moniker that he is remembered for today. In "Black Power, Name Choices, and Self-Determination," Edward Onaci

writes of the names that Malcolm X assumed at the bequest of others, and the ones that he chose for himself. Nicknames, ones he earned during his time in Harlem and Boston like "Detroit Red," and racial slurs, hurled at him by "white teachers and classmates," were the first instances of names other than the one given to him (Onaci), but Malcolm's membership in the Nation of Islam led him to choose names that would evoke his cultural history, rather than reflect the ideas of those around him. Before he was murdered, Malcolm "took on the name El-Hajj Malik El-Shabazz" (Onaci), inspired by the African and Arabic origins of his culture and religion that had been denied to his forefathers in America. Malcolm X, and those who continued his work in the Black Power movement, saw "self-definition through naming as an important step in a multi-faceted Black revolution" (Onaci).

My grandfather's grandfather was given a new name upon his arrival in America. This name was one he did not choose; the name George was the first denial of Arabness that Mershed Massoud experienced in the country that was to bring his descendants splendor. Though this first small mark of cultural erasure was indeed forced upon him, the newly named Mershed George followed a path of assimilation that would lead him to an easier life in America, the only adverse symptom being a cultural amnesia that would be inherited by his children. While Malcolm X found power in reclaiming a name that he had never been known by, my great-great-grandfather found power in forgetting his; perceived whiteness, even if only on paper at first, helped Mershed George to attain quotidian success, a gradual assimilation into American whiteness that privileges me to examine my own histories today.

The ethnic confusion and diasporic disconnection that I've attempted to explore within my own creative practice is not unique to me—I am not the only Arab with the taste of his homeland lingering on his tongue, let alone the only descendent of a diaspora that has melted into the fondue of American culture. While my forefathers sacrificed the cultural identities of their children in exchange for a more equal-footed life—a successful seizure of whiteness by both my Sicilian and Lebanese ancestors—other ethnic and racial groups have fallen victim to a forced erasure of cultural memory. African Americans, as hooks has begun to show us, have perhaps

suffered the gravest and most total cultural detachment of the racial groups that make up America's contemporary palette. The consequences of centuries of capture and enslavement are irreparable; the African American identity has formed in a manner incomparable to the Arab American or Italian American, who willingly relinquished their ethnic identities as they assimilated into various shades of whiteness. No matter how far removed a person like me becomes from the culture of their ancestors, we cannot equate their experiences with those of African Americans, who oftentimes find it impossible to determine even the area in Africa from which their families descended, much less their specific lineages. While my grandfather lost his name at Ellis Island, many African Americans have not only lost theirs but bear the names of men that enslaved their very ancestors. Native Americans have suffered a similarly damaging cultural loss through boarding schools, forced reeducation, and genocide. It would be remiss to write about the detriments of diaspora without acknowledging the fact that my own was made in comparative privilege.

Though I provide my own examples of cultural memory loss to situate my practice and experiences in the broader context of America's ethnic makeup, I do so also to set the stage for the artists, scholars, and chefs that have found themselves grappling with the same issues of identity that I do. It is my goal to better understand the ways in which cultural agents across different disciplines engage with the themes of culture and ethnicity; what might I learn from the Palestinian cookbook author or the black chef from the American South, and how can I take this into my own artmaking practice? Are my fixations on Arabic script an example of self-stereotyping in an attempt to reconnect with a culture and community I don't truly know? Or are they unconscious whispers of the way that I perform my culture in my everyday life? Can the mechanisms used to disadvantage historically marginalized groups tell me anything about my own experiences, or does my earned whiteness mitigate their significance? I think that examining different perspectives of ethnicity and diaspora, from outside my own medium and from cultural groups other than my own, will help me to ground myself as I explore my own identity and to think about my artwork moving forward in ways that I haven't before.

Digital Diaspora and Disidentification

The Arab diaspora is a unique one because of the wide geographical scope that its origins encompass, let alone the number of locations that its emigrants have settled in. Like the label of 'African,' the Arab diaspora includes people from different countries, regions, and even continents: by some definitions, members of the Arab diaspora can trace their origins not only to Southwest Asia and the Arabian Peninsula but to North African nations like Morocco and Somalia. As a result, the racial composition of diasporic communities differs greatly. Some Arabs, like myself, have been afforded the privileges of whiteness that were earned by the Italian and Eastern European immigrants that arrived in settler nations like the United States alongside them; others are black, met with not only the structural disadvantages and discrimination that targeted African Americans but the xenophobia directed toward any non-white immigrant group. Still, other diaspora Arabs fall in between the cracks, too brown-skinned to benefit from the arbitrary demarcations of whiteness but not subject to the same level of prejudice as African Americans and black immigrants. Despite the racial and ethnic differences that mold the way in which diaspora communities develop all over the world, the broad 'Arab' label still serves to connect people across lines of national and continental origin. Róisín Tapponi writes that "the 'Arab region' is as decentralized geographically as it is institutionally," and it is for this reason that many contemporary Arab artists have taken to "social-media platforms and new digital mediums" to express themselves and build community (Tapponi). Though an affinity for Internet-based art is certainly not unique to Arab visual culture, the scattered distribution of origins and resting places within the group has likely played a role in contemporary Arab artists' taste for the digital. After all, "the place we call 'home' is rarely limited to one geography or nation," and we "can travel the internet more readily than we can cross borders" (Tapponi). Digital art collectives have emerged from the Arab diaspora along the axes of national origin, race, and gender: Edge of Arabia serves to connect and foster dialogue between Arab nations, specifically Saudi Arabia, and their ethnic communities in Western Nations, and Tapponi's own Habibi Collective curates film and "moving-image work by women

and non-binary folk" (Tapponi). The exhibition and creation of digital artwork does more than simply connect a fragmented global diaspora, however significant that effect is. I know well from my family's own history the cultural loss that is often experienced as immigrants transliterate their cultures and traditions to adapt to a new, usually Western and white, context. Assimilation is often a mechanism of survival, where first generation migrants sacrifice cultural vibrancy in exchange for better futures for their children. And while individual family histories can be muddled or lost entirely, this is also the case for broader narratives within diasporic communities; memories of cultural developments and social movements can be forgotten, as second- and third-generation descendants of immigrants long for the imagined antiquity of their homelands. With Habibi Collective, Tapponi attempts to speculate an "alternative cultural history" through a concept that she calls "becoming" (Tapponi). She writes that the "sharing and reassessment of…fragmented archives of women's liberation" can work to make Arab histories more visible, as the "digital connections marshalled by the collective" have placed work by Arab women and non-binary folks at "international film screenings and festivals" (Tapponi). I have engaged in my own kind of 'becoming' in my creative practice and in this essay, working to retrace the lineages of my family and to imagine what their experiences might have been like through my artwork. Though Tapponi uses the word to refer to the construction of alternative histories and memories, I wonder how the idea of 'becoming' might be used to project new futures? Can my artwork be a means of 'becoming' Arab again rather than just dissecting and examining that identity?

I promise to return to Arabness, and of course to the kitchen, before I close out this essay, but I will first take a final detour through the art world and its criticisms, here by Native American voices. "Native Perspectives" is a collection of responses by "contemporary Native artists and historians" to works in the American Wing of the Met Museum, which were created by European and American artists in the 18th and 19th centuries (Baker et al.). Many of the works in this collection engage directly with Native American subjects, who are often manipulated to serve the aims of the American artists that rendered—or entirely fabricated, in some cases—them. In response to

"Hiawatha," an 1874 statue by Augustus Saint-Gaudens, Jackson Polys writes that "to trace the arc of the hunched back of this imagined Indian is to follow the call for the end of the wild" (Baker et al.). Saint-Gaudens' hand depicts this Native American character with the Eurocentric flavor of the marble statues of Greece and Rome; not only is this character white, save for some facial features and scant traditional dress, but he is at rest, sitting resigned and peacefully alongside his quiver. In Polys's eyes, this depiction serves to bolster the idea of a "new national tribe," one that "relies on the neutering of the Indian problem" (Baker et al.). It is necessary to make the Native American white to make room for the settler American, "aided by the work of artists, [to become] Native" (Baker et al.), an idea reflected in art as well as in the real-world policies of Native boarding schools. Polys makes a somber conclusion that the work of Saint-Gaudens helps to "justify the foundational violence of wars unfortunately won" and shows how "Natives are incorporated into the national narrative" (Baker et al.).

Augustus Saint-Gaudens (1848–1907). "Hiawatha," 1871–72, carved 1874.

Through "Hiawatha" and many of the other artworks critiqued by Native American contemporaries, we can see how the dominant structures of the art world have and continue to work genocidally, propping up policy and structure that effectively erases certain racial groups while tokenizing their images in favor of colonial discourses. It seems almost futile to make a play at resistance and rebellion against white dominance if it must make use of these existing structures. While we aren't the same segregated society that bell hooks recalls, where museums and galleries refuse to display work by black artists, the memories of this systematic exclusion still linger in the unchanging conventions of the art world. Though museums now make space for artists of color to exhibit within their walls, what does this matter if they are juxtaposed against stolen artifacts and appropriated artworks? Even in the digital world that Tapponi and her contemporaries inhabit, we must point out "the paradox of using platforms that sensationalize on POC, female-identifying and queer bodies as a means of advancing urgent political or cultural dialogue" (Russel). There is certainly something to be said for using the art world against itself, for working within existing frameworks to take what works and dismantle the rest; disidentification, described by José Esteban Muñoz as "a strategy that tries to transform a cultural logic from within, always laboring to enact permanent structural change while at the same time valuing the importance of local or everyday struggles of resistance" (Muñoz 11-12), lets symbols and structures be appropriated into cultures that have been historically disadvantaged by them. We have seen this idea at work in the same forum where Jackson Polys examined "Hiawatha." Jeffrey Gibson recalls a bronze cast by James Earle Fraser entitled "End of the Trail", "an image of a shamed, defeated Indian" that populated the Cherokee gift shop he visited as a child (Baker et al.). Like the Arabs that dilute their ethnic cuisine in order to serve a palatable menu to a white public, the reproduction of this sculpture might have operated to materially improve the lives of the Cherokee gift shop owners that sold it. However, the intentional reproduction of a harmful stereotype can lead to its reinvention and recontextualization. Gibson describes the powwows where "End of the Trail" was printed on ceremonial flags as a "symbol of resilience and strength—characteristics traditionally associated with the warrior" and far from the message likely intended by Fraser (Baker et al.).

In my own work, I grapple with lesser demons than the methodical and mechanical emasculation that Native Americans faced in the white artist's imagination. I look back now, though, with a newfound understanding of my visual affinity toward certain stereotypical aspects of Arabness. It is an unseen disidentification with the common imagination of Arabic ceramics that leads me to the porcelain whites and abyssal blues that dress my own vessels. In the earliest stages of this essay, I found myself perplexed by my attraction to this color scheme, which immediately came to mind when I imagined the works that my distant ancestors might have created but was nowhere to be found in my research. I know now that this urge—often an unconscious one—to engage with the intuitive association and the stereotype comes from a desire to reclaim them. To incorporate cultural aesthetics, real or imagined, into my work is to use them as the foundation for a cultural identity that I am continually constructing, stripped of their connotations and simmered into a roux that will become my own unique take on Arabness.

Stepping further away from the hegemony of the art world, past disidentificatory modes of making and representation, is not an entirely new frontier. In fact, this is exactly the type of image making that hooks describes in the counterhegemonic photo galleries that adorn the walls of black homes. Instead of repatriating symbols originally used for oppression, ethnic communities can create new spaces entirely outside the confines laid by white structures. In the home, we can break boundaries and write new histories ourselves. Even for me, an English-Irish-French-and-Dutch-and-Lebanese-Sicilian-American who has benefited from attained and inherited whiteness more than once over, the home can be a place for experimentation that doesn't happen in the studio or the gallery. Perhaps it is for this reason that I eat every dish I cook, no matter how strange, while I can't bring myself to continue past a collapsed clay vessel. The household and the kitchen as a gallery and studio all in one—perhaps it is in the context where I began that my creative practice will end up. I engage in the act of becoming every time I skin an eggplant, drain a can of chickpeas, sprinkle *zaatar* on a fried egg, more so than I do in the ceramic studio. This private rewriting of my own culture, a culinary intimacy shared between family and friends, seems only to naturally

exist in the home. And thus I arrive at the same questions that puzzled me at the beginning of this exploration: how can I meld the cultural experiences that I write with each meal I cook with my creative practice? Should I, if I knew how? It might be the case that the two are doomed to exist separately, always informing each other but never truly mixing—you'd never take a bite out of your tableware. It is at this point that you might expect me to shout *khara*, to step away from the keyboard with a frustrated hunger that would lead me nowhere but to the stove. That is not the case, however, and this deep dive that has ended up somewhat circular has caused me to look at both my creative practice and my cuisine in a new light. The answers that I have arrived at are nowhere near definite, and I've begun to grow an appetite for experimentation that I hope will lead me somewhere grandiose.

Works Cited

Baker, Joe, et al. "Native Perspectives." *The Met Museum*, www.metmuseum.org/
    about-the-met/collection-areas/the-american-wing/native-perspectives
    #contributors.

hooks, bell. *Art on My Mind: Visual Politics*. New Press, 1998.

Muñoz, José Esteban. *Disidentifications: Queers of Color and the Performance of
    Politics*. University of Minnesota Press, 2013.

Onaci, Edward. "Black Power, Name Choices, and Self-Determination." *Black
    Perspectives*, AAIHS, 24 Aug. 2016, www.aaihs.org/black-power-
    name-choices-and-self-determination/.

Russell, Legacy. *Glitch Feminism: a Manifesto*. Verso, 2020.

Tapponi, Róisín. "How Arab Art Collectives Are Using the Internet to Cross
    Borders." *Frieze*, 4 Dec. 2020, www.frieze.com/article/how-arab-
    art-collectives-are-using-internet-cross-borders.

Excellence in Upper-Level Writing (Humanities)

Progress in Explainable AI for Deep Neural Nets
by Anastasia Warner
From WRITING 400: Writing and Research in the Sciences
Nominated by Jimmy Brancho

Ana's review of explainable AI research embodies what science writing can do when it breaks free of its traditionalism. Written more like a TV segment or podcast episode than a quote-unquote literature review, Ana's piece shows us the unknown complexities of artificial intelligence and the stark need to firmly understand its inner workings even as we innovate on it. Ana makes clear the stakes of allowing black-box AIs - algorithms whose inner workings we can not interpret - to propagate into systems controlling hiring, distribution, or law enforcement. Ana sacrifices no scientific rigor in this sober and eminently readable review.

-- Jimmy Brancho

Progress in Explainable AI for Deep Neural Nets

ABSTRACT

The use of opaque neural network machine learning models, referred to as black-boxes, is constantly increasing as more and more fields apply artificial intelligence (AI) to difficult and time-consuming problems. Generally, the more complex the model, the better it can learn trends in data – and the harder it is for users to understand. More literature is being published detailing attempts to explain the inner workings of these black-boxes; known as Explainable AI (XAI), these systems cover a range of approaches in increasing the transparency, interpretability, and safety of opaque models. Some methods being created today are novel and unprecedented, while others are inspired by the earliest, decades-old work in the field of XAI. The literature shows that previous work is constantly improved upon to create better explanations and clearer understanding of AI systems. In acknowledgement of this trend, this review first presents an overview of a selection of foundational approaches in XAI for deep neural nets, intended to serve as an entry point for interested practitioners to learn about relevant XAI goals and common techniques. We further aim to provide a picture of how the field has grown in recent years by detailing some of the approaches' subsequent iterations, illuminating the connections embedded in the literature and how new work addresses the shortcomings of prior methods while often preserving the original design. This consolidation of literature examining the advancements of several explainability techniques should be useful for putting the research field into perspective.

I. INTRODUCTION

Deep learning models such as deep neural nets are growing in popularity as the challenges that we ask artificial intelligence to solve continue to expand in application and complexity. These machine learning algorithms are capable of training themselves to classify data by tuning the internal weights of their tens to thousands of layers of neurons. Deep learning models can be used to input a patient's health

records and receive a prediction of the likeliness they might acquire liver cancer, lung cancer, or breast cancer. Or, in a different field, to input a batch of one thousand job applicants' resumes and receive a selection of the ones your company should interview. What characteristics is the program basing its decisions on? Sometimes, even the engineers who designed it don't know for sure.

An algorithm such as this is called a black-box: the functions of the algorithm that detect meaningful features of an input dataset and determine their importance are hidden in the numerical weights inside the model, which is not inherently interpretable to humans. This problem is especially significant in the case of deep learning, where the intermediate outputs from many filtering and weighting layers interact before producing a final prediction.

Understanding why and how the model makes the decisions can help practitioners validate the model's success, discover new connections in datasets, and put more trust in the model. This is especially relevant, for example, in medical fields where doctors use AI to predict patients' risk of certain diseases or detect cancer cells.

On the other hand, the use of poorly understood models can result in harmful effects; they can be breeding grounds for unwittingly biased predictions or impractical weighting of trivial features. These problems can go undetected in the model testing and evaluation processes if the dataset, accuracy measures, or testing is somehow flawed. Gaining insight into what features are being discovered by the deep network's complex and noninterpretable weight interactions can help verify or sanity-check the model for fair and reasonable performance.

That's why many researchers and software designers are turning to the field of Explainable Artificial Intelligence, or XAI. XAI encompasses any features of artificial intelligence that produce results that can be understood by humans. Explainability can take on a number of approaches, including the design of inherently interpretable model structures, the probing of certain properties in a black-box by providing a range of related inputs, or the creation of justifications for a single prediction at a time.

In this survey we will provide an organized overview of the advancements made to a selection of technical approaches to illuminating these black-boxes in

recent years; these approaches include a method of rule extraction to explain a model's decision-making, along with two methods of explaining the influences for a single prediction. In section 2, we will discuss relevant background concepts in AI and neural networks. In section 3, we present the origin of the selected approaches and discuss some subsequent iterations of each.

II. BACKGROUND

Deep learning is a family of models for machine learning that are based on artificial neural networks with many layers. Neural nets are made of layers of matrices that apply nonlinear transformations to the input space data, converting it to the output predictions. The elements in the matrix are called neurons or nodes, and they store numeric weights that are tuned during training to better fit the data and to optimize the objective. This objective function specifies the score, called the loss, at the end of a training step and usually takes the form of minimizing the overall error in predictions. The objective can also be set to factor in other characteristics important to the designers, often with the goals of minimizing the model's complexity, avoiding certain types of incorrect predictions, and others. At each iteration of a training step, the weights are updated by a small amount in a direction that will further minimize the loss. To find the right direction and amount to update each weight, the algorithm finds the gradient of each neuron in the system with respect to the final class score or prediction. In a process called backpropagation, the gradients are calculated from the last layer all the way back through the first.

The internal complexity of deep learning models allows the many weights to be tuned to respond to extreme subtleties in input data, sometimes to a degree that humans cannot detect. This same complexity is also a problem: the influence of human-interpretable features in the model is hard and sometimes impossible to understand, as the decision-making in the model can be spread across thousands of nodes, each applying a mathematical transformation and sending the output to many other nodes. Models such as a deep neural net where the internal processes are uninterpretable to the observer are called black-boxes, a designation which is also sometimes applied to models whose overall design is unknown to the user.

Explainable AI

Explainable AI seeks to address several goals for users of AI. *Explainability* can be accomplished through using an interpretable model, or through applied explainability techniques. In machine learning, *interpretability* is defined as the ability to explain or to provide the meaning in understandable terms to a human [13] . Thus, interpretable models refer to machine learning models that are inherently human-interpretable due to their low complexity. Meanwhile, explainability refers to applied methods to support user understanding of complex models by providing explanations for their predictions. These explanations can have a number of benefits – Adadi and Berrada categorize 4 uses [14]:

1. Justification, by providing reasons for why a prediction was made, which can help improve trust in the model in critical situations.

2. Control, by providing visibility of flaws when debugging.

3. Improvement, by illuminating weak or illogical reasoning for predictions.

4. Discovery, when models pick up on trends not yet known to humans.

Interpretable Models

Interpretable models are sometimes referred to as transparent boxes, in contrast to black-boxes. Interpretable models include decision trees, linear models, and falling rule lists. A decision tree is a logic flow represented as a tree graph, depicted in Fig. 1. At each node, a test is done on an input feature comparing it to a threshold, and each leaf node assigns a predicted label. A path through the tree would pass through a collection of conditions which form the rules for a specific label.
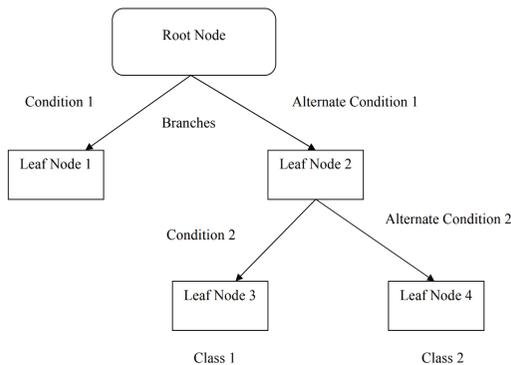

Fig. 1: A sample decision tree [12].

A falling rule list is a similar rule-based structure that checks just one condition at a time: the first condition assigns a probability for a certain label if it is true. If not, the next condition assigns another label with a smaller probability, until the end of the list. Often, to maintain simplicity and readability, designers will set a maximum size of the rule list or decision tree.

A linear model is a very commonly used method when the influence of each of the variables in the input, known as features, is direct. Each feature is given a numeric weight which indicates its contribution to the final output, a simple weighted average.

Model Explanation vs Outcome Explanation

Explainability can be achieved by solving any of a variety of relevant problems. The model explanation problem as defined by Giudotti et al. [15] provides a global explanation of a black-box through a second, interpretable model, an approach sometimes also known as rule extraction. This may take the form of a decision tree designed to approximate a black-box, as in the TREPAN algorithm, which we discuss in section 3.

The outcome explanation problem is the attempt to provide an explanation for the prediction of a single input. The explanation for the prediction is not necessarily generalizable to the whole model's decision-making, but it can be used to understand what features of the given input were most significant. The ultimate form of this explanation would be a human-readable sentence detailing how and why the values for influential variables in the input led to the output. However, it is common to find simpler strategies that require less semantic translation. In section 3, we detail the developments in building Class Activation Maps, which provide visual mappings of significant regions of input images, and in LIME, which designs locally-accurate models approximating the decision criteria of the neural net for the given input (and similar samples).

III. SELECTED APPROACHES

Rule Extraction

Creating a global interpretable approximation of a deep neural net has been an objective of the research community for decades. In one early work by Craven and Shavlik in 1996, the authors designed a supervised learning algorithm that would learn the rules within a neural net, which then was an improvement over a time-consuming search-based technique [6]. Rules can be conjunctive, i.e., made of and/or clauses, or M-of-N, meaning they define N criteria of which M need to be met. To create rules, example inputs and their class predictions are generated. For each example, if the example is not covered by the current rule for its class, it is used to contribute a new clause to the class's rule. This addition is done by generating a rule solely for that example and merging it with the current class's overall rule. The result of this processing is the creation of a set of rules for each class label – a user can check an input against the rule list and get the most likely prediction based on specific features.

In the same year, Craven also published a paper on the TREPAN algorithm, which produces a decision tree that approximates the neural network. TREPAN generates new examples (instead of drawing from training data) called queries, then expands the decision tree by adding nodes to account for the classification of new samples, prioritizing adding to branches that have the highest potential for improving the tree's fidelity to the network.

Two and a half decades later, De et al. expanded on TREPAN by incorporating hidden-layer clustering, a technique which groups/clusters the outputs of just the last hidden layer of neurons rather than the final predictions [8]. The last layer is where all the previous layers' knowledge is consolidated into a set of high-level features or patterns in the input space. It is hoped that this set represents similar features to those used by humans to distinguish the inputs, such as ears or a tail on an image of an animal. The hidden-layer clustering technique produces sets of input examples grouped together by their similar activations of the hidden layer neurons, which can be used to determine what internal features the hidden layer is detecting before they are weighted and added to find the final prediction. In the authors' hybrid approach, they

first cluster the hidden layer outputs as described, then build a TREPAN tree for each cluster, resulting in explanations more specific to each group of inputs that are deemed similar. This hybrid TREPAN method requires being able to access the intermediate values in the neural net as opposed to an entirely closed black-box model, but provides important insight into the significant features detected at the last level of the network.

In the original TREPAN, the decision tree model was created by an essentially blind following of the outputs of the original deep learning model. This produces a faithful explanation per input, but not necessarily the most interpretable one, because the features included in the decision tree could be tangential to the classification problem or simply counterintuitively ordered in the tree graph. With their approach called TREPAN Reloaded [10], Confalonieri et al. apply domain knowledge in the form of ontologies to create a tree that is informed by relevant concepts and interactions in the subject the model is classifying. The ontologies describing relationships between concepts and categories in the field are encoded via a commonly used method of terminology classification for computation called Description Logics. While new nodes are being added to the tree, their inclusion is also evaluated by a comparison to the domain knowledge connection, favoring branches that identify known relevant features.

The authors verified that the tree approximation models created with the influence of an ontology were more interpretable to users than those created without. However, a drawback of the TREPAN Reloaded method is that it requires the encoding of domain knowledge for every new application.

Although there are many approaches to rule extraction, TREPAN is one example that has stood the test of time and become a springboard for adaptions and improvements to this goal. Its iterations largely retain the same identity, but seek to address its shortcoming and blend it with outside strategies to create more accurate and interpretable model approximations.

Class Activation Maps

One of the most common approaches to understanding a prediction for a specific input is to create a saliency mask, which is a subset of the input that identifies

what was responsible for the prediction. This visual explanation method is most often applied to images or text in classification problems – the saliency mask is overlaid on the image and highlights the influential pixels.

Since 2016, the Grad-CAM algorithm designed by Selvaraju et al. has dominated image recognition with Convolutional Neural Nets [1]. Grad-CAM generates Class Activation Maps, a type of saliency mask, for each class label to identify the influential regions of images in CNNs of any architecture. The algorithm takes the final convolutional layer of the model and computes the gradients of the class scores with respect to each element in the last layer's feature activations to obtain how each class score depends on the final features detected in the model. This strategy of probing the final layer is similar to that used by De et al. in [8]. The gradients are then backpropagated through the layers to obtain the neuron importance weights, representing the importance of each feature map with respect to each class. Summing the importance of each feature map along with the map's activated (or highly influential) regions, the algorithm produces the final class activation map for the image, as seen in Fig. 2, which can be used to verify the model is using valid data to make its prediction.

A Grad-CAM problem noted by practitioners is that the resulting saliency map is not perfectly sensitive to the image's key features: the way that the gradients are backpropagated, they diminish and can all but disappear except for only the most salient pixels. In Fig. 2, this is seen in the highlighting of the dog's head but not its body, which also contributes to the prediction, albeit less strongly.
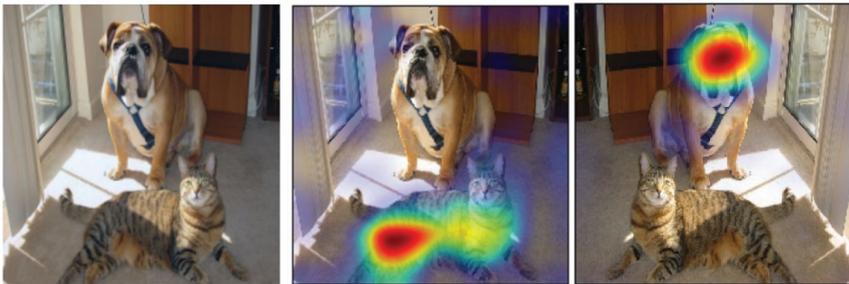

Fig. 2: (a) Original image of a cat and a dog. (b) Grad-CAM heatmap identifying influential regions of image for the class 'Cat,' applied to original image. (c) Heatmap identifying influential regions of image for the class 'Dog.'

In the following years, several works have iterated upon the Grad-CAM algorithm. One of these was Chattopadhay et al.'s Grad-CAM++ in 2018 [3]. Grad-CAM++ further generalizes Grad-CAM, allowing the saliency maps to have better accuracy and the ability to localize multiple instances of a class in an image (such as two cats). In Grad-CAM++, the derivatives are still used to create a weight for each class activation map, but now each element in the derivative matrix has its own weight, allowing for more fine-grained treatment of the maps. In Fig. 3, the authors provide examples of how Grad-CAM++ achieves better results than its predecessor.

Fig. 3: (a) The original images of swans. (b) Grad-CAM++'s heatmaps. (c) Grad-CAM's heatmaps, which fail to detect more than one swan or the entire body of a swan.

Axiom-Based Grad-CAM (XGrad-CAM) [9] was developed to improve the results of Grad-CAM and provide a simpler implementation than Grad-CAM++. The authors of XGrad-CAM are motivated by two axioms: sensitivity and conservation. Sensitivity is achieved when removing a feature from the input results in equal changes in the output and explanation. For example, if removing a feature from an image causes a large change in a class score, the feature's place in the class activation map should be of high importance. Meanwhile, the axiom of conservation is achieved when the explanation responses sum to the magnitude of the output. This implies that the resulting output can be wholly described by the maps and not other untracked causes. The creators of XGrad-CAM formulate these axioms mathematically and apply them

as constraints while learning the saliency map explanation as in Grad-CAM. This results in maps that are specifically trained to ensure they achieve the desired axioms, providing more explainability and accuracy.

LIME

Besides Grad-CAM, another foundational and often-iterated approach to the outcome explanation problem in XAI is LIME (Local Interpretable Model-agnostic Explanations), created in 2016 by Ribeiro et al [2]. Because the algorithm is model-agnostic, LIME can approximate any deep model "locally" by creating an interpretable model such as decision trees, linear models, or falling rule lists, centered on a specific input, to serve as an explanation for the result. The interpretable model is iteratively refined based on an objective function, which seeks to maximize the faithfulness of the explanation to the result, while minimizing the complexity of the model. In most uses, the explanation is a linear model, fit to the larger model at a point, similar to a tangent line.
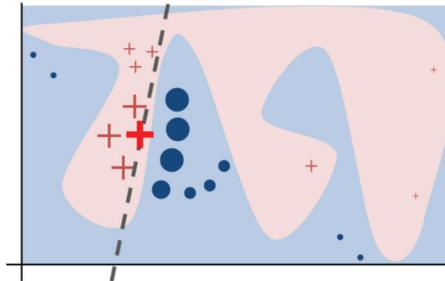

Fig. 4: Example to provide intuition for LIME.

In Fig. 4, a simple example of LIME is provided. The black-box model's decision boundary is represented by the blue/pink regions, and the bold red cross is the instance being explained. The other red crosses and blue circles are the instances sampled by LIME, which are used to learn a simple, more interpretable model (the dashed line) which explains the overall model's prediction for the given instance.

LIME requires an interpretable representation of the original input, such as a binary vector based on some features, set by the researchers. To generate the explanation for the sample, the algorithm samples instances near and far from the original instance, and generates a local model based on the predictions for these inputs.

This approach, the authors acknowledge, has limitations because the interpretable representation of the datapoint is not always relevant to the original classifier. The authors cite the example of detecting sepia-colored photos as retro; an interpretable representation of the input could be a binary vector of meaningful groups of pixels which indicate the presence of a certain feature, which does not apply to detecting sepia color.

The method can be useful in evaluating the trustworthiness of a trained model: one SVM with RBF kernel trained to classify whether a document is about atheism or Christianity ended up classifying one instance based on arbitrary features like the words "Posting", "Host", "Re". This model had 94% accuracy on the test data but the explanation revealed that the model is not trustworthy, likely because the dataset was too easy to classify with trivial features.

Like Grad-CAM, many iterations upon the LIME algorithm have been developed in recent years, which often focused on addressing the known flaws in the original algorithm – namely, its instability: researchers found that the samples surrounding the given instance, which are randomly generated each time the algorithm is run, created variance in the final explanation across runs. This can happen when the collection of samples fails to account for the full complexity of the model decision boundary surrounding the given instance. One example of an attempted improvement is S-LIME, which was designed to stabilize LIME by producing similar results each time when run on the same training data. In S-LIME, to guarantee that the samples are more representative of the model, the samples are verified for significance and appropriate variety before the explanation learning step [4].

OptiLIME, designed by Visani et al. [11], similarly seeks to address LIME's instability but takes a different direction: instead of finding an explanation after constraining its search to stable sample points, this algorithm does the inverse by guaranteeing stability while constraining its search to find relevant local explanations. Importantly, OptiLIME allows the user to choose the level of adherence, meaning the accuracy of the local explanation. It then maximizes the stability of the local explanation model by choosing hyperparameters that produce the most reproducible

method to generate sample points, all while enforcing that the resulting local model meets the desired level of adherence. OptiLIME's framework gives practitioners direct access and control over the priorities of the explanations (adherence or stability) which can help them put trust in the results.

The LIME and Grad-CAM approaches to justifying the prediction for a certain sample have served as successful starting points to the outcome explanation problem. Their adaptations in the last five years demonstrate both their strength and the high potential for improvement in the field. Many researchers have taken up this challenge, often choosing to work closely with the original design while producing meaningful improvements.

IV. CONCLUSION

In this article, we have presented three foundational approaches to augmenting deep neural nets to improve their explainability. We provided an overview of the approaches' methodology and goals to create an introduction to how XAI is implemented. Further, we identified and explored recent advancements to these algorithms' core designs, providing insight into the collaborative and iterative approach to improving XAI in the field. Through the literature for TREPAN, Grad-CAM, and LIME, we attempted to create a representative look at the current state of artificial intelligence interpretability. We have seen that although progress continues, there is no perfect solution for overcoming the complexity and opacity inherent in black-box neural networks. As society increasingly relies on AI predictions for real-world applications with potential consequences, it is critical to drive strategies like XAI that ensure these systems are fair and correct, drawing from novel inspiration as well as improving promising existing works.

REFERENCES

[1] Ramprasaath R. Selvaraju, Abhishek Das, Ramakrishna Vedantam, Michael Cogswell, Devi Parikh, and Dhruv Batra. 2016. Grad-cam: Why did you say that? Visual explanations from deep networks via gradient-based localization. In *2017 IEEE International Conference on Computer Vision (ICCV)*. DOI: 10.1109/ICCV.2017.74. 618-626.

[2] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. Model-agnostic interpretability of machine learning. *arXiv preprint arXiv:1606.05386* (2016).

[3] Aditya Chattopadhay, Anirban Sarkar, Prantik Howlader, and Vineeth N Balasubramanian. 2018. Grad-CAM++: Generalized Gradient-based Visual Explanations for Deep Convolutional Networks. In *2018 IEEE Winter Conference on Applications of Computer Vision*, 2018, pp. 839-847. DOI: 10.1109/WACV.2018.00097.

[4] Zhengze Zhou, Giles Hooker, and Fei Wang. 2021. S-LIME: Stabilized-LIME for Model Explanation. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '21)*, August 14–18, 2021, Virtual Event, Singapore. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3447548.3467274.

[5] Sheng Shi, Yangzhou Du, and Wei Fan. 2020. An Extension of LIME with Improvement of Interpretability and Fidelity. *arXiv preprint arXiv:2004.12277* (2020).

[6] Mark Craven and Jude Shavlik. 1994. Using sampling and queries to extract rules from trained neural networks. *In Proceedings of the International Conference on Machine Learning (ICML'94).* 37–45.

[7] Mark Craven and Jude W. Shavlik. 1996. Extracting tree-structured representations of trained networks. *In Proceedings of the Conference on Advances in Neural Information Processing Systems.* 24–30.

[8] Tanusree Dea, Prasenjit Giria, Ahmeduvesh Mevawalaa, Ramyasri Nemania, and Arati Deoa. 2020. Explainable AI: A Hybrid Approach to Generate Human-Interpretable Explanation for Deep Learning Prediction. In *Procedia Computer Science*, Volume 168, 40-48. DOI: https://doi.org/10.1016/j.procs.2020.02.255.

[9] Ruigang Fu, Qingyong Hu, Xiaohu Dong, Yulan Guo, Yinghui Gao, and Biao Li. 2020. Axiom-based Grad-CAM: Towards Accurate Visualization and Explanation of CNNs. *arXiv preprint arXiv: 2008.02312* (2020).

[10] Roberto Confalonieri, Tillman Weyde, Tarek R. Besold, and Fermin Moscoso del Prado Martin. 2020. TREPAN Reloaded: A Knowledge-driven Approach to Explaining Black-box Models. In *24th European Conference on Artificial Intelligence (ECAI 2020)*.

[11] Giorgio Visani, Enrico Bagli, and Federico Chesani. 2020. OptiLIME: Optimized LIME Explanations for Diagnostic Computer Algorithms. *arXiv preprint arXiv: 2006.05714* (2020).

[12] Awudu Karim and Shangbo Zhou. 2015. X-TREPAN: A Multi Class Regression and Adapted Extraction of Comprehensible Decision Tree in Artificial Neural Networks. *arXiv preprint arXiv: 1508.07551* (2015).

[13] Sina Mohseni, Niloofar Zarei, and Eric D. Ragan. 2021. A Multidisciplinary Survey and Framework for Design and Evaluation of Explainable AI Systems. *ACM Trans. Interact. Intell. Syst.* 11, 3-4, Article 24 (August 2021), 45 pages.

[14] Amina Adadi and Mohammed Berrada. 2018. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, 6, October 2018, 23 pages. https://doi.org/10.1109/2870052

[15] Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, and Dino Pedreschi. 2018. A Survey of Methods for Explaining Black Box Models. *ACM Comput. Surv.* 51, 5, Article 93 (August 2018), 42 pages. https://doi.org/10.1145/3236009

ESSAYS BY
Gabriel Consiglio
Ella Olesen
Josephine Schmidt
Ethan VanValkenburg
Anastasia Warner
Grey Weinstein