The University of Texas at Austin

Undergraduate Research Journal

Volume 5 • Number 1
Spring 2006

The University of Texas at Austin Undergraduate Research Journal is a student edited and published multidisciplinary journal of undergraduate research.

The University of Texas at Austin

UNDERGRADUATE RESEARCH JOURNAL

Volume 5 • Number 1 • Spring 2006

Undergraduate Research Journal

Inquiries should be addressed to:

The 2005-2006 staff presents the fifth edition of the University of Texas at Austin Undergraduate Research Journal (URJ). This was an exciting year for the journal, chiefly because of our structural changes. We named Dr. Jay Koehler as our Faculty Editor, who formed a Faculty Editorial Board to streamline the faculty review and revision processes. Dr. Koehler's work freed the staff to focus on soliciting submissions from a variety of new sources, and that focus was reflected in the diversity of submissions.

In addition to Dr. Koehler, I would like to thank Ms. Annie Elderbroom for her invaluable advice and resourcefulness this year. Their enthusiasm and passion for the Journal over the past few years have really solidified its presence on campus and made things much easier for the staff.

The changes we made this year are clearly demonstrated in the selected articles. Not only did we receive more quality submissions than ever before, but we were also able to evaluate and revise these submissions more closely. The end result is a collection of thoroughly researched and clearly written articles that showcases the quality of work done by our undergraduates. I hope you enjoy reading our most varied and interesting (in my opinion) articles yet.

The URJ still has a lot of work to do in trying to find the best research done on campus. We especially would like more submissions from the Colleges of Engineering and the Natural Sciences. If you would like to join our staff or submit a paper please visit our website at *http://www.utexas.edu/student/urj* for more information.

Sincerely,

Zachary Smith, Editor-in-Chief
URJ Editorial Staff

Contents

Applicability of the Geneva Convention (III) to the Global War on Terrorism: History and Perspectives[1]

Ryan Kingsbury

Abstract

The steady rise in terrorist activities in the past few decades and the recent outbreak of a global war on terrorism have forced a drastic change in the strategy and policy with which our nation fights war. The terrorist is a chimerical class of combatant who does not neatly fit into any criminal or military category, and as such, have forced policymakers to revisit nearly a century of policy and practice in the conduct of warfare. Recently, the reasoning that terrorists are substantively different from traditional armed combatants has been used to deny the applicability of one of America's most highly valued treaties—the Geneva Convention on Prisoners of War—to their capture and detention. Such a drastic step demands critical evaluation. This paper examines the reasoning

for rejecting the Conventions in detail, and then searches the history of the drafting of the Conventions to see whether or not "unconventional" combatants were accommodated in its provisions. Historic interpretations and applications of the Conventions by the United States are then presented. In the end, it appears that the architects of Geneva were fully aware of the problems posed by unconventional warriors and made efforts to address those problems in the texts of the Convention and subsequent agreements. Furthermore, many of the key arguments used by the U.S. to dismiss the treaty contradict the intent and purpose expressed by the drafters. Historically, however, the U.S. has displayed a tendency to apply the treaty only insofar as it does not conflict with its own specific security interest. As such, while selectively ignoring certain rules of the Conventions is not a novel policy move for the U.S., questioning the applicability of the entire treaty to a conflict is absolutely unprecedented, and displays a break with the values of the international community and a severe misinterpretation of the original meaning and purpose of this very important treaty.

Introduction

The steady rise of terrorist activities over the last two decades, leading ultimately to the outbreak of the Global War on Terrorism, has forced policymakers to confront a chimerical class of new enemies—partly criminal, partly military—whose conduct does not appear to fit neatly within traditional legal or strategic frameworks. Prior to September 11, terrorist attacks were few enough and small-scale enough to be considered but exceptions to the overriding rule of regular military and criminal activities in the world, marginalizing the problems introduced by their prosecution. However, the launching of the War on Terror has inverted that calculus, making terrorist suspects our primary enemies, while regular military battles fade slowly into the pages of 20th century history.

That inversion has required policymakers to revisit nearly a century of policy and practice in the conduct of warfare in order to adequately protect their nations against the new terrorist threat. Or so the argument goes.

After all, terrorists are indeed substantially different from traditional military forces—a "new paradigm" in the eyes of some[2]—so certainly the tactics, policies, and treaties that were developed over the last century to deal with regular units will not be effective or appropriate for fighting the War on Terror. This line of reasoning, which we will examine in detail later, led the United States to question and finally to deny the applicability of one of its most highly-valued treaties—the Geneva Conventions of 1949—to the war against al-Qaeda and the Taliban.[3] Such an extraordinary measure, and one that is demonstrably out of step with traditional U.S. policy, has enormous potential consequences for the safety of our own soldiers as well as the credibility America has with her allies. As such, this decision demands a critical evaluation that takes all these factors into account: is the War on Terror really distinctive enough to justify such a drastic move? What circumstances render it so completely new that we must revisit and in some cases rescind the agreements made in the conventional war era? Does global terrorism really constitute a "new paradigm," or is it merely a new, monstrous incarnation of a century old problem?

As new and threatening as these quasi-military, quasi-criminal, highly dangerous groups seem, they do have a history, and a thorough knowledge and examination of that history is notably absent from much of the debate over the Geneva Conventions within the Executive Branch.[4] It is frequently taken for granted that the Conventions were written only with traditional wars in mind, and so are relevant only to large-scale conflicts between nation states.[5] But a study of the treaty's roots demonstrates the reverse: the original drafting committee had much to say regarding the status of rogue groups and non state actors.[6] Additionally, the United States' experiences in unconventional conflicts stretching from Vietnam to the present contain highly relevant precedents in the interpretation of that treaty. A closer look at how Geneva was drafted and how it applied to various types of combatants reveals that the Convention and other international legal agreements remain relevant to conflicts against terrorists. Moreover, concerns about over-protecting dangerous people such

as terrorists were present even at the drafting of the Conventions in 1949,[7] and were incorporated into the treaty's terms. Finally, although American military units' depriving detainees of some rights and otherwise "bending the rules" of the treaty is not a new phenomenon, questioning the applicability of the Geneva Conventions as a whole constitutes a novel and dramatic policy move that will be difficult to justify in the eyes of the international community.

Background on the Conventions

Any informed discussion of the relevance of the Geneva Conventions to the present conflict must necessarily go beyond specifics of their terms and address the more broadly intended purpose of those agreements. The ancestors of the Third Convention—the series of accords and treaties that preceded it—demonstrate a clear intent to broaden the category of combatant and protect as many people as possible. The founding of the International Committee of the Red Cross and the drafting of the original Geneva Convention in 1864 were both motivated by a need to aid the wounded,[8] who before that time were frequently left to die by their enemies. It was followed by the Hague Conventions of 1899 and 1906, which expanded the 1864 Convention's scope to include wars at sea, and established new rules for the conduct of lawful warfare.[9] Another agreement— the 1907 Hague Convention IV—defined combatant status and introduced the idea of prisoner of war (POW), providing the first rudimentary protections for captured personnel, but did not actually define who qualified as a POW.[10] When it became apparent during World War I that this agreement was too ambiguous to protect soldiers adequately, several nations clarified it by private agreements at Berne in 1917-18.[11] These ad hoc remedies exposed some of the Hague Convention's deficiencies, and ultimately led to the passage of the Geneva Conventions of 1929, one of which was devoted entirely to POWs. It attempted to clarify the definitions in the Hague convention and addressed a number of loopholes that had been used to deprive soldiers of their rightful status during wartime.[12] However, even the 1929 Convention proved too vague to prevent widespread

outrages upon the dignity of prisoners in World War II, and this fact led to the drafting of the 1949 Conventions to which the U.S. is now party.[13]

The revisions and augmentations of the 1929 agreement that form today's Geneva Convention appear designed to accomplish two goals. The first is to prevent abuses of the treaty. The official Red Cross Reporter to the drafting convention notes that "it had become necessary to widen the scope of the term 'prisoners of war' so as to include members of the armed forces after capitulation and in order to avoid arbitrary loss of that status at any given moment."[14] He refers, as an example, to the situation in which a soldier surrenders, and is subsequently denied POW status because the 1929 Convention defines a POW as someone who is "captured by the enemy."[15] This type of abuse occurred frequently during WWII, according to the Reporter.[16]

Secondly, the 1949 treaty includes irregular combatants in its provisions for the first time. The activities of organized resistance movements during WWII informed the drafter's understanding of what types of combatants should be protected. These fighters were frequently denied POW status, or any type of protection, as a result of narrow interpretations of some of the treaty's provisions. For instance, resistance groups might be treated merely as domestic criminal bands, since they operated strictly within the territory of a single country, and owed allegiance to a dissident authority that the *de facto* government was unlikely to acknowledge.[17] This abuse prompted one revision specifying that a soldier need not belong to a government recognized by the power that captures him to qualify as a POW.[18] This is but one example.

The development and revision of these conventions ought to make clear that the contracting parties desired a definition of POW that would not only be precise enough to prevent abuse resulting from narrow state interpretations, but also that would include as many soldiers and irregular military units as could fairly be called combatants. What began as a movement to protect wounded soldiers gradually developed into a large body of accords designed to protect individuals from unfair punishment for their actions in military or

quasi-military units. It is worth noting that through all of the revision and expansion of the qualifications for prisoner of war status, the conditions for lawful combat were never changed, but remained those established by the Hague Conventions.[19] Like its predecessors, the Geneva Convention of 1949 required combatants to comply with four conditions:

- That of being commanded by a person responsible for his subordinates
- That of having a fixed distinctive sign recognizable at a distance
- That of carrying arms openly
- That of conducting their operations in accordance with the laws and customs of war[20]

These requirements were absolutely constant throughout the development of all the treaties detailed above, so the ambiguity that the drafters tried repeatedly to correct was strictly that associated with interpretation of those four conditions. Once it became clear that different nations would interpret them differently to suit their purposes, more and more specific qualifiers had to be added to the treaties to guide that interpretation.

It is important to understand that the drafting history of the Conventions demonstrates a contempt for this kind of narrow interpretation of a treaty's provisions: every new agreement came about precisely because the older one could not prevent such misuse. Consequently, as we enter the debate over whether the 1949 Convention (III) ought to apply in the War on Terrorism, it is critical to avoid these types of strict readings, as they appear to violate the intent and purpose of the treaty and its ancestors.

The Executive Debate

In the distressing weeks following September 11, as we all know, the government felt tremendous pressure to find a way to respond to the attacks, struggling to determine an effective way to deal with the terrorists. Among its responses were a declaration of National Emergency by President Bush on September 14, and then a

Presidential Military Order dated November 13. In the Order, President Bush established the groundwork for what would ultimately ignite the Geneva debate: it first granted the Secretary of Defense broad authority to detain any individual suspected of terrorist activities or ties, and second directed such individuals to be tried before military commissions for violations of the Law of War.[21] The President concluded his memo by saying that "an extraordinary emergency exists for national defense purposes. . . [which] . . . constitutes an urgent and compelling government interest, and that issuance of this order is necessary to meet the emergency."[22] In this order we already see elements of the central argument against applying Geneva—that the terrorist threat is extraordinary and new, and justifies extraordinary measures.

But putting those measures into practice would raise further debate, of course, because the Presidential Order did not spell out exactly how the military commissions should work, or under precisely what conditions prisoners should be kept. The war in Afghanistan presented American forces with a hodgepodge of detainees that could be either innocent civilians, members of the quasi-governmental Taliban militia, or al-Qaeda operatives, and raised practical questions about precisely how to treat each category. Unfortunately, the Presidential Order said only that prisoners were to be "treated humanely," "afforded adequate food, drinking water, shelter, clothing, and medical treatment," and be allowed "free exercise of religion consistent with the requirements of . . . detention."[23] These measures alone may not have satisfied all the requirements of the Geneva Conventions, however, so on January 9, 2002, a detailed memorandum was issued by Deputy Assistant Attorney General John Yoo to clarify America's legal obligations under the treaty.[24] This memo constitutes by far the most detailed treatment of the Geneva question by any member of the administration; it effectively provided all the arguments that later officials accepted or refuted.[25]

Yoo's memo first approaches the Geneva question by distinguishing between Taliban militia and al-Qaeda detainees, and treating each group separately. We will first address the arguments against applying Geneva to the al-Qaeda terrorists.

The rationale for denying al-Qaeda fighters POW status comes strictly from within the Convention itself. Yoo first notes that Article 2 of the Convention, which defines the scope of application of the treaty, covers only "High Contracting Parties" and other non-signatory state actors who pledge to abide by the treaty's requirements.[26] Al-Qaeda did not sign the treaty, nor are they a state, nor have they expressed any willingness to respect the Geneva Conventions; therefore the Convention should not apply.[27]

Even so, it is possible that the principles of the Convention would be called for if the conflict fell under the jurisdiction of Article 3, which requires that various minimal provisions be applied in armed conflicts "not of an international character occurring in the territory of one of the High Contracting Parties."[28] Here Yoo claims that the global war against al-Qaeda clearly *is* an international conflict, and so would not fall under Article 3, and that it is not confined to the territory of one Party.[29] Finally, Yoo reminds his readers that even if Geneva somehow did apply to the war with al-Qaeda, their operatives all seem to violate the rules for lawful combat, and so would not be entitled to POW status anyway.[30]

Each of these arguments makes a very strong case against granting al-Qaeda fighters POW status. But while each point presents an apparently reasonable reading of the treaty's terms, Yoo acknowledges that such strict readings may not be consistent with recent trends in international law:

> This [recent] approach gives central place to individual human rights. As a consequence, it blurs the distinction between international and internal armed conflicts, and even that between civil wars and other forms of internal armed conflict. . . . In this conception, common Article 3 is not just a complement to Article 2; rather, it is a catch-all that establishes standards for any and all armed conflicts not included in common Article 2.[31]

While recognizing that he is at odds with this trend in the international community, Yoo ultimately concludes that the above reading is inconsistent with both the language of the treaty and the historical context in which it was drafted,[32] leaving apparently little doubt in the minds of his readers that al-Qaeda operatives do not deserve POW status.

Just how many protections the Taliban militia are due is a far more complicated question for Yoo. Unlike al-Qaeda, the Taliban could possibly be considered a regular military force, representing the *de facto* government of Afghanistan. If that were the case, they would be plainly within the jurisdiction of the Convention. Yoo uses two lines of reasoning to argue that this is not the case. The first suggests that Afghanistan should be considered a "failed state," and therefore should not be considered a High Contracting Party, even though it signed the Conventions and was held to them in the past.[33] This argument depends on classifying the Taliban not as a legitimate government, but merely as an unlawful military presence inside the country. If one accepts these premises, Geneva is rendered inapplicable both on the grounds that there is no conflict "in the territory of a High Contracting Party," and that it is impossible for the Taliban to constitute a regular military force when there is no legitimate government in place. If the Conventions as a whole do not apply, then the Taliban would only be due POW protections if they expressed a desire to abide by the terms of the Convention, which they plainly did not.[34] Secondly, he argues that the Taliban militia fail to meet the four requirements of lawful combat outlined in Article 4 by their lack of uniforms and disregard for the Laws of War.[35] Therefore, even if the Conventions do apply to the conflict, Taliban warriors could not qualify as POWs.

John Yoo's arguments and analysis constitute the meat of the evidence cited in the administration's debate over the Conventions. In a series of five more memos between various department leaders, virtually no one challenges the quality or legitimacy of the legal analysis in the Yoo memo. Rather than trying to determine America's legal obligations under the treaty, most leaders dwell on the pros and cons of determining that Geneva does not apply.[36] Whether or not to apply the Conventions becomes a political rather than a legal question. Those who argue against applying Geneva essentially rehash Yoo's arguments and add some of their own

Ryan Kingsbury

opinions on the advantages of that approach, such as increased flexibility in handling terror suspects.[37] Two others, however, do criticize the prevailing analysis. Secretary of State Colin Powell points out that a determination of Afghanistan as a failed state contradicts the official position of the State Department. He also broadly challenges Yoo's analysis with the claim that the "[Office of Legal Counsel's] views are not definitive on the factual questions that are central to its legal conclusions."[38] Although Powell does not specify exactly which facts he is referring to, he does directly confront several of the assertions in the other memos, such as the idea that ignoring the Conventions provides greater flexibility in detainee treatment.[39] Additionally, Department of State Legal Adviser William Taft takes issue with separating al-Qaeda from the Taliban, claiming that "the Conventions call for a decision whether they apply to the conflict in Afghanistan. If they do, their provisions are applicable to all persons involved in that conflict. . ."[40]

All these arguments culminated in the Presidential memorandum of February 7, 2002. This order effectively separates the Geneva question into two issues—the applicability of the Convention to the war in Afghanistan as a whole and the eligibility of various detainees for POW status. It rules that the Geneva Conventions apply to the conflict in Afghanistan, but not to the global conflict against al-Qaeda. Further, it asserts that Taliban warriors are unlawful combatants within the Convention framework, and so are not entitled the rights of POWs, even though the broader protections of the treaty still apply.[41]

This decision has had tremendous consequences for the United States and its soldiers. The government is now embattled over the apparently indefinite detentions and lack of due process being granted to detainees at Guantanamo Bay, while the decision that no protections are due to al-Qaeda operatives has essentially opened the door for arbitrary deprivation of POW status: it has provided a simple and efficient way to sidestep international legal guarantees for anyone suspected of terrorism.

The dialog leading up to this momentous decision, as we have seen, fixated so much on the political and military implications of a decision not to apply Geneva that it marginalized consideration of the history of U.S. detention policy, the drafting and intent of the Conventions, and more recent additions to international law. Further exploration of each of these factors is certainly in order if we are to adequately understand and assess the President's decision.

The Drafting of Article Three

> It must not be forgotten that the Conventions have been drawn up first and foremost to protect individuals, and not to serve state interests.[42]

The humanitarian ideals and practical challenges that led to the 1949 Conventions have already been discussed in brief, but in light of the arguments above, a more thorough examination of how the provisions of Geneva were developed is in order. We have seen that the Executive determination about the treaty rested largely on interpretations of common Articles 2 and 3 of the Conventions. The drafting history of Article 3 in particular contains a number of relevant insights into its intended purpose.

In keeping with the gradual widening of the scope of the Conventions that we noted earlier, the drafters of the 1949 Convention—representing nations from around the world— proposed including some conflicts "not of an international character" in the treaty's jurisdiction, thus extending the authority of international humanitarian law inside state borders for the first time by including civil wars and the like. Their attitudes are clearly reflected in the quote at the head of this section; as the international community placed greater emphasis on individual rights as opposed to state security, the categories of protected individuals became more and more inclusive. So, in drafting Article 3, the committee took special care to keep their provision from being interpreted too narrowly. They chose to use the phrase "armed conflict" in paragraph 1, rather than "war," to avoid any doubt about the fact that "the occurrence of *de facto* hostilities is sufficient"[43] for the Convention to apply. They also defined "prisoner of war" as a person who had "fallen

into the power of the enemy" rather than who was "captured." Official commentary explains that this wording was chosen to ensure that no soldier who surrendered would be deprived of his POW status on a technicality, as frequently occurred during the Second World War.[44] The official Reporter noted that the scope of application of Article 3 was intended to be "as wide as possible."[45] It is, therefore, undeniable that the attitude of the contracting parties was one of universal respect for the authority and provisions of the Convention, even in dubious cases. When, in Article I, the signers pledge to "ensure respect" for the Convention "in all circumstances,"[46] this intent is made plain.

Further, the ways in which the historical circumstances of the 1949 Convention led the drafters to emphasize individual rights illuminate the real purpose of Article 3. World War II highlighted the importance of organized resistance movements and brought new attention to their status, as noted above. The frequent capture and summary execution of such freedom fighters weighed heavily on the minds of the drafters, who approached the new treaty with every intention to broaden, not limit the category of combatants who qualified for protection.[47] As such, the provisions of Geneva that relate to irregular combatants must be interpreted in this context, as the treaty's provisions were fashioned particularly, though not exclusively, with the protection of this type of combatant in mind.

But while the overall attitude was unquestionably one of broadening the POW category and protecting individual rights, strong concerns about granting terrorists, anarchists, and criminals undue protections were nonetheless quite prevalent at the drafting convention.[48] These concerns were very similar to the concerns expressed by the Executive even today: the formulation of common Article 3 was plagued by the same controversy and debate that divides policymakers now. When the International Committee of the Red Cross proposed applying the full force of the Geneva Conventions to civil wars and the like, numerous states protested that such a move would create a shelter for all sorts of criminal groups, not unlike today's terrorists. In the words of the Reporter:

It was said that it would cover all forms of insurrections, rebellion, and the break-up of States, and even plain brigandage. Attempts to protect individuals might well prove to be at the expense of the equally legitimate protection of the State. To compel the Government of a State in the throes of internal conflict to apply to such a conflict the whole of the provisions of a Convention expressly concluded to cover the case of war would mean giving its enemies, who might be no more than a handful of rebels or common brigands, the status of belligerents, and possibly even a certain degree of legal recognition. There was also a risk of ordinary criminals being encouraged to give themselves a semblance of organization as a pretext for claiming the benefit of the Convention, representing their crimes as "acts of war" in order to escape punishment for them.[49]

The committee entertained many proposed solutions, some of which aimed at specifically differentiating between internal conflicts that constituted actual wars and others that were merely hostile criminal acts. Ultimately, they decided to apply only the principles of the Convention to the case of internal conflict. However even this suggestion brought controversy, because of the term "armed conflict." "The expression is so general, so vague, that many of the delegations feared that it might be taken to cover any act committed by force of arms—any form of anarchy, rebellion, or even plain banditry."[50] So serious were these concerns that a list of criteria for qualifying conflicts was proposed.[51] Ultimately, however, these qualifiers were dropped, as it was agreed that a true terrorist group, by its disregard for humanitarian principles and the law of war, would automatically forfeit the protections of the Conventions, even in spite of a broad statement demanding their application to internal conflicts.[52] Thus in this matter we find clear support for Yoo's assessment of the Convention—that Article 3 was never intended to cover true terrorists like al-Qaeda members. At the very least, it is worth noting that Article 3 of the Conventions was framed with the presence of unlawful terrorist groups in mind.

Broadening the Conventions into such controversial territory created a need to build in protections against abuse by criminal and terrorist groups in order

Ryan Kingsbury

to protect state security. One such provision appears in Article 5: "should any doubt arise as to whether persons… belong to any of the categories enumerated in Article 4, such persons shall enjoy the protections of the present Convention until such time as their status has been determined by a competent tribunal."[53] This provision illustrates once more the desire of the drafters to emphasize individual rights over military expediency. The presence of this clause allows POW status to be denied to terrorists and other criminals, but effectively requires a presumption of that status until they have received proper process. Unfortunately, neither the commentary to the Conventions nor the treaty itself says anything about what constitutes a "competent tribunal," or how to determine whether "doubt" exists, leaving the practical execution of this Article to the judgment of individual states.

This clause provides a procedure for resolving cases of doubt regarding POW status that would apparently obviate all the debate within the administration. Why then, has the Executive chosen not to avail itself of this provision, but instead to attempt to sidestep the Conventions? It seems that no clearer case of detainees whose "status is in doubt" could possibly exist, yet the obligation of convening "competent tribunals" comes up only once in the entire debate, when Assistant Attorney General Bybee argues that "a presidential determination [that the Taliban are not POWs] would eliminate any legal "doubt" as to the prisoner's status, as a matter of domestic law, and would therefore obviate the need for Article 5 tribunals."[54] A look at the commentary to this article, however, demonstrates serious flaws in this interpretation of the treaty. The Reporter recounts that originally, Article 5 called for a "responsible authority" to determine the detainee's status, but this wording was changed to "military tribunal" because the drafters did not want a decision of that magnitude left to a single individual. Given this insight, Bybee's proposal that the President make such a broad determination regarding the status of Taliban fighters is demonstrably against the intent of Article 5. Furthermore, the drafters feared that the consequences of a sentence by a "military tribunal" could be more severe than those

resulting from a loss of POW status, so the drafters changed the language again to read "competent tribunal."[55] Thus, the President's order to use military commissions to try detainees without properly determining their status goes even further against the will and motivation of Article 5. If the arguments that Taliban warriors cannot be POWs are so persuasive, it seems that it would have been wiser for the government to apply the Conventions, while availing itself of the provisions of Article 5 to deny POW status through a tribunal, as was done in Vietnam for a time. That conflict, along with several others, provides a great deal of insight into how, or if, this and other protective measures of the Geneva Convention have been used to prosecute dangerous individuals.

Historical Application of Geneva

The first conflict that truly tested American respect for the Geneva Conventions was Vietnam, and it was really the only major conflict after the signing of the treaties in which U.S. forces encountered unconventional combatants on a large scale. The Vietcong had a great deal in common with the Taliban in the way they waged warfare, and so their treatment is particularly relevant to this discussion. Vietcong were often indistinguishable from the civilian population. It is also highly questionable that they conducted their operations according to the commonly accepted Law of War. But, they did generally display their weapons openly, and obviously obeyed some type of central command structure. And in Vietnam, as in Afghanistan, the multitude of enemy factions and the confusing status of different groups generated frequent doubts about some captive's rights. The initial American response was to institute "Article 5 Tribunals," which granted a detainee fundamental process rights and were weighted in favor of finding prisoner of war status.[56] Later, the United States issued a directive to its armed forces indicating that all detainees were to be treated as POWs whenever there was any reason whatsoever to believe they belonged to a military unit.[57] It must be remembered that this broad application of the treaty came at a time when the North Vietnamese were denying U.S. airmen POW status because they considered pilots who bombed civilian

targets war criminals.[58] Furthermore, the Vietcong / NLF were not a signatory party to the treaties, so in a strict sense Geneva did not apply to their soldiers, and they refused to apply it to ours.[59] Thus, the United States' decision to apply the Conventions was a matter of policy, not legal requirement: it was decided to afford POW protections to warriors that clearly were not entitled to such treatment, even while American soldiers were being denied the protections they deserved. This application of the Geneva Conventions earned high praise from the international community. In the words of an International Committee of the Red Cross representative: "this [action] could very well be a most important one in the history of the humanitarian law, for it is the first time (. . .) that a government goes far beyond the requirements of the Geneva Conventions in an official instruction to the armed forces."[60]

Lest this response appear too progressive an interpretation of Geneva, however, it should be noted that this decision could be explained merely as a protective measure for our own soldiers. In two cases in 1966, before the directive was issued, Vietcong detainees were executed by South Vietnam as terrorists (presumably after facing an Article 5 tribunal), since they obviously failed to meet the criteria for lawful combat. Both executions prompted the immediate killing of American hostages.[61] These murderous reprisals communicated an effective message that surely played some part in America's decision to apply the Convention so broadly. Nonetheless, Vietnam was the first of several conflicts in which the United States chose to apply the Conventions as a matter of policy, even when there was not necessarily a legal obligation to do so.

Despite its exemplary interpretation of the Convention in Vietnam, in two other conflicts involving "high value" detainees, the United States has displayed a willingness to ignore certain provisions of the treaty in the interest of detaining and prosecuting a prisoner. First, when Marines invaded Grenada in 1983, they detained the former Deputy Prime Minister Bernard Coard incommunicado for some two weeks aboard a warship, blindfolding him and his pregnant wife and binding his feet, both of which violated the Conventions.[62] He was suspected of a number of heinous crimes, and was presumably detained so that he could be brought to trial, but was never allowed to see a lawyer during his detention. Additionally, the U.S. claimed the right to detain civilians whom it deemed dangerous.[63] Other prisoners, who were held for being combatants, were restricted from access to the media. Although the status of many of the combatants was not completely clear, the U.S. did claim to be holding them in accordance with Geneva.[64]

In another case, following the invasion of Panama in 1989, the Executive sidestepped technical issues of interpreting the Convention by deciding in its favor. When General Noriega was captured and put on trial for drug charges, he claimed POW status and alleged that the U.S. courts had no jurisdiction to try him.[65] After initially rejecting the claim altogether, the prosecution eventually decided to classify him as POW as a matter of policy, whether or not he actually qualified.[66] After his trial a Federal District Judge did rule that he qualified as a POW, but the effect of this status on his trial and conviction was insignificant.

Noriega's case is interesting because it demonstrates the extent to which applying the Geneva protections is perfectly compatible with criminal prosecution. Indeed, the Reporter to the 1949 Convention said "no sort of immunity is given to anyone under [Article 3]. There is nothing in it to prevent a person presumed to be guilty from being arrested and so placed in a position where he can do no further harm; and it leaves intact the right of the State to prosecute, sentence and punish according to the law."[67] The conviction and imprisonment of Manuel Noriega is a striking example of the extent to which it is possible for the U.S. to apply all the provisions of Geneva while still protecting its own interests by prosecuting prisoners accused of crimes. It is interesting to note, however, that in the case of "high value" detainees like Noriega and Coard, this attitude has never prevailed. The prosecution's initial reaction to Noriega's POW claim was a flat denial; only later did they agree to treat him as a POW as a matter of policy.[68] Similarly, Coard could easily have been tried for war crimes under the provisions of Geneva, but was instead denied the

rights a POW should expect. These two cases illustrate that the U.S. has shown a preference for bending the rules of the Conventions over utilizing the provisions it supplies for the prosecution of criminals. Despite that tendency, however, the absolute authority of the Geneva Conventions has never been questioned. Nowhere was it suggested that the Conventions might not apply in a broad sense, but only in a few cases the rights of individuals were curtailed unlawfully in the interest of security. Apart from those few individual exceptions, however, it appears that virtually every combatant detained by the U.S. has been treated (at least nominally) according to the Geneva Conventions and granted POW status, regardless of their qualification.

The Rejection of Protocol I

Despite that trend, however, the U.S. has been hesitant to codify any binding presumption of POW status. So important was that presumption to the international community that it was finally written into the Protocol Additional to the Geneva Conventions of 12 August 1949, Relative to the Victims of International Armed Conflict (Protocol I) in 1977. Article 45(1) plainly states "a person who takes part in hostilities and falls into the power of an adverse Party shall be presumed a prisoner of war, and therefore shall be protected by the Third Convention, if he claims the status of prisoner of war, or if he appears to be entitled to such status. . . ."[69] The same article reiterates the provisions of the Third Convention regarding the need for a "competent tribunal" to resolve doubts about the status of a combatant. In general, Protocol I removes whatever ambiguity the Third Convention contained regarding what types of combatants are POWs: the Protocol makes POW status the default condition of a detainee, placing the burden of proof on a tribunal to remove that status. Needless to say, this Protocol encountered all the same debate as common Article 3 regarding the abuse of such protections by criminal groups. Ultimately however, the signers of Protocol I determined individual rights to be their paramount objective, and crafted a treaty which further broadened the definition of "combatant," while attempting to maintain mechanisms

to ensure that those protections were not abused by terrorists. Interestingly, the United States signed this protocol, but after a decade of consideration opted not to ratify it on the grounds that it granted too much protection to terrorists.[70] That decision puts America in the company of only three other nations—Iran, South Africa, and Israel—that are not bound by the Protocol. The specific reasons for this rejection are telling and illuminate current policy.

When the Reagan Administration chose to reject the treaty in 1987, one of its major objections had to do with Article 1(4), which broadened the scope of qualifying conflicts to include "conflicts in which people are fighting against colonial domination and alien occupation and against racist regimes in the exercise of their right of self-determination."[71] As with the first Convention, historical context is important. According to George Aldrich, who represented the U.S. at the drafting meetings for the Protocol, the sponsors of the above paragraph had the wars in South Africa and Palestine in mind specifically when they authored it.[72] The Reagan Administration worried that granting protections to the PLO would be seen as legitimizing terrorist activities.

A second objection related to parts of Article 44, which removed the requirements for a distinctive sign and respect for the laws of war to qualify a group as a military organization. Aldrich tells us that the latter requirement was dropped precisely to avoid the tendency for nations to refuse an entire *group* POW status because a few of its individuals had committed war crimes.[73] Another part of the Article prohibits the commission of war crimes from depriving a combatant of his POW status. The U.S. was among the strongest supporters of this provision because of their experience during Vietnam, in which pilots were uniformly denied POW status because a few had hit civilian targets.[74] Again, the Reagan Administration feared that these relaxed requirements would make it easier for terrorist organizations to claim protection for criminal acts. On the other hand, it was the understanding of the delegation that by recognizing the impracticality of uniforms for irregulars and guaranteeing POW status to irregular groups more

broadly, they would create a greater incentive to comply with the Laws of War.[75] The Protocol does provide for criminal punishment of any violation of the familiar conditions for lawful combat, but this punishment cannot include deprivation of POW status.

Aldrich goes on to argue the various ways in which other provisions of the Protocol actually render its protections inapplicable to terrorists, and provides a point by point counter argument to the Administration's objections. The technical provisions of Protocol I are less important, however, than recognizing that once again the international community believed it could broadly guarantee POW status without improperly privileging criminals, and indeed they included protections from such abuse in the Protocol.

The theme that emerges from the drafting of Protocol I and the ultimate rejection of it by President Reagan is a fundamental difference in priorities between the United States and much of the rest of the international community. The drafting history should make it clear that the activities of terrorist groups do not in any way constitute a "new paradigm," but have been of concern to the international community as long as the Conventions have been in existence. Since the signing of Protocol I, however, the U.S. has clearly placed a greater emphasis on limiting the rights and protections available to criminal groups than broadening those available to combatants in general. Then, as now, the majority of signatory parties seem to have agreed that broadening the lawful irregular combatant category is not necessarily at odds with prosecuting criminals. However, the last half-century of U.S. policy has demonstrated that we are often unwilling to use the avenues provided by the Geneva Convention for prosecution, preferring instead to ignore them and pursue our own prerogative. This fundamental difference in understanding is precisely the cause of controversy surrounding the Guantanamo detainees today.

Ideologically, then, the position of the present Administration is consistent with a policy at least two decades old, and simply reflects a long-standing difference of priorities. In practice, however, this difference has seldom had a chance to express itself so sharply, and in this regard, perhaps the Global War on Terror is something new.

Conclusions

Today, the United States Government clearly places a higher priority on the capture and prosecution of terrorists than it does on the individual liberties of warriors. As the source of so many controversies, this condition certainly need not sit well with every reader, but neither should it appear an original creation of the Bush Administration. The history of the 1977 Protocol and the Reagan Administration's decision not to ratify it betray a pronounced departure from the prevailing philosophy in the international community that is at least two decades old. Similarly, the arbitrary detention of civilians, incommunicado imprisonment of certain individuals, and trial of unlawful combatants as terrorists are not without precedent in post-Geneva history. As such, the current situation in Guantanamo does not constitute a novel abuse of power by the Executive, but in many respects is a logical extension of U.S. security policy over the last several decades.

Consistency with past policy, however, is not a substitute for sound rationale. The refusal to designate the Taliban warriors as POWs is plainly at odds with the attitudes and laws of the rest of the world concerning the rights of combatants: international law on the subject has continually evolved in the direction of broader protections, even in cases of doubt. Had the U.S. chosen to ratify the 1977 Protocol, which some scholars argue is approaching the status of customary international law,[76] there would be no question that the Taliban should be designated as POWs. The fact that America subscribes to an out of date treaty is no excuse in the eyes of many.

Furthermore, despite the fact that the *de facto* results of the President's decision are not unprecedented (incommunicado detention, etc.), the formal questioning of the applicability of the Geneva Conventions to an entire group is absolutely novel. In the past, exceptions have been made and rules broken, but only on an individual basis. The idea that one person (The Commander in Chief) may broadly determine the POW

Ryan Kingsbury

status of an entire group, such as al-Qaeda, is demonstrably against the intent of the Conventions and the body of more recent humanitarian law. As such, this view counts among the most critical misinterpretations of the U.S. legal obligations under Geneva; it will undoubtedly continue to draw criticism from abroad and could undermine the legal cooperation other nations are willing to afford us.

There is no doubt that the Global War on Terrorism is markedly different from any past conflict, as all wars are. It demands an increased emphasis on state security that is vital to the safety of America's citizens. But the superficial novelty of this war should not prevent us from recognizing that many of the circumstances and challenges it presents were understood and provided for by the drafters of the Geneva Conventions, and have even been managed in our own recent history through respect for the treaty. As such, we must resist dismissing our treaty obligations merely because they are old, and trust that the intent and evolution of international humanitarian law can and will provide the best protection against the terrorist threat.

Bibliography

Alrich, George. "Prospects for a United States Ratification of Additional Protocol I to the 1949 Geneva Conventions." *American Journal of International Law* 58, no. 1 (1991): 1-20.

Annex to Convention (IV) respecting the Laws and Customs of War on Land: Regulations concerning the Laws and Customs of War on Land. The Hague, 18 October 1907.

Ashcroft, John. Memo to The President. Feb 1, 2002, in Danner, Mark. *Torture and Truth: America, Abu Ghraib, and the War on Terror*, pp. 92-93, NYRB, Inc. 2004.

Berke, Richard. "Calling Himself a War Prisoner, Noriega Says U.S. Can't Try Him." *New York Times,* 27 January 1990, p. 9.

Bush, George. Memorandum "Humane Treatment of al-Qaeda detainees." Feb 7, 2002, in Danner, Mark. *Torture and Truth: America, Abu Ghraib, and the War on Terror*, pp. 105-6, NYRB, Inc. 2004.

Bybee, Jay S. Memorandum for Alberto Gonzalez, Counsel to the President. Feb 7, 2002, in Danner, Mark. *Torture and Truth: America, Abu Ghraib, and the War on Terror*, pp. 96-104, NYRB, Inc. 2004.

Casella, Alexander. "The Politics Of Prisoners of War:The politics of P.O.W.'s." *New York Times*, 28 May 1972, p.SM9.

Cushman, John. "New Focus Is on Status Of the 3 As P.O.W.'s." *New York Times*, 5 April 1996, p. A6.

Danner, Mark. *Torture and Truth: America, Abu Ghraib, and the War on Terror.* New York: NYRB, 2004.

de Preux, Jean. *The Geneva Convention Relative to the Treatment of Prisoners of War Commentary.* Edited by Jean S. Pictet. Geneva: International Committee of the Red Cross, 1960.

"Denied: A Shield for Terrorists." *New York Times*, 17 February 1987.

Emerson, Gloria. "48 American Civilians in Vietnam Charge U.S. Violates Treaties." *New York Times*, 29 December 1970, p.12.

Geneva Convention Relative to the Treatment of Prisoners of War, 27 July 1929.

"Geneva Convention On Prisoners of War." *New York Times*, 30 December 1983., p. A8.

"Goal of the Geneva Treaties: Helping the Victims of War." *New York Times*, 27 January 1990, p.9.

Goldman, Francisco. "In a Terrorized Country." *New York Times*, 17 April 1995, p.A17.

Gonzales, Alberto R. Memo to The President. Jan.25, 2002, in Danner, Mark. *Torture and Truth: America, Abu Ghraib, and the War on Terror*, pp. 83-87, NYRB, Inc. 2004.

"Judge Expresses Doubt On Status of Noriega." *New York Times*, 14 November 1992, p.6.

Lewis, Anthony. "The American Way." *New York Times*, 17 November 1983, p. A27.

Lewis, Neil. "Noriega May Be Prisoner of War But Can Still Face Trial, U.S. Says." *New York Times,* 3 February 1990.

Mallison, Sally and Thomas. "The Juridicial Status of Irregular Combatants under the International

Humanitarian Law of Armed Conflict." *Case Western Reserve University Journal of International Law* 9 (1977): 39-78.

Margolick, David. "Judge Rules Noriega Is Prisoner of War." *New York Times*, 9 December 1992.

Margolick, David. "Noriega: Public Enemy Becomes Prisoner of War." *New York Times*, 13 Decembe 1992.

May, Clifford. "Terrorist or Insurgent? Diplomacy and the Law." *New York Times*, 27 May 1988, p. B5.

Memorandum from Deputy Assistant Attorney General John Yoo to Department of Defense General Counsel William J. Haynes II (January 9, 2002) ("Application of Treaties and Laws to al Qaeda and Taliban detainees.")

Myers, Steven Lee. "U.S. Weighs Repatriating 2 P.O.W.'s To Yugoslavia." *New York Times*, 5 May 1999.

Naqvi, Yasmin. "Doubtful Prisoner of War Status." *International Review of the Red Cross* 84, no. 847 (2002): 571-595.

"Operation Just Cause." *Global Security.org. <http://www.globalsecurity.org/military/ops/just_cause.htm>* (15 March 2005).

Powell, Colin L. Memo for Counsel to The President. Jan.26, 2002, in Danner, Mark. *Torture and Truth: America, Abu Ghraib, and the War on Terror*, pp. 88-91, NYRB, Inc. 2004.

Presidential Military Order of November 13, 2001, 66 C.F.R. §222 (2001), reprinted in Danner, Mark. *Torture and Truth: America, Abu Ghraib, and the War on Terror*, pp. 78-82, NYRB, Inc. 2004.

Prial, Frank J. "Report Cites Campaign of Terror Since Soviet Entered Afghanistan." *New York Times*, 17 December, 1984.

Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts, 8 June 1977.

Rohter, Larry. "Legal Vacuum in Haiti Is Testing U.S. Policy." *New York Times*, 4 November 1994, p.A32.

Rohter, Larry. "More Than an Ex-Dictator's Future At Stake as Trial of Noriega Begins." *New York Times*, 5 September 1991.

"Significant dates in the history of international humanitarian law and the Red Cross and Red Crescent Movement." *Red Cross.* 12 January 2004. *<http://www.icrc.org/web/eng/siteeng0.nsf/iwpList304/7A44BA8D2CA C0CD2C1256B6 600599A60>* (10 March 2005).

"States Parties and Signatories." *Red Cross. <http://www.icrc.org/ihl.nsf/WebNORM?OpenView& Start=52.1.97&Count=150&Expand=52.2#52.2>* (18 April 2005).

"Syria Not Obligated To Free U.S. Airman." *New York Times,* 6 December 1983.

Taft, William H. Memo for Counsel to The President. Feb 2, 2002, in Danner, Mark. *Torture and Truth: America, Abu Ghraib, and the War on Terror*, pp. 94-95, NYRB, Inc. 2004.

Taylor, Stuart Jr. " Treatment of Prisoners Is Defended." *New York* Times, 29 October 1983, p. 7.

Taylor, Stuart Jr. "U.S. CLAIMS A RIGHT TO HOLD CIVILIANS:State Dept. Also Says Soldiers Can Be legally 'Detained'." *New York Times*, 4 November 1983, p. A17.

"Text of Resolution Adopted in U.N. on War Prisoners." *New York Times*, 2 December 1970, p.18.

"Texts of U.S. Resolution and Draft Convention on Terrorism." *New York Times*, 26 September 1972.

"The History of Humanitarian Law." *Red Cross. <http://www.redcross.lv/en/conventions.htm>* (15 April 2005).

"Treaties in Force." *U.S. Department of State.* 1 January 2000. *<http://www.state.gov/www/global/legal_affairs/tifindex.html>* (19 March 2005).

"U.N. Speaks Up on Prisoners." *New York Times*, 3 December 1970, p.46.

Wedgwood, Ruth. "Al Qaeda, Military Commissions, and Self Defense." *Political Science Quarterly* 117, no. 3 (2002): 357-372.

Endnotes

1. The original research for this article was completed for a term paper in the Plan II seminar "Legal Perspectives on the War on Terror" during the spring of 2005.
2. Gonzales, Alberto R. Memo to The President. Jan.25, 2002, in Danner, Mark. *Torture and Truth: America, Abu Ghraib, and the War on Terror*, pp. 78-82, NYRB, Inc. 2004.

Ryan Kingsbury

3. For a primary-source description of the Executive debate, see the series of memos in Danner, p.78-107

4. Danner, p.78-107.

5. Gonzales, Alberto R. Memo to The President. Jan.25, 2002, in Danner, 78-82.

6. Jean de Preux, *The Geneva Convention Relative to the Treatment of Prisoners of War Commentary*, ed. Jean S. Pictet (Geneva: International Committee of the Red Cross, 1960), 30-38.

7. de Preux, 30-38.

8. "The History of Humanitarian Law," *Red Cross, <http://www.redcross.lv/en/conventions.htm>* (15 April 2005).

9. "History of Humantarian Law," *<http://www.redcross.lv/en/conventions.htm>*.

10. Annex to Convention (IV) respecting the Laws and Customs of War on Land: Regulations concerning the Laws and Customs of War on Land. The Hague, 18 October 1907. Chapter II.

11. de Preux, 1-6.

12. de Preux, 1-6.

13. de Preux, 6.

14. de Preux, 6.

15. Geneva Convention [III] Relative to the Treatment of Prisoners of War, Aug. 12, 1949, Art. 1, *6 UST 3316, 3328,* 75 UNTS 135, 150.

16. de Preux, 50.

17. de Preux, 60-2.

18 Geneva Convention [III] Relative to the Treatment of Prisoners of War, Aug. 12, 1949, Art. 4, *6 UST 3316, 3328,* 75 UNTS 135, 150.

19. Compare Annex to Hague Convention IV of 18 October 1907, Art. 1, 2, and 3; Geneva Convention Relative to the Treatment of Prisoners of War of 1929, Art. 1; Geneva Convention Relative to the Treatment of Prisoners of War of 1949, Art. 4.

20. Geneva Convention [III] Relative to the Treatment of Prisoners of War, Aug. 12, 1949, Art. 4, *6 UST 3316, 3328,* 75 UNTS 135, 150.

21. Military Order of November 13, 2001 in Danner, Mark. *Torture and Truth: America, Abu Ghraib, and the War on Terror*, pp. 78-82, NYRB, Inc. 2004.

22. Military Order of November 13, 2001 at 1(g). in Danner, p. 79.

23. Military Order of November 13, 2001 at 2(b)-(d). in Danner, p. 80.

24. Memorandum from Deputy Assistant Attorney General John Yoo to Department of Defense General Counsel William J. Haynes II (January 9, 2002) ("Application of Treaties and Laws to al Qaeda and Taliban detainees.") [handout from Professor Amon Burton]

25. Reading the Yoo memo in connection with the series of memoranda in Danner, p. 78-107 easily demonstrates the extent to which the Yoo memo provides a laundry list of ways to argue that Geneva does not apply. The later memos build upon this list.

26 Geneva Convention [III] Relative to the Treatment of Prisoners of War, Aug. 12, 1949, Art. 2, *6 UST 3316, 3328,* 75 UNTS 135, 150. The article provides for the convention to be applied "in all cases of declared war or any other armed conflict which may arise between two or more of the High Contracting parties," "all cases of. . . occupation of the territory of a High Contracting party," and in any conflict with a non-signing nation, so long as that nation accepts and applies the provisions of the Convention.

27. Yoo, p.11

28. Geneva Convention [III] Relative to the Treatment of Prisoners of War, Aug. 12, 1949, Art. 3, *6 UST 3316, 3328,* 75 UNTS 135, 150.

29. Yoo, 12.

30. Yoo, 13.

31. Yoo, 9.

32. Yoo, 10.

33. Yoo, 16-22.

34. Yoo, 25.

35. Yoo, 25.

36. Series of memos in Danner, p. 78-107

37. Memoranda from Gonzalez, Ashcroft, and Bybee in Danner p.78-87, 92-3, 96-104.

38. Powell, Colin. Memo to to the Counsel to the President. Jan 26, 2002, in Danner, p. 91.

39. Powell, 88-91.

40. Taft, William. Memo to the Counsel to the President. Feb 2, 2002, in Danner, p. 94-5.

41. Military Order of February 7, 2002, in Danner, Mark, p. 105-6.

42. de Preux, 23.

43. de Preux, 23.

44. de Preux, 50.

45. de Preux, 36.

46. Geneva Convention [III] Relative to the Treatment of Prisoners of War, Aug. 12, 1949, Art. 1, *6 UST 3316, 3328,* 75 UNTS 135, 150.

47. Thomas and Sally Mallison, "The Juridical Status of Irregular Combatants under the International Humanitarian Law of Armed Conflict," *Case Western Reserve University Journal of International Law* 9 (1977): p.39-78.

48. de Preux, 33.

49. de Preux, 32.

50. de Preux, 35.

51. These extensive criteria are listed in de Preux, 36.

52. de Preux, 32.

53. Geneva Convention [III] Relative to the Treatment of Prisoners of War, Aug. 12, 1949, Art. 5(1), *6 UST 3316, 3328,* 75 UNTS 135, 150.

54. Bybee, Jay S. Memo to the Counsel to the President. Feb 7, 2002, in Danner, p. 97.

55. de Preux, 72.

56. Yasmin Naqvi, "Doubtful Prisoner of War Status," *International Review of the Red Cross* 84 no. 847 (2002): p. 586.

57. Naqvi, 586-7.

58. Alexander Casella, "The Politics of POWs," *New York Times*, 28 May 1972, SM9.

59. Casella.

60. Naqvi, 587.

61. Casella.

62. Anthony Lewis, "The American Way," *New York Times*, 17 November 1983, p. A27.

63. Stuart Taylor Jr., "U.S. Claims a Right to Hold Civilians," *New York Times*, 4 November 1983, p. A17.

64. Stuart Taylor Jr., "Treatment of Prisoners Is Defended," *New York Times*, 29 October 1983, p. 7.

65. Richard Berke, "Calling Himself a War Prisoner, Noriega Says U.S. Can't Try Him," *New York Times,* 27 January 1990, p. 9.

66. Neil Lewis, "Noriega May Be Prisoner of War But Can Still Face Trial, U.S. Says," *New York Times,* 3 February 1990.

67. De Preux, 40.

68. Berke.

69. Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), 8 June 1977. Art. 45(1)

70. "Denied: A Shield for Terrorists," *New York Times*, 17 February 1987.

71. Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), 8 June 1977. Art. 1(4)

72. George Aldrich, "Prospects for a United States Ratification of Additional Protocol I to the 1949 Geneva Conventions," *American Journal of International Law* 58, no. 1 (1991): p. 6.

73. Aldrich, 8.

74. Aldrich, 8.

75. Aldrich, 8.

76. Naqvi, 591-4.

Communal Care of Offspring in the White-faced Capuchin Monkey (Cebus capucinus)

Kristin A. Hook, College of Natural Sciences

Abstract

The purpose of this study was to document whether a communal infant care system existed in the white-faced capuchin monkey, *Cebus capucinus* (Cebidae), in the Monteverde area of Costa Rica. The intention was to compare time allocations spent between infants and other individuals in the troops followed. In addition, approach and leave times in interactions involving the infant and other members of the troop as well as the sexes of those individuals involved were documented. It was found that a communal infant care system does in fact seem to exist for this species of primate. Infants spent the majority of their time on the backs of a primary female in addition to the backs of juveniles. The majority of approaches made were to engage in social interaction. The

majority of approaches made by individuals other than the primary female were either solicitations or offers for rides. There was no difference in numbers of approaches initiated by female or male individuals; however, females spent more time involved with the infant during their interactions than did the males.

Introduction

Alloparenting, defined as the care of offspring by non-parental individuals, has been reported in over 30 mammal species (Goldizen 1986). Although literature on this phenomenon is scarce for white-faced capuchin monkeys, *Cebus capucinus*, this form of helping behavior is known to be common among the cebines, whereby adult females and juveniles of both sexes show great interest in infants from birth until about two months of age and will attempt to touch or pull an infant off its mother's back (Freese and Oppenheimer 1981). Alloparenting behaviors include carrying or keeping watch over infants, active or passive sharing of food, and even nursing, the latter of which exists only for *Cebus* (Robinson 1987). This study attempts to offset the paucity of data on communal care of infants in wild capuchins, a system that probably subsists because of their active nature and seasonal movement to exploit fruiting trees (Robinson 1987). There have been no studies published on the time budget of infant interactions to date. A time budget is simply a measure of what an organism spends its time doing on any given day. This study hypothesized that the majority of the infant's time is spent with a primary female presumed to be its mother because the relationship with the most frequent affiliative interaction observed in all cebine species is that between a mother and her youngest offspring (Robinson 1987). Moreover, it was theorized that females would make more infant approaches than males and that time spent by females during the interaction would be longer based on the premise that they gain a delayed benefit of experience in parental care (Goldizen 1986). Lastly, helping behavior in males was assumed to be relatively minimal in comparison to the females in any given troop because of the significant loss of foraging or hunting time that would be incurred (Goldizen 1986).

Kristin A. Hook

Methods

This study was conducted in Monteverde, Costa Rica, from April 13 through May 5, 2005, in the latter part of the dry season. The study site was in the area of El Bajo del Tigre, which at 1300m comprises 29 hectares of protected premontane wet forest, as well as adjacent areas on and around the properties of Frank Joyce, Susie Newswanger and Bob Law (Barnert 1999). The site was chosen based on previous studies done in the area and past predicted travel patterns based on repetitive sightings of these organisms (Filipkowski 1998). Common sightings were made on the Sendero de Los Murciélagos, Sendero Arboretum, and Sendero de Los Monos (Fig. 1).

The capuchin troops were located by arriving at El Bajo del Tigre at eight in the morning, where the information center employees and volunteers were questioned for news on their latest whereabouts. The majority of the search occurred on the three aforementioned trails. The behavioral method of continuous focal animal sampling was used due to the nature of the hypotheses that emphasized individual behaviors. Once a troop was found, time and date were recorded and a pair of binoculars was used to visually scan for the infant's location. Infant was defined based on relatively tiny size and frequent presence on the back of another individual. Its age and sex could not be determined. Observed infant locations included the back of a primary female, free and alone, or on another individual's back. Once the infant was found, data collection of its interactions and time budget began and continued until the troop was out of possible following range. Behaviors cited include approaches and leaves made by all individuals within the troop and inferred purpose of the interactions (see APPENDIX 1 for types and definitions). Approaches were defined based on close proximity judged to be less than one meter away. Leaves were defined based on leaving the one-meter proximity area of the infant. The individual who initiated the approaches and leaves, either the primary female, infant, or the individual involved in the approach, were recorded as well.

Figure 1 | *Map of the trails where the capuchins were located.*
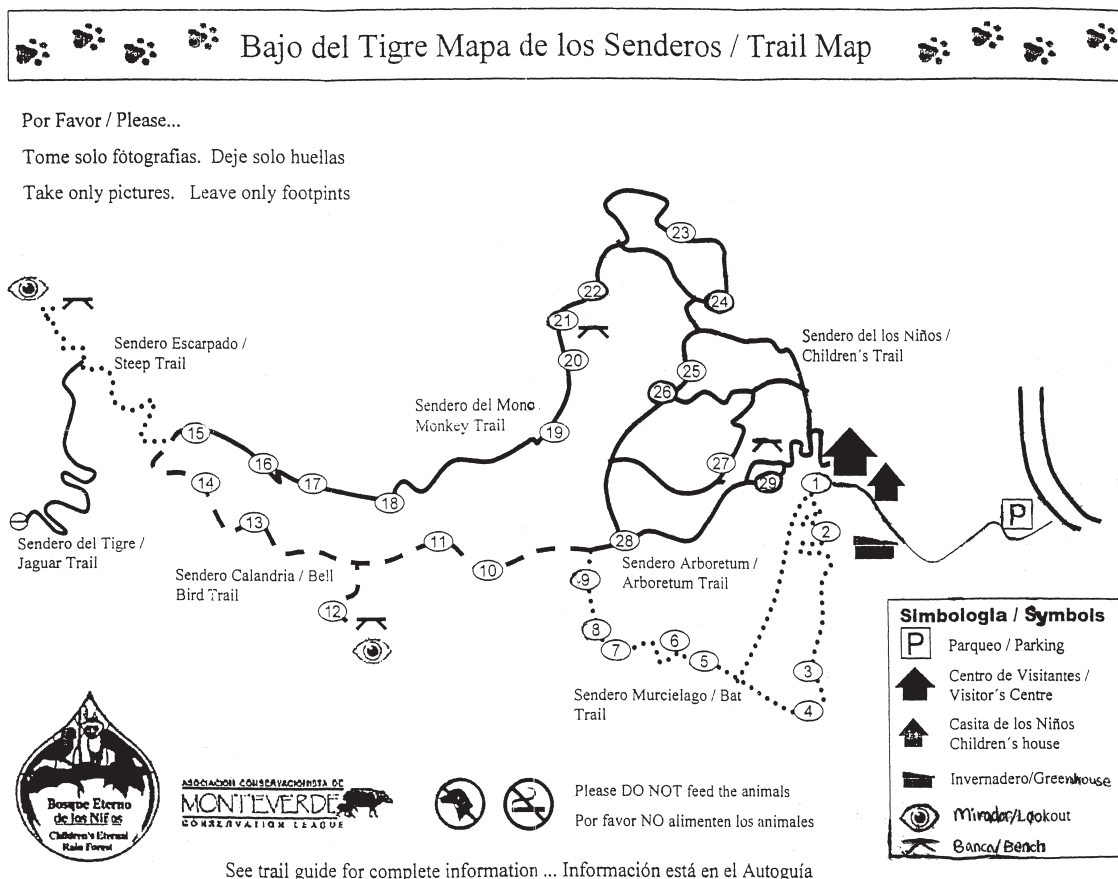
Finally, the number of individuals within the troop was estimated and recorded. Over time it was easier to distinguish between the two troops observed based on the composition and ratio of juveniles to adults. Recorded data was then broken down into percentages of how often certain behaviors occurred to get an overall picture of the communal infant care system and the individual interactions between the infant, primary female, and other troop members. Adults were identified as either male or female, and sex was disregarded in juveniles due to the small sample size. Juveniles were defined based on smaller body size, relative frequency of play they were engaged in, and a youthful appearance based on fewer face wrinkles in comparison to the adults. A Chi-square test was used to determine similarity in comparing male, female, and juvenile interactions. While observing *C. capucinus* for

indications of a communal infant care system, notes were taken on other miscellaneous behaviors within the two troops studied (see APPENDIX 2).

Results

Observations were made on eleven separate contact days, all of which occurred within El Bajo del Tigre and the surrounding properties. In over 80 hours spent in the field, approximately thirteen hours were spent in contact. Only 183 minutes of these were actually spent observing and collecting data on the infants. The monkeys observed on the first nine days were believed to be of the same troop; however, it is highly probable that the last two days in the field were spent following a different troop with an overlapping range. The best estimate of the number of individuals in the first troop was sixteen, which

Kristin A. Hook

includes the two infants followed and observed. The best estimate of the number of individuals in the second troop was eighteen, which also includes the two infants followed and observed. The sex ratio of the troop as a whole, as opposed to the individuals of focus, was hard to determine due to their inconspicuous sexual organs and fast pace behind dense foliage. Ratio of juveniles to adults was nearly even for the first group, while the second group seemed to have a higher ratio of juveniles to adults (7 adults, 9 juveniles, 2 infants).

This field study involved unmarked individuals and no distinctions were made as to the identity of the individuals due to the density of foliage and constant movement of the troop. Hence, data was collected per individual infant per day, even though there were only two individual infants observed from one day to the next. Hence, the total sample size was seventeen individual infant observations.

Over the course of this study, the infants were on the backs of a primary female 54% of the time and were alone or free from another individual approximately 18% of the observed time (Fig. 2). The infants were on the backs of another individual 28% of the time (Fig. 2), with 60% of those interactions being on the backs of juveniles (Fig. 3). Individual males and females exhibited no difference in the number of approaches made (Fig. 4), however, females interacted with the infant for a longer period of time than did the males (Fig. 5). Juveniles made approximately 25% of approaches that males and females did (Fig. 4) with considerably less time spent with the infant (Fig. 5). Primary females approached males and females in even numbers (Fig. 6), but interaction time was considerably longer for the males at 76% of the time (Fig. 7).

Individuals other than the primary female initiated half of the approaches made to the infant (Fig. 8), and the majority of approaches made were to engage in a nuzzle (APPENDIX 1) with the primary female and infant (Fig. 9). Another interaction behavior commonly observed was when an individual approached the infant in order to carry it or when the infant approached another individual in order to solicit a ride (APPENDIX 1). Leaves were initiated by the primary female 59% of the time, the other individual 37% of the time, the other individual 37% of the time, and the infant 4% of the time (Fig. 10).

A Chi-square test was used to determine significance of differences in male, female, and juvenile

| Figure 2 | Infant Interactions as Time Percentage |
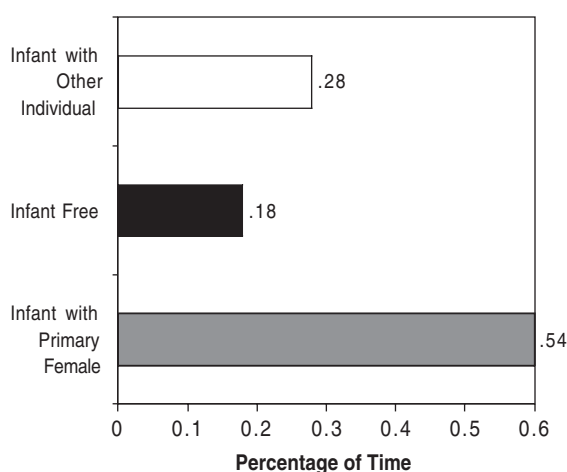
Chart of infants' time distribution spent with individuals within troop. Infants were observed to be either on the back of a primary female, alone and free from others, or on the back of an individual other than the primary female.
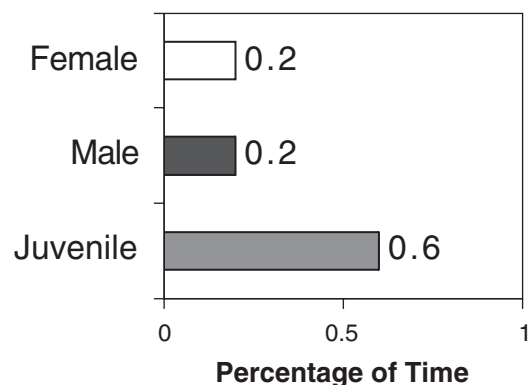
Kristin A. Hook

| Figure 3 | Individuals Involved in Infant Interactions as Time Percentage |

Chart of distribution of and the infant 4% of the time (Fig. 10). time percentage among individuals involved in interactions with infants.

Figure 4 | *Approaches Made By Individuals*

Number of Approaches Made

Legend: ☐ Male, ■ Female, ▨ Juvenile
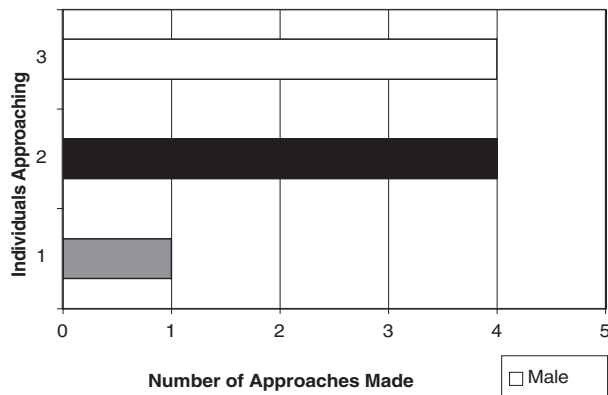
Number of approaches made by individuals other than the primary female. Categorized by adult males, adult females, and juveniles.

Figure 5 | *Individual Initiated Interactions*

Percentage of Time Approaches Made

Legend: ☐ Male, ■ Female, ▨ Juvenile
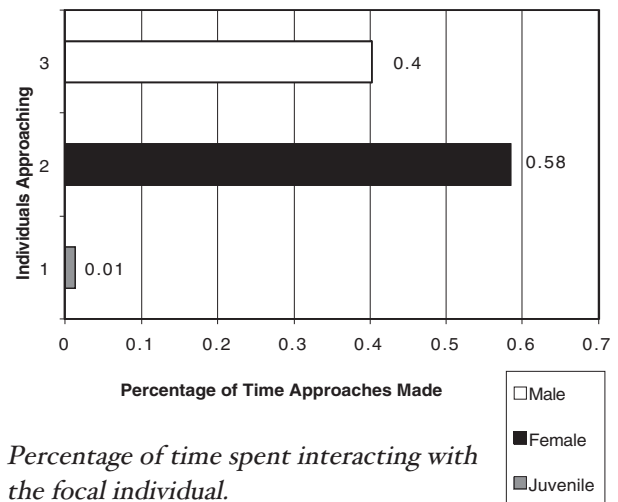
Percentage of time spent interacting with the focal individual.

Figure 6 | *Approaches Made by Primary Female*

Number of Approaches Made

Legend: ☐ Female, ■ Male
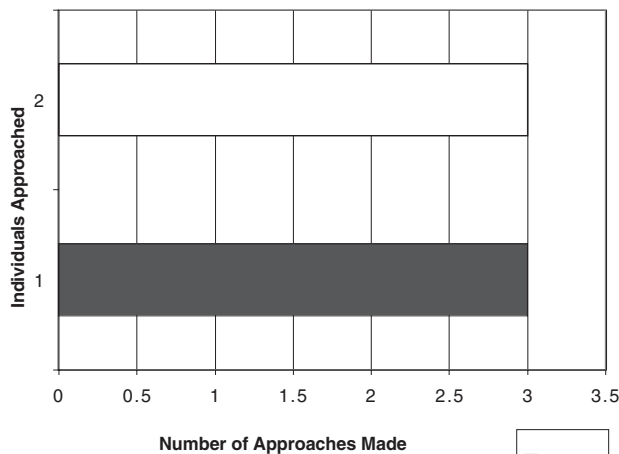
Approaches initiated by the primary female with infant to other individuals. Primary females did not approach juveniles.

Figure 7 | *Primary Female-initiated Interactions*

Percentage of Time Approaches Made

Legend: ☐ Female, ■ Male
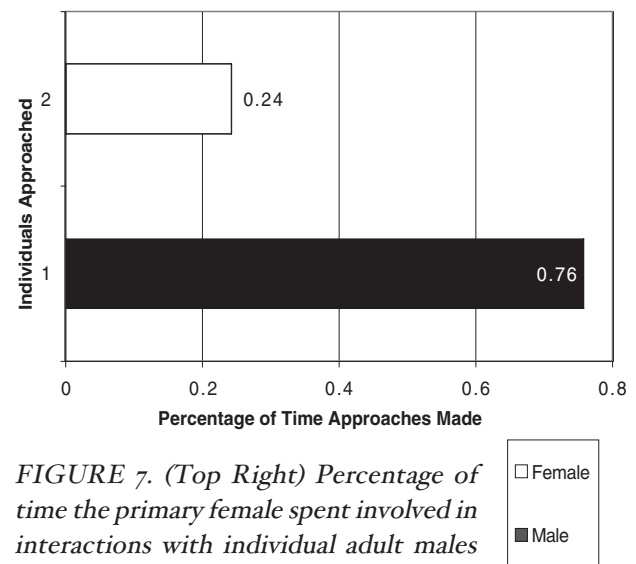
FIGURE 7. (Top Right) Percentage of time the primary female spent involved in interactions with individual adult males and adult females.

interactions with infants, using the data recorded in Figure 5. The results showed there was a significant difference among the percentages of approach times allocated by these individuals (df = 2, p < 0.05) as well as between varying percentages of time for approaches made by the primary female, as recorded in Figure 7 (df = 2, p < 0.05).

Discussion

The purpose of this study was to determine whether a communal infant care system exists for this species of primate. Many observations of individual interactions which may be construed as helping behaviors were recorded (APPENDIX 1). In general, many individuals within the troop expended a lot of

Kristin A. Hook

| **Figure 8** | *Distribution of Initiation* |
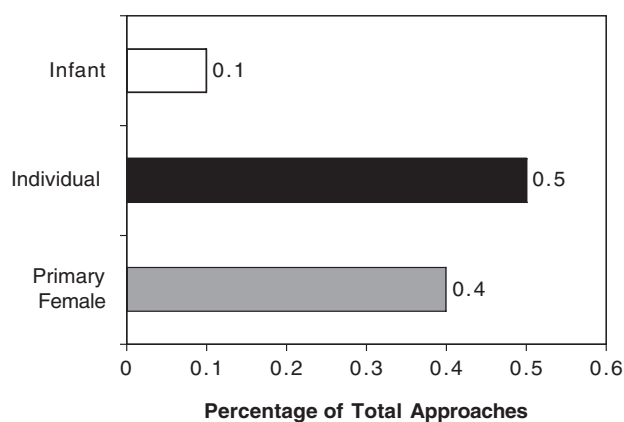
Distribution of total approach initiations. Individuals categorized by primary female, individual, and infant.

| **Figure 10** | *Distribution of Initiation of Leaves* |
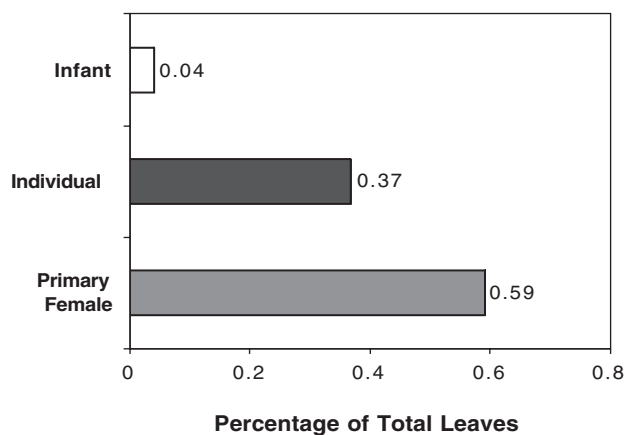
Chart of percentage distribution of leaves initiated by either primary female, individual other than the primary female, or the infant.

energy involved in interactions with the infants. Using approaches and interaction with the infant as an indicator of such a system, the data supports that one does in fact exist.

Moreover, the data supports that the majority of the infant's time is spent with a primary female, as hypothesized. Although maternity of the infants was impossible to know due to lack of DNA testing, the

Kristin A. Hook

| **Figure 9** | *Approach Types and Relative Percentages* |
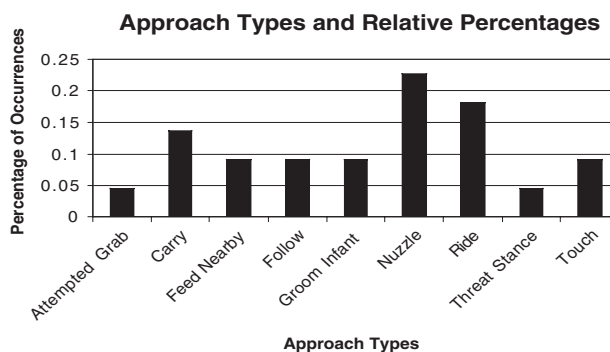
Approach types and percentage of occurrences between individuals and infants within the troops.

primary female is assumed to be the mother of the focal individual, since *Cebus* mothers are known to have a frequent, affiliative relationship with their young (Robinson 1987). Although nursing behavior was observed several times during the nuzzle social interaction (FIG. 8), nursing is known to be a form of helping behavior in cebines and, thus, can not infer maternity of the infant (Robinson 1987). The fact that the infants spent the majority of their time with a primary caretaker demonstrates that they may be physically dependent on one individual to provide them with most of their biological needs. Therefore, it seems logical that female capuchins only produce singletons at one to two year intervals because the young are so dependent until they reach five to six months of age when they achieve locomotory independence (Freese 1983). Overall, assuming that the primary female is in fact the biological mother, such a system probably subsists in order to lighten the strain placed on mothers due to their active nature and seasonal foraging movements.

Furthermore, the hypothesis that females would make more approaches to infants was rejected because there was no difference between number of approaches made by female and male individuals. However, the hypothesis that females spend considerably more time involved in infant interactions was supported by the data;

of the total time spent interacting with infants, females were involved 58% of the time while males were involved only 40% of the time (FIG. 5). Additionally, the fact that males approached infants the same number of times as females does not rule out the idea that males would have more of a time constraint due to significant loss of foraging and hunting time. In fact, this is possibly supported due to the fact that they spent considerably less time involved in the interaction with the infant than did the females. Although female individuals lose considerable foraging time as well, they are presumed to spend more time involved in infant care than are males due to the increased gain of experience in parental care (Goldizen 1986). A possible explanation for why males would invest some of their time to helping in infant care is to enhance their repertoires in order to qualify as next-in-line for alpha position. As the literature reveals, alpha male is the only one to achieve substantial mating success (Robinson 1987). Therefore, behaviors which accommodate this position benefit the individual by increasing their reproductive success. In order to fully assess the gains for capuchin males to assist in this communal infant care system, however, further research is necessary.

Future studies should incorporate DNA sampling and marked individuals as well as marked sexes of all individuals involved to get a better idea of the conspecific interactions that occur among members within a given capuchin troop. Such testing could also be important to help assess whether or not the primary female is in fact the biological mother of the focal individual. The Chi-square test results signified that approach times were not random for individuals and could in fact be dependent on sex or age of individuals approaching. Based on these results, it would be interesting to study whether or not the infants exhibit a preference for those individuals involved in the infant communal care system and whether or not this is dependent on the characteristics of sex or age.

Acknowledgements

Literature Cited

Barnert, E.S. and C.L. Hendricks. 1999. Home range, travel paths, and energy allocation of a troop of white-faced monkeys (*Cebus capucinus*) in Monteverde. UCEAP—IMV Monteverde Tropical Biology. Costa Rica: Monteverde Biological Station Library.

Filipkowski, A. and E.W. Olson. 1998. *Activity Patterns of the White-faced Monkey (Cebus capucinus)*. CIEE Symposium on Tropical Ecology and Conservation. Costa Rica: Monteverde Biological Station Library.

Freese, C. H. 1983. *Cebus capucinus* (Mono Cara Blanca, White-Faced Capuchin). In: *Costa Rican Natural History,* D. H. Janzen, ed. The University of Chicago Press, Chicago, IL, pp. 458-460.

Freese, C. H., and Oppenheimer, J. R. 1981. The capuchin monkeys, genus *Cebus. In* Ecology and behavior of Neotropical primates, vol. 1, ed. A. F. Coimbra-Filho and R. A. Mitermeier. Rio de Janeiro: Academia Brasileira de Ciencias. As cited in Robinson, J.G. and C.H. Janson. 1987. Capuchins, squirrel monkeys, and atelines: sociological convergence with old world primates. Cited in Smuts, B.B., et. Al. Primate Societies, (eds.) Chicago, 1987: University of Chicago Press.

Goldizen, A.W. 1986. Tamarins and marmosets: communal care of offspring. Cited in Smuts, B.B., et. al. Primate Societies, (eds.) Chicago, 1986: University of Chicago Press.

Robinson, J.G. and C.H. Janson. 1987. Capuchins, squirrel monkeys, and atelines: sociological convergence

Kristin A. Hook

with old world primates. Cited in Smuts, B.B., et. Al. Primate Societies, (eds.) Chicago, 1987: University of Chicago Press.

Terborgh, J. 1992. Diversity and Tropical Rainforest. New York: Scientific American Library.

Appendix

1. Interaction Types

Attempted Grab—the individual attempts to grab the infant from the primary female's back.

Carry—the individual carries the infant on its own back after an approach.

Feed Nearby—the individual feeds nearby the infant, or the primary female with the infant feeds nearby the individual, both within one meter proximity .

Follow—the individual follows the primary female with infant or the primary female with infant follows the individual, both within one meter proximity.

Groom Infant—the individual is observed grooming the infant, whereby the individual comes up behind and picks at its hair.

Nuzzle—the individual, along with the primary female and infant, lie down on a branch in a ball-shape form with the infant in between them or on their backs; infant is assumed to be nursing during this time because it is attached to primary female's stomach when she gets up from position.

Ride—the infant is on the back of an individual other than the primary female.

Threat Stance—the individual stands in front and the primary female with infant positions directly behind this individual with teeth exposed in a threatening stance to appear bigger in size and ward off a predator, in this instance a coati.

Touch—the individual is observed touching with the infant.

2. Miscellaneous Observations

The majority of the time individuals foraged and moved along the trails.

The time the troops foraged seemed to be random between 8am and 4pm.

Troops seemed to partition foraging space by splitting off into smaller groups.

Main activities were to rest, eat, travel, and engage in social interaction.

Troops were observed eating *Cecropia sp.* (Cecropiaceae), *Symplocos limoncillo* (Symplocaceae), and *Persea Americana* (Lauraceae) fruits along with random tree leaves and contents of Bromeliad tanks.

Quetzalcoatl: A Mestizo Myth

Sara Enright

Abstract

Quetzalcoatl is a Mesoamerican god that was known variously as the plumed serpent, the life-giver, the trickster, the god of the winds, or even as the king of the ancient city of Tula. A compelling figure, he was appropriated into the pantheon of gods of various conquering civilizations. The Spanish conquerors were no exception. Spanish priests became fascinated by the parallels between Quetzalcoatl's mythology and Christianity. Was he the devil, encouraging a parody to mock Christianity? Was he the Apostle Thomas, sent hundreds of years ago to evangelize the Americas? For some, Quetzalcoatl represented a Christian presence in the New World, far before the arrival of the Spaniards. In any case, because Quetzalcoatl's mythology has been passed through so many

hands, ultimately put into print by Spanish priests whose interpretations include an ethnocentric bias, the Quetzalcoatl we know today incorporates the worldviews of both the Old World and the New. His current mythology is a bridge between two diverse cultures, a symbol of Mexico's mestizo identity.

The Mesoamerican wind god, Quetzalcoatl, was introduced to the people of the Old World through a disastrous case of mistaken identity. During the conquest of the New World, Spanish texts document how the Aztecs mistook the Spanish conquistador Hernán Cortez for their long awaited god, Quetzalcoatl, an error that is commonly cited to explain why the Aztecs naively welcomed Cortez and his troops into the capital city of the Aztec empire. Though this version of the story may not necessarily be true, it does give Quetzalcoatl a distinctive role in conquest history, and marks the beginning of the Old World's fascination with the mythology of this New World deity. Catholic priests took special interest in Quetzalcoatl, in whom they saw unsettling echoes of Christianity; for hundreds of years after the conquest, they debated whether Quetzalcoatl was the devil or a Christian evangelizer, perhaps even the Apostle Thomas. Though these explanations may in retrospect appear to be bizarre misinterpretations of a non-Christian culture, what results is the formation of a new mythology for Quetzalcoatl, one that draws equally from Mesoamerican and Christian beliefs[1]. With each new interpretation, the Christian invaders provided Quetzalcoatl with an ever more complex mythology: the story of how the Old World incorporated itself into the New through religion, contributing to the redefinition of Mexico as a mestizo nation.

Pre-conquest Identity: Scattered Among the Ashes

There are very few resources available that document Quetzalcoatl's role in pre-conquest Mesoamerican civilization. By the time the Spaniards took an interest in the original culture and belief systems of the societies they conquered, they had already killed most of the indigenous priests and burned all but sixteen of their sacred books.[2] Instead, the Spanish scholars relied on oral histories, on reproduced manuscripts, and on the analysis of the remaining picture codices—so there was a lot of room for misinterpretation, ethnocentric misunderstandings, and manipulation of the text.[3]

It is not surprising, then, that the resulting texts that discuss Quetzalcoatl provide complicated and often contradictory information about him. He sometimes appears as a trickster god who rules over an unlucky calendar day; in others texts, he plays the role of the creator god, who forms mankind out of kernels of corn, and who instigates the practice of agriculture. Sometimes he is a man, and sometimes a god. Some texts portray him as a principal deity, and others hardly mention him. The complexity of the original Quetzalcoatl myth explains, in part, why the Spanish develop so many interpretations of Quetzalcoatl's role in the Americas. It also explains the Quetzalcoatl figure was so easily adapted to different personas.

Franciscan priest Bernadino de Sahagún was charged by his order in the 1560s with the task of recording indigenous culture so that the Franciscans could better convert the Indians. Because he wrote so soon after the conquest, and because he translated the mythology directly from oral histories of the survivors of the conquest, Aztec texts, and Mayan glyphs, Sahagún provides one of the most complete and accurate descriptions of the Pantheon of Mayan gods, including a complex interpretation of Quetzalcoatl.[4] In Sahagún's *A History of Ancient Mexico*, Quetzalcoatl is a human king who ruled over a prosperous and fertile age in Tula—the ancient capital city of the Toltecs. He is described as a man with fair skin and a long beard, a historic figure who continues to symbolize the legitimacy and power of Mesoamerican priests.[5] In *The Florentine Codex,* Quetzecoatl appears as a deity, the god of the wind: ". . . he was the wind; he was the guide, the roadsweeper of the rain gods, of the masters of the water, of those who brought rain."[6] In this, the most significant of Sahagún's writings, the feathered serpent is only one of many gods, and hardly the most important.

Sahagún represents Quetzalcoatl the god as a trickster. Babies born under the sign of the ceacatl (one

reed)—Quetzalcoatl's sign—lead a life of misfortune; everything they own will be carried off with the wind.[7] One disturbing anecdote tells about the special protection of burglars who are born under the sign of ceacatl; in the middle of the night, if they carry the left arm of a woman who died in her first childbirth to the house they will rob and pound the arm on the threshold, they can put all the occupants of the house in a trance to facilitate the robbery. Protected by the magic of Quetzalcoatl, they can then thieve in safety.[8]

In other interpretations of indigenous texts, Quetzalcoatl appears as a life-giver and culture bearer. In the *Leyenda de los Soles*, Quetzalcoatl goes on a dangerous journey to the underworld to collect human bones. He grinds the bones in a jade bowl and bleeds his penis over the powder, bringing the bones to life.[9] He revives the bodies by feeding them with masticated kernels of corn, which he stole from an anthill.[10] This symbolic offering of corn also represents the beginning of agricultural practices, with which Quetzalcoatl is credited.

Perhaps most fascinating to the Spaniards was the god's mysterious disappearance from the continent sometime in the 10th century. According to the *Leyenda de los Soles*, Quetzalcoatl left Tula under questionable circumstances. The other gods had become angry at Quetzalcoatl because he refused to allow human sacrifice in his kingdom. They tricked him into drinking too much, after which he had a sexual encounter with his sister. Upon sobering, he felt such anguish and embarrassment for his actions that he decided to exile himself to Tlapallan, a city in the Yucatan, leaving Tolan to ruin. Before leaving, he promised to return, so his followers continued to propagate his memory.[11]

By emphasizing different points in the story, one can interpret this exile myth in several manners: Quetzalcoatl, a god revered for his cleverness, is outtricked by other gods, symbolically representing his fall from power among Mesoamerican gods, but will return to reestablish his authority and to win back his honor; Quetzalcoatl—a righteous god—believes in the sanctity of life and in the value of repentance and self sacrifice for spiritual cleansing. Both of these interpretations play a role in the god's post-conquest identities.

Because the Spaniards destroyed so many of the original sources of Mesoamerican religious doctrine and mythology, this patchwork story is the most complete rendition of the Quetzalcoatl myth available to scholars today. Yet Quetzalcoatl continued to be an important figurehead in Mexico far after the conquest, albeit in a form that morphed along with the whims of the Spaniards.

Quetzalcoatl and the Conquest: The Return of the King

In book 12 of the *Florentine Codex*, Sahagün provides one of the few descriptions of Cortez's conquest of Tenochtitlan, in which he develops Quetzalcoatl's most popular myth: that Montezuma initially mistook Hernán Cortez, the Spanish conquistador, for Quetzalcoatl, returning from the Yucatan. According to Sahagün, there were a combination of many coincidences that explain this unusual misidentification: Cortez was a light skinned man with a beard, who came to the Americas on the year 1 Reed, which was the Aztec calendar date on which prophesies scheduled Quetzalcoatl's return. Beginning ten years before the arrival of the Spaniards, the Aztecs observed and reported strange phenomena such as the appearance of a meteor, the falling of comets, and the discovery of monsters, which they interpreted to be signs for the return of their god and of the end of their empire.[12] Sahagün was the first Spanish historian to record these omens.

Modern historians disagree about the legitimacy of this version of the story. While some believe that the Aztecs did, in fact, mistake Cortez for their returning god, others argue that this is one of the Spaniards' most elaborate myths. In either case, we begin to see how Quetzalcoatl's identity in post-conquest society becomes intimately tied with how the Spaniards chose to interpret this myth.

In his book *Quetzalcoatl and the Irony of Empire,* Davíd Carrasco traces Quetzalcoatl's role in the construction of Mesoamerican cities, and concludes that the feathered serpent—the mythological ruler of a golden age—represented a legitimizing power for Mesoamerican rulers and empires: "Quetzalcoatl was a

Sarah Enright

symbol of authority. . . in terms of the origin and sanctification of authority in capital cities." [13] When the Aztecs conquered the Toltecs, they stole Quetzalcoatl for themselves, thus legitimizing their power and dominance over Toltec society. According to Carrasco, the Aztecs felt insecure because they feared that their power was illegitimate. They became obsessed with fear that Quetzalcoatl, whose power they had seized, would return from the Yucatan to destroy their empire and take back his throne. If they were so self-conscious of their illegitimacy, it is not surprising that they were quick to conclude that their god had returned when Cortez arrived on Mexican shores. As support for this argument, Carrasco examines several instances where Montezuma, instead of defending his empire from invaders, appears instead to be preparing to hand over the empire to Cortez. Also, several original sources support the idea that Cortez was indeed mistaken for Quetzalcoatl. In 1536, Fray Toribio de Benavente, a Franciscan priest who adopted the Indian name Motolin'a, recorded from his indigenous informants that when the Indians saw the four Spanish ships appear off the coast, they believed that Quetzalcoatl was returning with his four temples across the sea.[14] Further, Hernán Cortez wrote in one of his letters to King Charles I in 1520 about how Montezuma received him in the name of their lord (though it is left unclear which lord he speaks of, Quetzalcoatl, the Christian god, or King Charles I):

> ". . . we know that a chieftainÉbrought our people to this region. And he returned to his native landÉbecause of the place from which you claim to come, namely from where the sun rises, and the things you tell us of the great lord and king who sent you here we believe and are certain that he is our natural lordÉSo be assured that we shall obey you and hold you as our lord in place of that great sovereign of whom you speak. . ."[15]

There is some suspicion that Cortez left the dialogue ambiguous in order to please his king and benefactor, so it is possible that Cortez was himself received as the Aztec's "natural lord," Quetzalcoatl.

Jacques Lafaye, author of *Quetzalcoatl and Guadelupe,* also believes in the Quetzalcoatl-Cortez myth. Lafaye argues that while the Spaniards believed that the misidentification was a clear sign of God's favor, the Indians found justification for their downfall within their own mythology.[16] Under this scenario, Quetzalcoatl legitimizes the beliefs of both the Spaniards and the Aztecs, giving each culture spiritual guidance on how to interact: the Spaniards as rulers and the Indians as subjects. This is reflected in the writings of Fray Diego Durán, a contemporary of Sahagün, who titled one of his chapters in *Historia de las Indias* "Of the Idol Called QuetzalcoatlÉWho Was the Father of the Toltecs, and of the Spaniards Because He Announced Their Coming." [17] Because of the coincidences of Cortez' arrival, as recorded and propagated by the texts of Sahagün, Quetzalcoatl would become intimately tied to the history of the Spaniards in Mexico.

Camila Townsend, author of the article "Burying the White Gods: New Perspectives on the Conquest of Mexico," disagrees with the Quetzalcoatl-Cortez myth. In her reanalysis of the story, the Aztecs were quite aware that the conquistadores were human, and that they had far superior war technology. They were not in awe of the Spaniards' godliness, but rather tried to defend themselves from their arrows and guns by keeping out of range, choosing to observe them from a distance. When Montezuma commanded that his people not attack the Spanish when they approached Tenochtitlan, it was not out of cowardice or awe, but rather out of concern for the wellbeing of his people, who were no match for the Spaniards: "If the Aztecs could not deliver a quick victory on the outskirts of their own capital, they were doomed; so if his army could not win quickly here—and Moctezuma knew they could not—then they could not fight."[18] Townsend portrays an Aztec empire that is only too aware of the advantages of its enemy, and not dumbstruck by fear and awe of their returning god.

Townsend deconstructs Sahagün's version of the conquest by using the very same sources that he had available at the time—she just emphasizes different parts of the text. For instance, she points out that, by law, Carlos V could not annex any lands that were not given to him voluntarily or won through a just war. So Cortez was pressured to provide testimony that Montezuma

handed over the land willingly, which may have compelled him to fabricate Montezuma's famous welcoming speech in his favor.[19]

It is possible that Townsend misinterprets the story by applying a western worldview to Mesoamerican actions; Montezuma could very well have acted according to the mandates of his religion rather than as a strategizing military commander when he received Cortez and his troops. But there seems to be just as much textual evidence for the mistaken identity story as there is against it. After all, the Aztecs did kill a few Spaniards in several battles, and had many opportunities to observe their humanity.[20] So why did the Spaniards propagate Sahagún's version of the conquest, where Cortez masquerades as a Mesoamerican god? Townsend suggests that the Spanish were satisfied by Sahagún's portrayal because they wanted to think of themselves as blameless, righteous victors: ". . . it perhaps comes as no surprise that the relatively powerful conquistadors and their cultural heirs should prefer to dwell on the Indian's adulation for them, rather than on their pain, rage, or attempted military defense."[21] It must have been very flattering to think that the peoples of the New World had been expecting the arrival of the Spaniards for many years, and that they were willing to give up their empire to the spiritually superior civilization. The Spaniards might also have been pleased by the idea that the natives were so na•ve and innocent that they mistook a man for a god, which seemed like a marvelous argument for why they needed Christian spiritual guidance.

As further evidence against the Quetzalcoatl-Cortez myth, Sahagún's version of the conquest may have been tainted with the biases of his indigenous sources. He relied on the descriptions provided by elder storytellers from Tlatelolco, who were descendents of a people conquered by the Aztecs. They might have wanted to reveal the Aztecs as cowards and fools out of old feelings of envy and resentment.[22] In fact, Sahagún's Book Twelve of the *Florentine Codex* is the first text that shows that Cortez was mistaken for the returning god (and in that, only twice), which suggests that this part of the story may easily have been altered after the conquest.[23]

If the Aztecs used Quetzalcoatl to legitimize their own dominance, as Carrasco argues, they were brought to their knees by the same deity.[24] The power of Quetzalcoatl's legitimacy then passed into the hands of the Spaniards. It was Sahagún's version of the conquest— the winner's version—that was adopted as historic truth.

Post-Conquest Interpretations: Be He Devil or Saint?

Regardless of whether or not the Aztecs viewed the conquest through a lens of religious fervor, the Spaniards certainly did. The conquistadors passionately believed that they played a significant role in the apocalypse.[25] They were advancing Christ's standard in a world that was quickly coming to an end. Before his third voyage to the Americas, Christopher Columbus commissioned and edited the writing of *The Book of the Prophesies*, which pulls together all of the Biblical texts that predict—or could be interpreted to predict—the discovery of the New World. The extended title of the book is: "Book or collection of authoritative writings, sayings, opinions, and prophecies concerning the need to recover the Holy City and Mount Zion, and the finding and conversion of the islands of the Indies and of all peoples and nations."[26] Many of the conquistadors and the priests that trailed Columbus's discovery interpreted the conquest of the New World as a harbinger to the second coming of Christ: "The Turkish victories in Europe, the decadence of the Roman church, the appearance of the false prophet Luther (regarded by some as the Antichrist), the spiritual harvest promised and, it seemed, reserved by providence for the Franciscan pioneers of the conversion of the New World, were so many converging signs of the approach of the millennium. . ."[27]

The first twelve Franciscan priests to land in the New World in 1524 believed that their mission was "the beginning of the last preaching of the gospel on the eve of the end of the world."[28] The Franciscans established the *Inglesia Indiana* in which they converted the Indians as quickly as possible—overlooking the fact that, while the Indians converted to Catholicism in the thousands, they often secretly continued to worship their old gods.[29] A few decades after the conquest, however,

Sarah Enright

millennial fever began to cool. With the passing of time, not all religious orders turned out to be as millennial as the Franciscans. Several important Jesuits, including Jos de Acosta, argued that there were still many lands left where people had not yet converted to Christianity, such as China, which meant that the conversion of the Americas, though important, was not the last push for the second coming of Christ.[30] When the Jesuits and the Dominicans came to the New World, they changed the focus of religious communities from mass conversion to maintenance and education of the indigenous population, trying to find more practical ways to incorporate them into Spanish colonies.[31] This shift in religious focus is also reflected in how the Spaniards interpret the Quetzalcoatl myth.

The Spaniards noticed many familiar Christian symbols and sacraments mixed within Indian idolatrous practices. For instance, the Aztecs practiced infant-bathing ceremonies that closely resembled baptism, and they also ate corn biscuits that were shaped in the form of their gods on certain religious holidays, calling to mind the Eucharist. On the other hand, the Aztecs also appalled the Spaniards with orgies, bloodletting rituals, and bloody human sacrifices.[32] Ethnocentric as they were, the Spaniards were unable to believe that the religious similarities could have been developed separately from Christian history, but they found it difficult to explain the bloody pagan rituals. They concluded that either God or the devil must have influenced the Indians before the Spaniards ever set foot on American soil.[33]

Quetzalcoatl had Christian parallels as well. Among a horde of bloodthirsty gods, he allegedly refused human sacrifice. Some texts assert that Quetzalcoatl was born of a virgin, that he initiated the practice of mortification, and that he taught that there was only one god who created all things. And, of course, he prophesized the conquest.[34] A question emerged among priests and scholars as to the origin of Quetzalcoatl: was he the devil, or was he connected to Christian doctrine in some other way?

For Sahagún and other early priests, the existence of human sacrifice in Mesoamerican tradition provided clear evidence that the indigenous peoples were pagan devil-worshippers, and that their souls needed to be saved:[35] "Sahagún and his companions interpreted Christianity as history; they had come to the Indies to fulfill the Word and thus achieve the prelude to the last days. . . In the eyes of the Twelve, they themselves were the true apostles of the Indians. . . everything in the Indian beliefs that had a distant resemblance to the true faith was a devilish parody. . ."[36] For these early Franciscan priests, the New World was the last battle ground between God and the devil. Each conversion aided God's struggle, and brought Spain greater glory.[37]

Sahagún was mainly interested in recording Mesoamerican beliefs in order to find ways to more thoroughly convert the Indians to Christianity. In the *Florentine Codex*, he argues that Quetzalcoatl was not a god, and that worshiping him is a sin:

> He was a common man; he was mortal. He died; his body corrupted. He is no god. And although a man of saintly life, who performed penances, he should not have been worshiped. What he did which was like miracles we know he did only through the command of the devil. . . His body died; here on earth it became dust, it became filth. And his soul our Lord God damned and thrust into the land of the dead. There it is. It will forever suffer in the flames.[38]

This is an example of how early Christian missionaries reframed New World myths and beliefs to favor the teachings of Christianity. Here, Quetzalcoatl is represented as a human who worshipped the devil—a figure worthy only of disdain.

However there were priests who had a difficult time believing that such a large portion of the human population had thrived thus far without having any knowledge of the existence of God and Christ. These priests referred to a verse of the Bible that talked about how Jesus' apostles were sent out to preach the Word throughout the world: "And this Gospel of the kingdom shall be preached in all the world for a witness unto all nations; and then shall the end come."[39] They believed that one of the original 12 Apostles must have come to the Americas after Christ's death fifteen centuries before the Spaniards arrived. St. Thomas, who went east to "the Indies," seemed to be the only Apostle that could have reached the New World. When priests looked

through Mesoamerican mythology for a figure that resembled a prophet, Quetzalcoatl stood out as the most likely candidate.[40]

Among the missionaries, the Jesuits in particular supported the Quetzalcoatl-St. Thomas version of the story. The Jesuits arrived to the New World in 1572. Unlike the other orders, they tended to focus on supporting the indigenous population and helping to establish a place for them in creole society rather than on their mass conversion.[41] Because a growing number of American-born Spaniards needed many laborers to work their land, they frequently forced indigenous people to do the labor. The Jesuits tried many ways to protect the indigenous peoples from being enslaved, and developing the Quetzalcoatl-St. Thomas myth was convenient for this objective. If indigenous people had already been evangelized, law dictated that they could not be used as slaves.

Lafaye describes how Fray Diego Durán, a contemporary of Sahagún, searched for evidence to support the Apostle of the Americas hypothesis in order to protect the Indians from enslavement: "The pro-Indian missionaries found in this Christian tradition of the New World a sign of God's grace to the Indies and their peoples and, consequently, as new metaphysical ammunition against the slave-hunters. . ."[42] Not surprisingly, Durán also was known for his strong sense of loyalty and admiration for the Americas, and the people in it. He believed that the Quetzalcoatl-St. Thomas myth validated the New World as valuable to God—and thus also deserving of special attention from the Church.[43]

The Apostle of the Americas gained popularity in the mid 1600s, when priests were searching for more proof to support the story about the appearance of Our Lady of Guadalupe, also known as Mary, the mother of Jesus Christ, to Juan Diego in 1532.[44] In 1662, there was an initiative to ask the Pope to recognize the growing cult of Our Lady of Guadalupe, and while a majority of priests focused on gathering details about the miracle of Juan Diego's vision of the Virgin Mary, some apparently believed that the appearance of an Apostle in the Americas would further validate the story.[45] Again, the Jesuits were especially interested in the Apostle. They wrote often about traces of St. Thomas in the New

World, which include a set of footprints he left in a rock in Brazil. They also argued that the many similarities between indigenous ceremonies and Christian ceremonies—including forms of baptism, fasting, and confession[46]—were vestiges of St. Thomas' evangelization rather than demonic parodies. Jesuits Manual Duarte and Sigüenza y Góngora wrote extensively on the subject in the late 1600s, continuing the work of Diego Durán.[47]

By the late 1600s, as a growing number of creoles grew to believe that God had bestowed his blessings on New Spain through the appearance of Our Lady of Guadalupe, who was a "divine gift to the creoles of Mexico City, who regarded themselves as a chosen people,"[48] writings about the Apostle of the Americas also reflected this growing sentiment of creole pride. Both the presence of St. Thomas and the Virgin Mary gave New World history and land greater significance in Christian terms. The appearance of St. Thomas was important because it showed that Christianity had a history in the New World prior to the conquest, which meant that the Americas had always been spiritually equal to Spain. Lafaye says, ". . . the creoles preferred Saint Thomas, who redeemed their American *patria* from the stigma of having lain in darkness for sixteen centuries, isolated from revelation."[49] Even so, the connections drawn between Quetzalcoatl and St. Thomas often sound fantastic and illogical, giving Quetzalcoatl a personality very distant from the trickster rain god of Mesoamerican legend, and while there is evidence of a growing interest in Our Lady of Guadalupe, it is hard to say how popular the Quetzalcoatl-St. Thomas myth actually was among the people of New Spain.

During the independence period in Mexico, the symbol of Our Lady of Guadalupe seems to eclipse the idea of Quetzalcoatl-St. Thomas. Her image is more directly Christian, more personable, and more accessible to the creoles of New Spain than the vague evidences provided for St. Thomas' evangelization of the Indians of the Americas, hundreds of years ago. Quetzalcoatl's Christian connection petered off, overwhelmingly obscured by the image of Our Lady of Guadalupe as proof of God's love for the people of the New World.[50]

Sarah Enright

However, while the plumed serpent has fallen out of favor within Christian doctrine, he continues to be a familiar figure in Mexican art and literature. While other Mesoamerican gods remain obscure, Quetzalcoatl resurges today as national symbol, this time representing the mystery and the glory of Mexico's Mesoamerican past.

Cloaked variously in the guise of a Toltec king, Cortez, the devil, and St. Thomas, Quetzalcoatl's original Mesoamerican identity lies in the ashes of the burnt books at Tezcoco. Though scholars lament the loss of so many valuable texts that were destroyed in the first years of the conquest, it is the ambiguity of the Quetzalcoatl myth that made it so attractive to Catholic historians and so malleable as one of the first national symbols: mixing past with present, Mesoamerican beliefs with Christian doctrine. If Catholic priests were responsible for the destruction of much of the wind god's pre-conquest identity, they were also responsible for his survival into modern times through their writings. Quetzalcoatl's current mythology originates neither from the New World nor the Old, but rather from a mixture of the two, making him a unique symbol of the mestizaje of Mexico. While the other gods of Mesoamerica are relegated to Mexico's ancient history, Quetzalcoatl rises like a phoenix—or like a plumed serpent—reborn with each new interpretation of his incomplete mythology.

Works Cited

Carrasco, Davíd. *Quetzalcoatl and the Irony of Empire: Myths and Prophesies in the Aztec Tradition.* Chicago: University of Chicago Press, 1982.

Cervantes, Fernando. *The Devil in the New World.* New Haven: Yale University Press, 1994.

Códice Chimalpopoca: Anales de Cuauhtitlán y Leyenda de los Soles. Translated Primo Feliciano Velázquez from Nahuatl. Mexico: Universidad Nacional Aut—noma de Mexico, 1992.

Columbus, Christopher. *The Book of the Prophesies, Volume III.* Translated Blair Sullivan. University of California Press: Los Angeles, 1997.

Lafaye, Jacques. *Quetzalcoatl and Guadalupe: The Formation of Mexican National Consciousness 1531-1813.* Translated by Benjamin Keen. Chicago: The University of Chicago Press, 1976.

León-Portilla, Miguel and Earl Shorris. *In the Language of Kings.* New York: W. W. Norton and Company, 2001.

Lockhart, James. *We People Here: Nahuatl Accounts of the Conquest of Mexico.* Berkley: University of California Press, 1993.

Phelan, John Leddu. *The Millennial Kingdom of the Franciscans in the New World.* Berkely: University of California Press, 1970.

Poole, Stafford. *Our Lady of Guadalupe: the Origins and Sources of a Mexican National Symbol 1531-1797.* London: The University of Arizona Press, 1995.

Ricard, Robert. *The Spiritual Conquest of Mexico: An Essay on the Apostolate and the Evangelizing Methods of the Mendicant Orders in New Spain: 1523-1572.* Translated by Lesley Bird Simpson. Berkeley and Los Angeles: University of California Press, 1966.

Sahagún, Bernardino de. *Florentine Codex.* Translated by Arthur J.O. Anderson and Charles E. Dibble. Santa Fe: The School of American Research and the University of Utah, 1970.

Sahagún, Bernardino de. *A History of Ancient Mexico.* Translated by Fanny Bandelier from the Spanish Version of Carlos Bustamente. Glorieta: The Rio Grande Press, 1976.

Townsend, Camilla. "Burying the White Gods: New Perspectives on the Conquest of Mexico." *The American Historical Review* 108 (June 2003): 659-687.

Endnotes

1. Mythology here means an oral or written religious story, not necessarily factual, that reveals the worldview and ideology of a people. In this case, because Quetzalcoatl's original mythology was interpreted and altered by Spaniards, it can be argued that the interpretations reflect the worldview of both Mesoamerican and Spanish societies.
2. Carrasco, Quetzalcoatl and the Irony of Empire (Chicago: University of Chicago Press, 1982), 13. Both the early priests and the conquistadores believed that the indigenous religions were heretical. While Juan de Zumárraga, the apostolic inquisitor of Mexico in 1535, ordered the destruction of all of the pictorial manuscripts in Tezcoco, Cortez and his troops defaced all of the religious monuments, idols, and images that they encountered on their way to Tenochtitlan. Later, European sponsors requested Indian scribes to reproduce what they could remember of their destroyed texts as "curios" from the New World.

Spanish priests also encouraged the recreation of indigenous religious texts so that they could better understand, and thus convert, the Indians. Though these reproduced texts are certainly less reliable than the originals, they are still valuable sources of information.

3. For more on this, see Carrasco 19-62.

4. Bernardino de Sahagún, *A History of Ancient Mexico*, trans. Fanny Bandelier (Glorieta: The Rio Grande Press, 1976) Sahagún's writings that are available in print today are not even completely accurate. His Franciscan superiors censored some of the things that he recorded about the conquest, and many of his other writings were conveniently "lost." See Jaoquin Icazbalceta's biography of Sahagún in *A History of Ancient Mexico*, pages 8-14.

5. Ibid., 179. Sahagún often wrote about the Indian gods in European terms, referring back to Greek mythology. One of the chapters in A *History of Ancient Mexico* compares Quetzalcoatl to Hercules, which shows Sahagún's ethnocentric bias: "Account of who Quetzalcoatl was, another Hercules, a great necromancer. . ."(179).

6. Bernardino de Sahagún, Florentine Codex, trans. Arthur J.O. Anderson and Charles E. Dibble (Santa Fe: The School of American Research and the University of Utah, 1970), 9.

7. Sahagún, *A History*, 68 and 218.

8. Sahagún, *A History*, 240.

9. León-Portilla, *In the Language of Kings* (New York: W. W. Norton and Company, 2001), 57.

10. *Códice Chimalpopoca: Anales de Cuauhtitlán y Leyenda de los Soles*, trans. Primo Feliciano Velázquez (Mexico: Universidad Nacional Autónoma de Mexico, 1992), 119-121. Even in this version of the story where Quetzalcoatl acts as a benevolent giver of life, he "tricks" the gods of the underworld into giving him the bones, and the ants into giving him the corn.

11. Carrasco, 175. See also Sahagún, *A History*, 180. and León-Portilla, 187.

12. Florentine Codex, in Lockhart, *We People Here*, 51-57.

13. Carrasco, 2.

14. Carrasco, 42.

15. Carrasco, 202.

16. Jacques Lafaye, *Quetzalcoatl and Guadalupe: The formation of Mexican National Consciousness 1531-1813*, trans. Benjamin Keen (Chicago: The University of Chicago Press, 1976), 151.

17. Ibid., 151.

18. Camilla Townsend, "Burying the White Gods: New Perspectives on the Conquest of Mexico," *The American Historical Review 108* (June 2003): 683. Camilla also quotes López de Gómara from his book *Cortes: The Life of the Conqueror* (page 134), who gives another reason for why Montezuma might have opened the gates of Tenochtitlan to the conquerors for reasons other than naivet: "It seemed unfitting and dishonorable for him to make war upon Cortés and fight a mere handful of strangers who said they were ambassadors. Another reason was that he did not wish to stir up trouble for himself, for it was clear that he would immediately have to face and uprising among the Otomá, the Tlaxcalans, and many others."

19. Townsend, 674.

20. Townsend, 684-686.

21. Ibid., 660.

22. Lafaye, 48.

23. We People Here: Nahuatl Accounts of the Conquest of Mexico, trans. James Lockhart (Berkley: University of California Press, 1993).

20. The Spaniards are, however, referred to in various texts as teotl, a Nahuatl word that is often used in the naming of various Mesoamerican deities. Some scholars propose that the Aztecs referred to the Spaniards as teotl because no other term sufficed to describe such foreign invaders.

24. Carrasco, chapter 4.

25. John Leddu Phelan, The Millennial Kingdom of the Franciscans in the New World (Berkley: University of California Press, 1970), 17.

26. Columbus, The Book of Prophesies. Trans. Blair Sullivan (Los Angeles: University of California Press, 1997), 5.

27. Lafaye, 32.

28. Phelan, 23. Twelve, of course, was intended to symbolize the Twelve Apostles—a role that the Franciscans took quite seriously (pg. 44). Franciscan Gerónimo de Mendieta, a missionary historian, writes elaborately on the millennial theory in Historia eclesiástica Indiana, wherein Cortez appears as the new Moses who liberates the Indians from the "slavery of the Devil" by taking them to the Promised Land of the Church (29-38).

29. Fernando Cervantes, *The Devil in the New World* (New Haven: Yale University Press, 1994), 13.

30. Phelan, 25.

31. Ibid., 109.

32. Cervantes, 15 and 30.

33. Ibid., 15 and 30.

34. Lafaye, 155.

35. Carrasco, 194.

36. Lafaye, 45.

37. Cervantes, 10.

38. Sahagún, Florentine Codex, 69.

39. Matt. 24 as quoted in Lafaye, 46.

40. Lafaye, 45. Also, pages 178-179, and 190 give a detailed history of how St. Thomas was chosen through process of elimination as the most likely candidate for having been the Apostle of the Americas.

41. Robert Ricard, *The Spiritual Conquest of Mexico*, trans. Lesley Bird Simpson (Berkeley and Los Angeles: University of California Press, 1966), 3. Ricard points out on page 285 that each mendicant order in the New World specialized in their own area of mission work. While the Jesuits attended to creole society, the Franciscans focused on studying and documenting native cultures and on training native priests; the Dominicans were more tied to orthodoxy and taught the Bible literally to the Indians; and the Augustinians constructed large communes, building monasteries where they provided spiritual training to the Indians.

42. Ricard. 161.

43. Ibid., 161.

44. Poole argues that the cult of Our Lady of Guadalupe remained relatively small before 1648, when two priests published the first accounts of Juan Diego's vision of the Virgin Mary, which instigated a "rapid spread of devotion" throughout the creoles of New Spain. By this time, he explains, the priests of New Spain were largely creole by birth, and tended to support ideas that proved God's support of their endeavors in the New World (100).

45. Stafford Poole, *Our Lady of Guadalupe: the Origins and Sources of a Mexican National Symbol 1531-1797* (London: The University of Arizona Press, 1995), 156.

46. Ricard, 32.

47. Lafaye, 187.

48. Poole, 152.

49. Lafaye, 183.

50. In the early 1900s, historians Mariano Jacobo Rojas and Ignacio Dávila Garibi argued that the word "Guadalupe" originates from the indigenous word "Cutlalopeuh," which translates as "she who drives away the serpent" (Poole, 33). The serpent here is often thought to be Quetzalcoatl, representing the devil or perhaps the vestiges of Mesoamerican religion. While there are many other readings of the name "Guadalupe," this particular interpretation is interesting considering Our Lady of Guadalupe's role in stamping out Quetzalcoatl within Christian.

Sarah Enright

CLC-4-TTS and Fire Vox: Enabling the Visually Impaired to Surf the Internet

Charles L. Chen

Abstract

This paper is about software architecture visualization and a specific software architecture example. Working for Dr. Dewayne Perry on the NSF CCR-0306613 grant "Transforming Requirement Specifications into Architectural Prescriptions," I created a software architecture visualization tool. Given an architectural prescription, this tool automatically generates several views of the software,s architecture. Using software architecture principles and the architectural visualization tool from this project, I created the CLC-4-TTS (Core Library Components For Text To Speech) suite of Firefox extensions which provides visually impaired users with screen reading capabilities built right into their Firefox web browser.

Introduction

Software architecture is the first step in transforming a problem into its software solution. I have been participating in Dr. Perry's NSF supported research on transforming requirements into architectures [1] as an undergraduate research assistant. As part of this investigation, I was tasked with creating an architecture visualization tool, SAPIRE, to provide various graphical views of architecture prescriptions [2]. These architecture prescriptions result from applying a specific method to a formally defined, goal-directed requirements specification that describes the specific problem requirements as well as the problem domain. The resulting architecture prescriptions delineate the basic solution structure and the constraints on the components and their interactions. SAPIRE [3] is an architecture visualization tool that takes architectural prescriptions as input. Based on the interactions of the components in the architectural prescription, it creates three views of the architecture: a single-layered network of components, a hierarchy of components and their subcomponents, and a multi-layered network of components.

I chose to explore the area of screen readers as an example for applying both the architecture design principles and techniques and the architecture visualization tool because I had some familiarity with that domain and with the solution space of Mozilla¨ Firefox¨ extensions [4].

Screen readers are software applications that read back the content that is displayed on a computer screen to users. These screen readers are the primary means by which the visually impaired have access to computers. With the rapid growth of the Internet, good support for web browsing using screen readers has become an increasingly important issue, and the World Wide Web Consortium (W3C) has issued a set of accessibility guidelines for web developers [5].

Unfortunately, visually impaired users are faced with a very limited number of screen readers to choose from. Not only are many screen readers on the market quite expensive, most of them do not offer much support, if any, for web browsers other than Microsoft¨ Internet Explorer¨. While there are screen readers for most operating systems, there have not been any cross platform compatible screen readers that offer the same user experience regardless of the operating system.

One popular web browser alternative to Internet Explorer is Mozilla Firefox. The number of Firefox users has grown recently thanks to Firefox's resistance to several of the vulnerabilities that Internet Explorer has. According to Mozilla, Firefox has been downloaded over 20,000,000 times [6], and an article from CNET confirms that Firefox usage is steadily increasing [7].

Firefox is especially appealing to users who like to tinker with their software since it allows them to create and distribute extensions, i.e., plug-in JavaScript applications that add new functionality to the Firefox browser. Despite its popularity and ease of extension, there is still very little screen reader support for Firefox.

I decided to explore a screen reading solution for Firefox rather than Internet Explorer as Internet Explorer is already well supported by existing screen readers. I had a choice between developing generic code for multiple Windows¨ applications (including both Internet Explorer and Firefox) using Microsoft's accessibility framework or developing code specific to Firefox which would work on multiple operating systems (Windows, Mac¨, Linux¨); I chose to do the latter. Focusing on just Firefox also provided a more appropriate scope for this research project, and I believe it is important to understand a particular instance well before trying to generalize to more generic approaches.

To address this issue of accessibility in Firefox for the visually impaired, the CLC-4-TTS open source [8] project provides 1) a set of JavaScript library functions and 2) an implementation of screen reading features for transforming Firefox into a talking browser using this library. The separation of the library from the implementation enables others to create their own accessible extensions. The CLC-4-TTS suite is also built with cross platform compatibility as a major goal to provide developers and end users with a consistent set of interfaces, regardless of which operating system they are using.

Architecture of CLC-4-TTS

The architectural style of the CLC-4-TTS suite is constrained by Firefox. Firefox has a framework for

Charles L. Chen

extensions to interact with it, and all extensions must follow this framework in order to function properly. My design decisions were how to decompose and organize the components needed to compose the system in the context of Firefox's style constraints. The design decisions were in part determined by which components were likely to be satisfied by multiple alternate implementations.

I divided the CLC-4-TTS suite into three extensions: one for handling text-to-speech; one for retrieving information from Firefox; and one for making Firefox into a talking browser by tying the functionality of the first two together. This division allows the first two of these extensions to be used by others who wish to implement their own screen reader or need to use some of the functionality (such as text-to-speech) in order to provide better accessibility of their extensions for users with visual impairments.

CLC-4-TTS—Text-to-Speech Core

This extension provides functions for taking a text string and speaking it. In order to have cross platform compatibility later on, I made the underlying text-to-speech engine a configuration option that is set at run time. The current implementation supports both Microsoft SAPI for Windows and FreeTTS [9] for any operating system capable of supporting Java. I provide an abstract API in my library that is not specific for any text-to-speech engine. Because of these abstract functions, applications that use these functions can switch between the SAPI and FreeTTS text-to-speech engines without needing to change any of their code.

In addition to being a building block for screen readers, this library is designed to be especially helpful for anyone who is trying to make their extensions more accessible. The main usefulness of some extensions is their ability to add an icon to the navigation bar in Firefox, making it easy to immediately see the status of a specific option and toggle it without having to go through all the menus. These extensions are not accessible since screen readers are unable to read back these status icons in a sensible way. However, their authors can provide

accessibility by using the CLC-4-TTS Text-to-Speech Core to announce the status of the setting when it is toggled. Since FreeTTS is a Java based text-to-speech engine, it works on Windows, Mac, and Linux computers; thus it provides a cross platform alternative to SAPI which only works on Windows.

CLC-4-TTS—Utilities

This extension provides functions that are useful in processing information that has been retrieved from the browser as well as in modifying that information so that the browser displays it differently (i.e., the high-contrast highlighting effect). In these functions, I have extended the concept of the Document Object Model (DOM) [10] by defining the abstract concepts of *atomic DOM* objects and the *lineages of objects* to simplify the process of locating the smallest logical unit of content on web pages and identifying the HTML elements that these units belong to. I also have functions that can identify HTML elements, highlight HTML elements, determine if an HTML element has text content, find the frame which currently has focus, and locate the DOM object that the cursor is placed at.

This library is useful for implementing a screen reader as well as for any general web development tool that needs to analyze the various HTML elements on a page. One of the major advantages of using the DOM instead of Microsoft's MSAA, which is more commonly used by commercial screen readers because it makes event handling easier for developers, is that the DOM is platform independent and available on all operating systems, not just on Windows.

CLC-4-TTS—Fire Vox

This extension rests on top of the other two and, as application layer software, provides coordination between them to give users the control needed to read back web pages in a sensible fashion. It retrieves the DOM object on the web page as requested by the user, processes this object by using the functions in the utilities library to extract the necessary information about it, and then speaks an appropriate audio message using the functions in the text-to-speech core.

Unlike the first two extensions, this extension is designed to be a program and not a library of functions. It provides visually impaired users with a freeware screen reading solution that can be used to access web content. Since the libraries that it depends on are cross platform compatible and all of its code is cross platform compatible, Fire Vox works on Windows, Mac, and Linux.

Visualization of the Architecture

Before implementing these extensions, I created an architectural prescription [2] for the system as a whole and generated some visualizations of it using SAPIRE. Figure 1 is one of these graphical views generated from this tool using the architectural prescription for CLC-4-TTS. The architectural prescription can be found in Appendix A; other graphical views of the architecture are presented in Appendix B. This architecture served as a blueprint to keep me on track as I did the detailed design and coded the extensions.

An architectural prescription is composed of components that have defined interrelationships and constraints. Ovals represent the data components of the system; double circles represent the processes that act on the data; and rectangles represent the connectors that hold the system together by capturing the interactions among these components. Components are grouped according to which subsystem they belong to. Components can be placed as Starting, Middle, Terminal, Input, Output, or Both Input and Output. Starting components use other components but are not used by any other components. Middle components use other components and are used by other components. Terminal components are only used by other components and do not use other components. Input components are subcomponents in a larger component; they are the parts of that larger component which is used when that component is used by other components. Output components are subcomponents in a larger component; they are the parts of that larger component that use other components when that larger component uses other components. Both Input and Output components are components that act as both Input components and Output components in a larger

component. The Detail Level is an indication of the level of granularity of the component; larger numbers indicate components which are more detailed and closer to the implementation level.

Users interact with the main Firefox browser through some user interface, typically a keyboard (and possibly a mouse). Users can also interact with extensions such as the CLC-4-TTS suite using the extension system in Firefox. When users request content from the Internet, the browser retrieves the content through commonly used protocols such as HTTP or FTP and then creates a DOM object from this content. Fire Vox uses the Utilities to process this DOM object and generates the text that should be voiced to the user. Fire Vox then sends this text to the Text-To-Speech Core which in turn passes the text to an external speech engine, either SAPI via an XPCOM component or FreeTTS via a LiveConnect component. Finally, the external text-to-speech engine generates the audio output for this text using the sound card on the computer.

Implementation

Writing the code for the text-to-speech core library and for the utilities library was relatively straightforward. The functions in these libraries were designed as small modules that performed simple tasks such as passing strings to the underlying text-to-speech engine, retrieving the tag name information from a specified node in the DOM of a web page, etc. Related functions were encapsulated in the same file so that it would be easier to find them later on. By following the principles of modularity and encapsulation, I was able to implement the initial versions of these two libraries in less than two weeks.

The Fire Vox extension proved to be more difficult. The major problem was that of parsing a web page. I had originally wanted to use a pre-parser strategy. My initial plan was to use the atomic DOM object and lineage functions in the utilities library to find the smallest logical units on a page and obtain enough information about the objects that they belong to and add special tags that would make it easier to locate them, announce identifying information about them, and feed

Charles L. Chen

Figure 1 | *A Visualization of the CLC-4-TTS Architecture.*
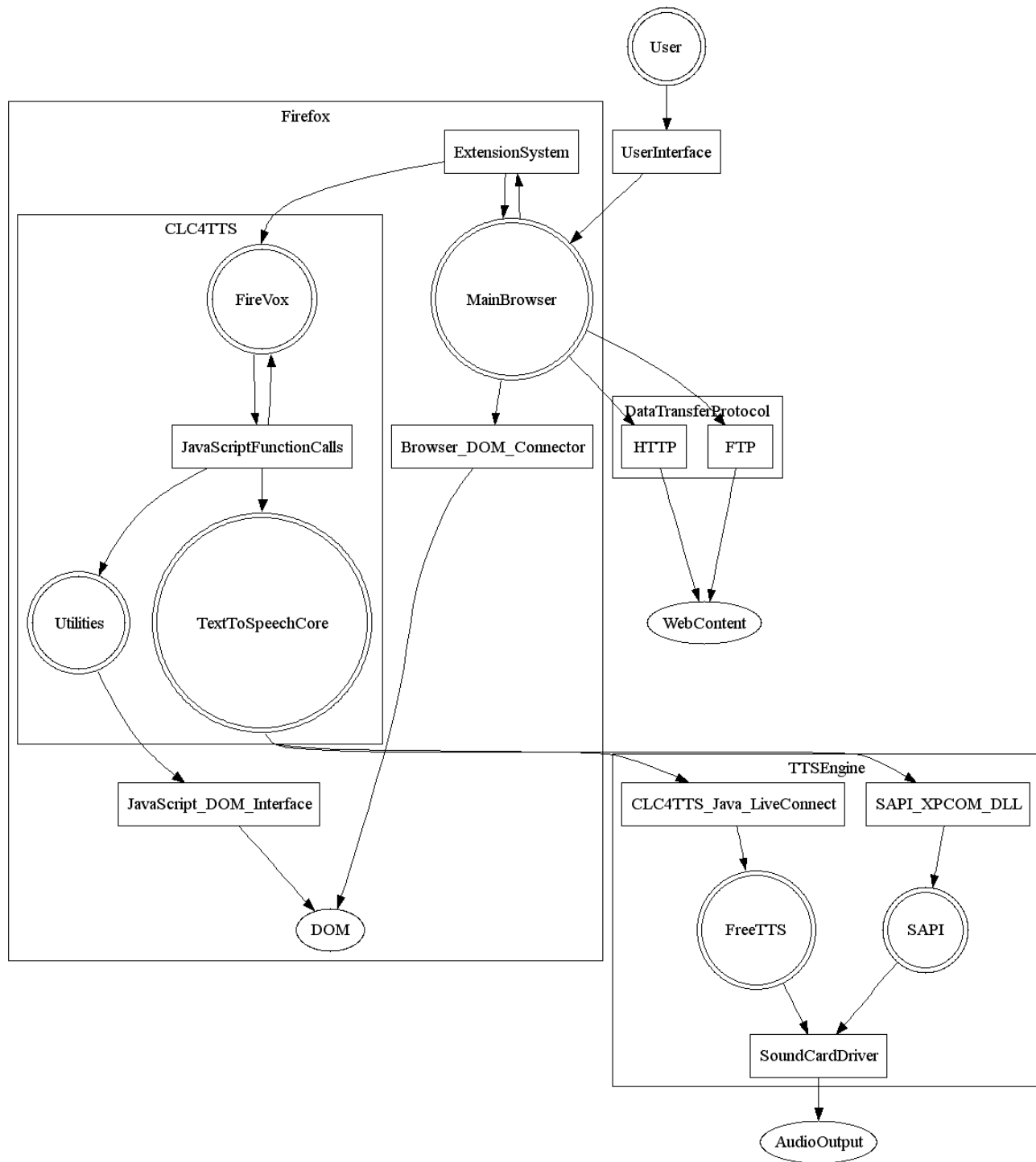
their content to the text-to-speech engine using the text-to-speech core library functions. This approach, however, had some problems. Most importantly, it relied on the underlying assumption that the web page content would remain static after the page has loaded. This assumption was not valid since web pages can have content added to them dynamically thus causing improper identification. Using a pre-parser method that changes parts of the web page can also create both minor annoyances such as some distortion to the page layout and major problems such as causing Firefox to crash. Because of these issues, I decided to switch to dynamically parsing web pages.

Dynamically parsing a web page is harder than pre-parsing because it requires more work to be done by Fire Vox and it also adds an element of uncertainty. There is more work involved since it is no longer possible to simply gather lists of all the different types of elements on a web page and add some tags in front of them. Instead, Fire Vox must go from one element to the next in an orderly fashion. This is complicated because these elements will be at different levels in the DOM. Dynamic parsing adds an element of uncertainty because there is no guarantee about the properties of the next object on the page—the next object could be text, a graphic, an input field, etc. Even more uncertainty is introduced by the internal links on a web page. Because these links allow users to jump directly to a specific section of the page, there is no guarantee that the next object that should be retrieved will be the one which comes immediately after the current object.

To deal with these challenges, I used the abstractions of *atomic DOM objects* and their *lineages* in a different way. Before moving to a new atomic object, Fire Vox saves the current atomic object. After moving to a new atomic object, Fire Vox compares this new object's lineage against the lineage of the atomic object that it had just left. At the HTML element where the lineages first diverge lies the first HTML element that should be identified and announced before reading the content of the new object.

Examples of how this functions are as follows. Note that these examples are not meant to represent an exhaustive set of possible scenarios. Rather, they are given to provide concrete instances that illustrate how the system behaves. These are relatively simple examples since more complex examples would be better explained with demonstrations rather than with descriptions.

Case 1. The user is at the table of contents of a page. The table of contents is created by having an unnumbered list of links that will allow the user to jump directly to the relevant section. The user is reading through the list and hears something that he would like to learn more about. He or she activates the link, and it just so happens that the section he has chosen is another unnumbered list element with list items inside. By comparing the lineages, Fire Vox determines that the current object and the previous object both share the same <BODY> HTML element but are not the same unnumbered list. Therefore, it announces the title of the new list and the number of items in this list.

Case 2. The user is going through a long set of directions which is created by having an ordered list with list items. At the end of one of the steps, there is a link that allows users to skip three steps and the text in that link is telling readers to skip the next three steps if they have met a certain condition. The user realizes that she has met the condition and activates the link. She is taken past the next three steps, and when she tries to read the next object, only the number of the step is announced and its contents read back. The title for this list of instructions and the total number of steps are not repeated. Fire Vox has detected that the previous <LI> element and the current <LI> element are both in the same ordered list according to their lineages and that there is no difference between their lineages until the <LI> elements themselves.

Related work

Although Fire Vox only provides accessibility for Firefox and thus has a narrower focus than other screen readers which provide accessibility for the entire system, it is comparable to (and in some ways, better than) these other solutions in making web content accessible.

For Windows, one of the leading screen readers is JAWS¨ from Freedom Scientific¨ [11]. It is a screen

Charles L. Chen

reader that works for the entire operating system and tries to support as many applications as possible; however, it only supports Internet Explorer and does not support Firefox. When reading back web pages in Internet Explorer, it will add additional information such as identifying headings, links, images, etc. - a feature that Fire Vox has as well for reading web content in Firefox. The main advantage JAWS has over Fire Vox is that JAWS will support accessible Flash in Internet Explorer. However, Fire Vox supports reading back MathML [12] in Nemeth code, a standard for conveying mathematical information to the visually impaired [13], whereas JAWS does not support MathML at all. JAWS costs $895; additional upgrades require a software maintenance agreement which costs an additional $120 at the time of purchase.

Mac comes with its own built-in screen reader called VoiceOver [14]. VoiceOver does not function very well with Firefox, but it does work with Safari[a] [15]. Since VoiceOver and Safari come with the operating system, it is much easier for users to get started; Fire Vox requires the installation of Firefox, the Java Runtime Environment, FreeTTS, and the CLC-4-TTS suite. Despite the extra effort needed in setting it up, Fire Vox has an advantage because it respects the cursor position and can start automatically reading at any point. VoiceOver is unable to use skip navigation links to jump to the main content and start reading there, and it also has problems with resuming automatic reading after the user has manually skipped over some content [16].

Linux has Speakup [17] and Emacspeak [18], which are both free open source projects. However, neither has support for Firefox; instead they support Lynx [19] and Emacs/W3 [20]. Lynx is a text only, terminal based browser, and Emacs/W3 lacks some of the more advanced HTML features that are standard in web browsers such as Internet Explorer and Firefox. These limitations mean that several web sites, especially sites that require forms, will not work properly. Firefox can handle forms without any problems, and Fire Vox provides special form handling features such as providing a list of all the forms on the page, reading back the content in a form field, echoing keys that have been typed in normal blanks, and only

saying "asterisk" when keys are typed in a password blank to preserve the secrecy of the user's password.

All of the above screen readers are also specific to their respective operating systems. JAWS will only work under Windows, VoiceOver is part of the Mac operating system, and Speakup and Emacspeak are only for Linux. In contrast, Fire Vox is a solution that works across all operating systems that support Java. In fact, Fire Vox is the first platform independent screen reading solution for a web browser; together with Firefox, it provides a consistent user experience regardless of the underlying operating system. This reduces the amount of learning needed to use different operating systems for browsing the web, a problem that the visually impaired face when needing to switch between systems (for example, using a Windows computer at work and a Mac at home).

Lessons Learned

Modularity, encapsulation, and abstraction proved to be powerful techniques for dealing with the complexity encountered in transforming Firefox into a talking browser. Modular components that perform only one simple, straightforward task can be tested and debugged without too much effort. Encapsulating related functions in the same file kept all of the library functions well organized and made it easy for me to locate the function I wanted. Abstractions such as atomic DOM objects and the lineage of a DOM object greatly simplified the task of moving through the DOM and retrieving the information that was necessary.

By intentionally creating all of the screen reading components as libraries designed for reuse, I was able to make changes quickly, easily, and correctly to Fire Vox. Changing some of the function calls being made, using their results in a slightly different manner, and writing a few lines in the Fire Vox code, allowed me to successfully switch from a pre-parsing strategy to a dynamic parsing strategy with less than two hours of coding. This confirms Brooks' statement that programmers can "radically [raise] the conceptual level, and [eliminate] the vast amounts of work and the copious opportunities for error that dwell at the individual statement level" by using "pre-built collections" [21].

Because library functions need to work in various situations, they must be generalized and have well defined interfaces. As a result, they are much harder to design than specific one-shot functions. Having a solid architectural prescription before I started the implementation was crucial in providing me with a clear roadmap for building the screen reading libraries.

Conclusion

The results of creating the CLC-4-TTS suite support the proposition that software architecture is useful as a "framework for satisfying requirements," as the "technical basis for design," as an "effective basis for reuse," and as the "basis for dependency and consistency analysis" [22].

Following an architectural approach simplified the design. It forced me to carefully consider the requirements that I was trying to meet before jumping to an implementation. In the long run, having a well thought out plan as the basis for design and code resulted in less work and faster progress. Generating the visualizations for the architectural prescription of this system made it easy to understand the structure, thus enabling me to better divide the system into reusable libraries and reason about the behavior of the interacting components.

This project has sparked considerable interest in both academia and industry. A software engineering class at the University of Wisconsin (Milwaukee) extended Fire Vox as one of their class projects for the Fall 2005 semester. Pamela Berman of Bloomsburg University of Pennsylvania interviewed me in February 2006 for a podcast regarding CLC-4-TTS and Fire Vox. I was invited as a speaker for the June 2005 meeting of the Public Electronic Services On-the-Internet (PESO) Working Group, a part of the Department of Information Resources of the State of Texas. Google has expressed interest in seeing the addition of CSS speech properties to the set of web page features which are supported, and Sun Microsystems is interested in incorporating CLC-4-TTS into their Orca screen reader for Linux. The CLC-4-TTS Suite was used as one of the screen readers at the Knowbility booth at SXSW 2006, and the volunteers doing web accessibility demonstrations felt that it was easier to use and performed better than JAWS. The Mozilla Foundation has invited me to present this project at the CSUN 2006 conference in California in March.

References

[1] Perry, Dewayne E. *NSF Project: Requirements to Architecture. http://www.ece.utexas.edu/~perry/ work/projects/nsf-r2a/index.html,* February 21, 2005.

[2] Brandozzi, Manuel and Perry, Dewayne E. "From Goal-Oriented Requirements to Architectural Prescriptions: The Preskiptor Process." International Workshop From Software Requirements to Architectures, May 2003, pp 107-113.

[3] Chen, Charles L. "Tech Report on SAPIRE." In preparation.

[4] Firefox Home Page. http://www.mozilla.org/products/firefox/, May 26, 2005.

[5] Chisholm, W., Vanderheiden, G., and Jacobs, I. *Web Content Accessibility Guidelines 1.0.* http://www.w3.org/TR/WAI-WEBCONTENT/, February 21, 2005.

[6] mozillaZine. *Mozilla Firefox Hits 20,000,000 Downloads. http://www.mozillazine.org/talkback. html?article=5953,* January 24, 2005.

[7] CNET News.com. *Firefox continues gains against IE. http://news.com/Firefox+continues+ gains+against+IE/*2100-1032_3-5545930.html, January 21, 2005.

[8] Open Source Initiative OSI, *The GPL:Licensing. http://www.opensource.org/licenses/gpl-license.php,* February 22, 2005.

[9] FreeTTS Home Page. http://freetts.sourceforge.net/docs/index.php, May 26, 2005.

[10] H garet, P., Whitmer, R., and Wood, L. *W3C Document Object Model. http://www.w3.org/ DOM/,* February 22, 2005.

[11] JAWS Home Page. http://*www.freedomscientific. com/fs_products/software_jaws.asp*, June 24, 2005.

[12] Ausbrooks et al. *Mathematical Markup Language (MathML) Version 2.0 (Second Edition).* http://www.w3.org/TR/MathML2/, June 24, 2005.

[13] Nemeth, A. *MATHSPEAK. http://www.rit.edu/ ~easi/easisem/talkmath.htm*, June 24, 2005.

Charles L. Chen

[14] VoiceOver Home Page. *http://www.apple.com/ macosx/features/voiceover/,* June 24, 2005.

[15] Safari Home Page. *http://www.apple.com/macosx/ features/safari/,* June 24, 2005.

[16] Johansson, Roger. "VoiceOver and Safari: Screen reading on the Mac." http://www.456bereastreet. com/archive/200505/voiceover_and_safari_screen_ reading_on_the_mac.html, June 24, 2005.

[17] Speakup Home Page. http://www.linux-speakup.org/, June 24, 2005.

[18] Emacspeak Home Page. *http://emacspeak.source forge.net/,* June 24, 2005.

[19] Lynx Home Page. *http://lynx.browser.org/,* June 24, 2005.

[20] Emacs/W3 Home Page. *http://savannah.gnu.org/ projects/w3/,* June 24, 2005.

[21] Brooks, Frederick P. Brooks, Jr. *The Mythical Man-Month. Anniversary Edition.* Massachusetts: Addison Wesley Longman, 1995.

[22] Perry, Dewayne E. and Wolf, Alexander L. "Software Architecture." August 1989.

Appendix A

Architectural Prescription for CLC-4-TTS

Name	Type	Placement	Detail Level	Contains	Uses
User	Process	Starting	0	-	UserInterface
UserInterface	Connector	Middle	0	-	Firefox
Firefox	Process	Middle	0	CLC4TTS, ExtensionSystem, MainBrowser, Browser_DOM_Connector, JavaScript_DOM_Interface, DOM	DataTransferProtocol, TTSEngine
DataTransferProtocol	Connector	Middle	0	FTP, HTTP	WebContent
WebContent	Data	Terminal	0	-	-
TTSEngine	Connector	Middle	0	SAPI, SAPI_XPCOM_DLL, FreeTTS, CLC4TTS_Java_LiveConnect, SoundCardDriver	AudioOutput
AudioOutput	Data	Terminal	0	-	-
MainBrowser	Process	Both input & output	1	-	ExtensionSystem, FTP, HTTP
ExtensionSystem	Connector	Middle	1	-	CLC4TTS, MainBrowser
Browser_DOM_Connector	Connector	Middle	1	-	DOM
JavaScript_DOM_Interface	Connector	Middle	1	-	DOM
DOM	Data	Terminal	1	-	-
CLC4TTS	Process	Output	1	JavaScriptFunctionCalls, FireVox, TextToSpeechCore, Utilities	JavaScript_DOM_Interface, TTSEngine
FTP	Connector	Both input & output	1	-	WebContent
HTTP	Connector	Both input & output	1	-	WebContent
SAPI	Process	Middle	1	-	SoundCardDriver
SAPI_XPCOM_DLL	Connector	Input	1	-	SAPI
FreeTTS	Process	Middle	1	-	SoundCardDriver
CLC4TTS_Java_LiveConnect	Connector	Input	1	-	FreeTTS
SoundCardDriver	Connector	Output	1	-	AudioOutput
JSFunctionCalls	Connector	Middle	2	-	Utilities, TextToSpeechCore
TextToSpeechCore	Process	Output	2	-	CLC4TTS_Java_LiveConnect, SAPI_XPCOM_DLL
Utilities	Process	Output	2	-	JavaScript_DOM_Interface
FireVox	Process	Input	2	-	-

Charles L. Chen

Appendix B

Visualizations of the CLC-4-TTS Architecture

Level 0:
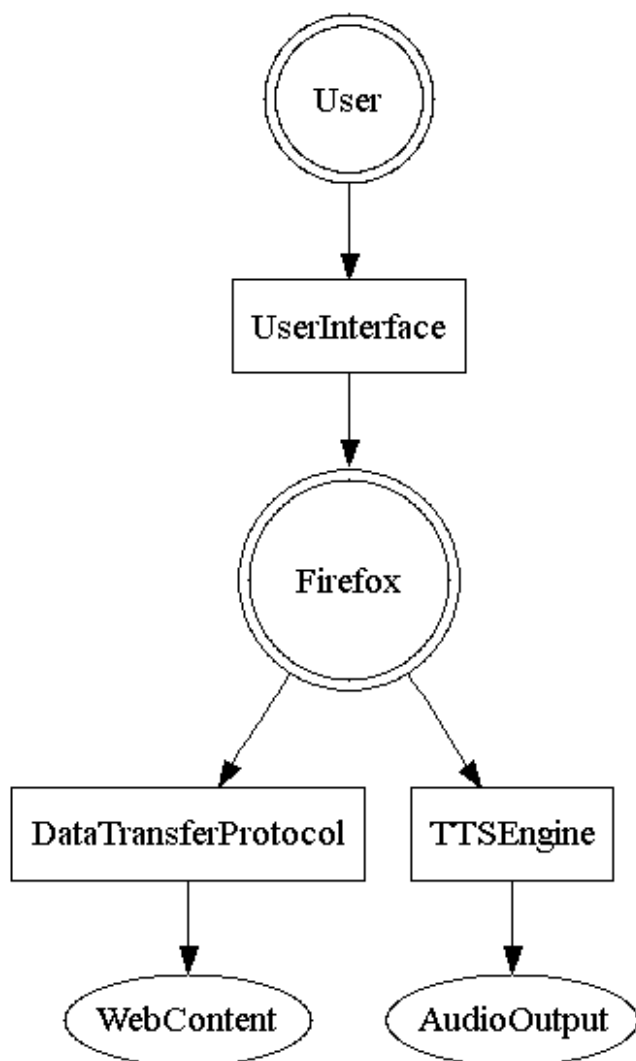
Level 1:

A Series of Stages:
A Theatrical and Historical Reading of Moscow and Yershalaim in The Master and Margarita

Anna E. Shoemaker

Abstract

Since its illicit initial appearance in 1940, Mikhail Bulgakov's *The Master and Margarita* has resonated with both Russian and foreign readers. A highly complex Soviet era retelling of Goethe's *Faust* conjoined with Pontius Pilate's account of the trial and execution of Jesus of Nazareth, the story has pervaded Russian culture by way of the page, stage and screen. Bulgakov, also a playwright, composed this novel with aims of producing it theatrically. By considering the theatrical aspects of the novel's structure in terms of specific Slavic folk traditions—specifically the *vertep* theater that evolved from the Nativity and mystery plays that spread across Europe after the Middle Ages and into the Enlightenment—the reader can achieve a greater understanding of its complexity. Such a reading is not

complete, however, without a historical consideration of its content and style in relation to poetics observed in works of literature created by other Slavic writers under totalitarian regimes. In this paper, I attempt to show that it is through these theatrical and historical readings that the cultural profusion of Bulgakov's chef d'oeuvre can be more fully understood. I additionally provide examples of how such readings have been successfully applied to critical and theatrical interpretations of the novel. By using these models of interpretation on what could be called the standard bearer of literature produced during the Soviet era, I hope to show that such readings could be applied to a broader spectrum of works created under totalitarian regimes for the purpose of critical interpretation and adaptation.

Mikhail Bulgakov's final work, *The Master and Margarita*, is a highly complex novel, both in structure and content. Its dual plotlines intermittently merge and diverge, creating a strange flight between worlds—or shift between stages—that ultimately serves to complement and integrate the novel's two halves. By examining *The Master and Margarita*'s structure, content, and style in theatrical and historical terms, a greater clarity can be brought to this challenging work.

In *The Master and Margarita*, Bulgakov presents his reader with two worlds. The first is that of Moscow during the 1930s, under Stalin's harsh regime. This world contains the title characters, the Master, and his lover, Margarita, along with figures from the official Moscow literary establishment—namely Misha Berlioz and the poet Ivan Bezdomny—and a whole cast of typical Muscovites. This world is turned upside down by a visit from Woland (who is actually Satan), a self-professed professor of black magic, and his devilish cohorts. Their incredible deeds astound the inhabitants of Moscow, many of whom will never be the same again.

The second world is that of Jerusalem, or Yershalaim, as Bulgakov refers to it, during the trial and execution of Yeshua Ha-Nozri and the aftermath of these events. This world is seen through the eyes of the fifth procurator of Judea, Pontius Pilate, as well as Yeshua's disciple, Matthew Levi, and Yeshua's betrayer,

Judas of Kiriath. This novel within a novel is the literary work of the Master, and is recounted to the audience by Woland, dreamed by Ivan Bezdomny, and read directly from the Master's manuscript.

Rita Giuliani compares the structure of Bulgakov's novel to that of Ukrainian popular puppet theater, or *vertep*.[1] *Vertep* dates to the nineteenth century, and, like the *betleika* of Belarus[2], has its origins in the Polish Nativity puppet theater *szopka*.[3] The *szopka* originates from the mystery plays that arose across Europe during the Middle Ages. The original biblical scenes performed were gradually enhanced with "theatrical effects, secular motifs, and inevitably, laughter."[4] This eventually led the plays to be banned from church interiors, outside of which they developed into the form of theater known as the mystery play. The mystery play was presented both by human actors and by puppeteers. However, as the performances were continually augmented the plays became so popular that they began to draw crowds away from the churches themselves, naturally drawing the wrath of the Church.[5] Throughout the course of the eighteenth and nineteenth centuries, these forms of folk theater became more associated with social commentary and the nationalist spirit that was taking Europe by storm. Touring puppeteers were welcomed in Poland, from which the form spread to Belarus and Russia.

Little information exists concerning the early repertory performed by these touring groups.[6] However, their influence was exerted on the traditional nativity play, the *szopka*, which is the name of both the stage construction and its associated performance. A *szopka* performance takes place in an often elaborate portable cabinet consisting of three levels, two of which are used for performing scenes from the Nativity, Herod's palace, and "profane interludes,"[7] which feature "characters concerned with contemporary social or political problems."[8] The Nativity scenes are performed on the upper level, the scenes of Herod and the comic interludes on the lowest level, and the middle floor between them is screened.[9] These divisions are seen to signify heaven, earth, and hell.[10] In its most evolved form, the *szopka* is only minimally concerned with the "mystery" of Christ's birth. The secular scenes were not the only

opportunities for subversion, as Herod himself was a symbol of corrupt political power due to his attempt to protect his authority from the Christ child by executing all baby boys born in his domain. However, most of the performance focuses on the secular and satirical scenes involving local characters, as mentioned above. These scenes primarily reference contemporary local life rather than biblical subjects. This part of the performance always includes a "Margareta" who dances with "a band of hussars." Pieces from popular vaudeville and farces are occasionally inserted as well, such as scenes from *Traveling through Warsaw* and *Birdcatcher of the Tyrol*.[11]

The *szopka*'s "episodic structure allowed the introduction of figures from contemporary Polish political and artistic life, with possibilities quickly exploited by patriotic circles."[12] For example, one such performance consisted of the following:

> In 1790. . . in one of the *szopkas*, after the inevitable passing of Herod and Death from the little stage, ingenious little figures representing dignitaries of the Polish Kingdom and associates of King Stanislav August appeared one after another, while the voice from behind the *szopka* sang a song. . . [which would have consisted of] a satirical riddle juxtaposed with a small rod puppet in the form of some well-known public personality. The spectators solved the riddles by shouting the name of the person represented, expressing hatred or approval. . . they would have recognized certain political intriguers and renegades.[13]

Another *szopka* presented the revolutionary French army fighting against the forces of Herod. The popularity of this particular performance alarmed the Polish king, as patriotic radicals were encouraging a similar solution to that of the French for Polish problems.

During the eighteenth century, Russia was oppressed by an authoritarian regime "which controlled the movements of every visitor in the same way as it controlled the life of its Russian subjects."[14] However, the Tsar encouraged the introduction of Western forms of entertainment in his country and encouraged foreign comedians to visit. Little information exists concerning the content of such performances, but the documents concerning a puppeteer named Jan Splavsky reveals that

his time in Russia was highly problematic for both himself and the authorities. Being of Polish descent, it is presumed that the content of his performances mirrored those taking place in Poland to the extent allowed by the censors.[15]

In the nineteenth century, as texts of the plays were published for the first time, the unity between the nativity play as it developed in the Ukraine and that of Poland became evident. Poland and the Ukraine were so culturally linked that in *vertep* the play of the Nativity was written in Polish and the interludes in Ukrainian. The most popular Nativity in Poland was the Krakow *szopka*, which consisted of fifteen scenes including "the shepherds' homage, the Herod story, and a revue featuring regional characters such as a Krakovian boy and girl, a Cossack boy and girl, a Jew, a Hungarian, a Witch and *Pan* (Mr.) Twardowski, the Polish Faust." The Warsaw *szopka* primarily consisted of scenes including local characters such as a Sand-dredger, a Tinker, and a Chimneysweep. The censors were stricter in Warsaw, however, and the patriotic themes seen in the Krakow *szopka* had to be abandoned for purely entertaining subjects.[16]

The Ukrainian *vertep* closely resembles the *szopka* in form. Although the shape and appearance of the theater differ in that it is less ornate, it still consists of the tri-partitioned stage signifying heaven, earth, and hell. *Vertep* also further reduces the number of Nativity scenes, focusing on the profane interludes. The main character, known as the Cossack "*Zaporozhets*" is included here, who was likely based on *Kozak* (Cossack) *Mamai*, a figure from Ukrainian folklore who was "a great warrior, defender of freedom, and honor."[17] It is likely that the Cossack character in *vertep* is based on this figure. *Zaporozhets* "fought with everyone" in these scenes, including the Polish Landlord character, who represented the old social conflict between the peasants and the "polonised gentry of the Ukraine."[18] Satan also figures here,[19] in scenes which "dealt with a variety of negative characteristics of man, i.e., cowardice [and] greed. . . the scenes changed rapidly and in them people lied, cheated, tricked one another, argued, and fought."[20] *Zaporozhets* also became a stage villain and a mouth-

Anna E. Shoemaker

piece of the Russians and their political program in the eastern texts. The point of vertep was thus to entertain all ages, while inserting scenes that could potentially subvert the current authority by revealing their negative characteristics, depending on the will of the puppeteer and the nature of the audience to which he catered.

In the early twentieth century, the *szopka* retained and expanded its prominence, and not just in Poland. Modernist artists and writers appropriated old forms and recycled them in experimental forms. Polish Symbolist writer Stanislav Wyspianski conceived of a satirical play, *Wesele*, which revealed the cowardice of the Polish: "When the moment arrives [to carry out the conspiracy, the Polish] are unable to act, because the conspirators lack the courage or the ability to make decisions." *Wesele* was performed in Paris and based on the traditional *szopka* that divided the characters into three groups: "the living characters of the drama, the Apparitions (historical figures representing "national consciousness"), and puppets [that replace] the first group when they [are] placed under a spell."[21] This form would later be revisited in the 1980s by artists under the Soviet regime seeking self-expression by use of techniques of folk theater. The traditional *szopka* also lived on, and was so entrenched in the culture of Eastern Europe that *Teatr Kukielek* (The Theatre of Rod Puppets) accompanied the Polish army in the Soviet Union.[22] Polish puppet productions were even presented in German prisoner-of-war camps and concentration camps across Eastern Europe during the Second World War.[23]

To return to Giuliani's idea, the structure of *The Master and Margarita* can thus be imagined as if the story of Pilate and Christ in Yershalaim takes place in the upper level of the *vertep* stage, while the story of Woland in Moscow takes places in the lower level. Such a comparison might have even come to mind to a Russian reader, as the Ukrainian *vertep* tradition became known even to the remote region of Siberia.[24] The two worlds of Bulgakov's novel possess many of the same characters of the two levels of *vertep*, particularly the scenes that take place in Moscow. The Moscow scenes include satirical episodes concerning the corruption of Soviet authority regarding the Moscow housing shortage, the

banning of foreign currency, and the "disappearances" perpetrated by the secret police. As in *vertep*, many of the scenes in the Yershalaim storyline take place in Herod's palace, like in *vertep*. Both *vertep* and the Moscow scenes include secular and satirical scenes of local characters, a Margareta character, inserts from theatrical performances, and deal with "a variety of negative characteristics of man," namely cowardice and greed.[25] The character Ivan Bezdomny could even be viewed as a *Zaporozhets* figure, as both experience conflict with the prevailing local power.

Bulgakov distinguishes between the two worlds he creates within the novel stylistically just as the separate levels of the stage serve to distinguish between the two worlds of *vertep*. In the story of Pilate and Christ, Bulgakov "simulates the style of the historical novel"[26] in a fashion reminiscent of another of his works, *The White Guard*.[27] The use of the names Yeshua Ha-Nozri, Yershalaim, and Judas of Kiriath in place of Jesus, Jerusalem, and Judas Iscariot serve to distance the text from the Western synoptic gospels and lend more historic authenticity to his account of Christ's judgment and crucifixion. The use of names closer to their Greek and Hebrew origins also reflects the influence of the Russian Orthodox tradition on Bulgakov rather than that of the Western tradition. This is due to the fact that the Russian Orthodox Church lacks an official modern translation of their bible, which is traditionally written in Slavonic, the Russian equivalent of the King James Version of the bible.[28] In contrast, the characters speak in a more colloquial language throughout the story of Woland's exploits in Moscow. Bulgakov's journalistic experience writing *feuilletons*—"savagely funny sketches written in a witty, urbane style, sometimes descending to the level of street speech, that deal with a variety of bizarre or tragicomic incidents in the daily life of Moscow"[29]—lent itself to this style, as Bulgakov employed almost identical tactics in composing the Moscow scenes in *The Master and Margarita*.[30] This stylistic differentiation lends consequence to the fictional reality of the story of Pilate and Christ when compared with the "reality" in Moscow. When contrasted with the humorous, satirical, and occasionally ridiculous episodes seen

in the Moscow plotline, the fictional world in Yershalaim seems more realistic than the "real" world in Moscow.

Yet Bulgakov also constructed these worlds as parallels, as the many similarities between the two plotlines illustrate. The two stories begin at the same time of the year, on Good Friday, and proceed under a full moon. When introducing the two sets of characters, a similar conversation about the existence of God takes place between Woland, Berlioz and Bezdomny in Moscow and in Yershalaim between Pilate and Yeshua Ha-Nozri. Woland is present in Moscow and claims to have been present at the events taking place in Yershalaim. As Ellendea Proffer notes:

> The connections between the Pilate chapters and the rest of novel are multifarious. Themes, phrases, and motifs are shared. Roses, the colors red, black and yellow (used as the colors of the Devil in Dante), the phrase "O gods, gods!" (which comes from *Aida*)—all of this suggests parallels between people and events through space and time.[31]

The greatest indication that "events of monumental importance are happening simultaneously in both places,"[32] however, is the merging of the two plotlines at the end of the novel and the fact that the last words of their final sentences are identical: "the cruel fifth procurator of Judea, the equestrian Pontius Pilate."[33]

Similar themes are also addressed in both storylines. The two plotlines both reflect worlds where corrupt materialism strongly influences the lives of their inhabitants. In the Yershalaim plotline, the arrest and ultimate execution of Yeshua is brought about because Judas of Kiriath is bribed by religious leaders to betray him. In Moscow, the Master is arrested because Aloisy Mogarych, motivated by the acute housing shortage in Moscow at the time, falsely claims that he possesses illegal literature.[34] His denouncer then moves into his apartment as the Master is taken away, first for questioning, then to Stravinsky's asylum. One of the greatest tricks Woland plays on the Muscovites begins at the s ance of black magic, and involves tricking the greedy crowd with currency that later disappears or

transforms into worthless bits of paper. Woland also takes advantage of the avaricious women of Moscow, presenting them with expensive foreign clothing that disappears as they leave the theater after the s ance. Berlioz's uncle, Maximilian Andreevich, rushes to Moscow upon hearing of his nephew's death not out of grief, but because he, motivated by the housing shortage like Aloisy, wants to lay claim to Berlioz's apartment.

Additionally, Bulgakov establishes a "pattern of confrontation" in both plotlines, in which "[an] individual who completely embodies the values of humanism encounters the institutionalized forms of power that are anti-humanist by definition."[35] Such confrontations occur between Pontius Pilate and Yeshua Ha-Nozri during their interview, Ivan Bezdomny and Dr. Stravinsky during his diagnosis in the mental hospital, and the Master's conflicts with the literary establishment. The result of such confrontations is often ambiguous. Yeshua is executed, but according to Christian faith, he is resurrected and triumphs over death. Ivan is confined to Stravinsky's asylum until he accepts the state's explanation for Woland's doings, but he ultimately recalls the truth once a year, under the full moon of the springtime. The Master is taken away, then left like Ivan in the asylum, but he is eventually reunited with Margarita, given peace, and his novel ultimately lives on in the mind—and as some critics posit, the writings—of Ivan Bezdomny.[36]

Furthermore, the truth is something that is easily manipulated and distorted by those in power throughout Bulgakov's two storylines. Yeshua Ha-Nozri first mentions such distortion when he says that Matthew Levi incorrectly transcribes his speech, to which he credits his arrest and charges of wrongdoing. Pilate himself organizes the murder of Judas of Kiriath and covers it up with the rumor that Judas committed suicide. The ultimate untruth of the Yershalaim storyline, however, is the condemnation of Yeshua Ha-Nozri, whom Pilate knows to be innocent. In Moscow, Woland's devilry is constantly dismissed and explained away, despite undeniable occurrences witnessed by many. In the epilogue, the "cultured people" of Moscow "adopted the view. . . [that the results of Woland's visit] had been the work of a gang of hypnotists and ventriloquists with a superb

Anna E. Shoemaker

command of their art."[37] Even Ivan Bezdomny "knows that as a young man he fell victim to criminal hypnotists and was afterwards treated and cured."[38] Yet Ivan still senses that "there are some things he cannot manage" to explain or completely forget, as he experiences flashbacks to his adventures each spring.[39] The truth is manipulated and falsified by those in power, despite the awareness of the general population that these falsifications exist.

This awareness of truth leads to one of the central themes of *The Master and Margarita*: cowardice. Characters in both storylines are guilty of this failing. Yeshua's only utterance during his execution is that "among human vices he considered cowardice one of the first."[40] This is the sin that haunts Pilate in his dreams, during which he tells Yeshua, "No, philosopher, I disagree with you: it [cowardice] is the most terrible vice!"[41] The phrase is even repeated a third time by Matthew Levi in his notes of Yeshua's teachings. Riita Pittman claims that the Master is also guilty of cowardice. As an author of a work about the historical Jesus, his abandonment of his novel in the face of critical opposition could be considered to constitute cowardice. Pittman asks, "Is the Master's attempt to burn his manuscript justifiable? Is his failure to defend his story in the face of adverse criticism tantamount to a betrayal of the truth? Is the Master guilty of cowardice?"[42] The novel's conclusion states that the Master "does not deserve the light, he deserves peace,"[43] "light" being a sort of "heaven" and "peace" being a sort of purgatory. He must, therefore, be guilty of some sin, and cowardice is the only sin posited for the Master in the novel.

All of these themes reveal aspects of the reality in which Bulgakov was living as he composed this work, and, like the puppeteers using their dualistic *vertep* as social commentary, he uses this "doubled" novel to make a point about the plight of artists living in the Soviet Union under Stalin. Tomislav Loginovic groups *The Master and Margarita* into his category of "borderline poetics," literature created by Slavic writers under totalitarian regimes. According to Loginovic, these writers used "experimental prose [to question] the limits of realism and [examine] the effects ideology has on the

formation of identity."[44] Loginovic claims:

> The parodic positioning and eccentric individuality of the borderline writer could have its origins in the Slavic popular tradition of the holy fool (*urodivoy*). This cultural heritage is an ancient model for the borderline poetics practiced by the intellectual dissent, taking the side of the uncanny, distorted, and abject part of 'reality' in order to demonstrate 'the unreal, deceptive nature of the external environment.'"[45]

This source is especially applicable when one considers the characters in *The Master and Margarita* who resemble the "Holy Fool," namely Yeshua and Ivan Bezdomny. The use of such characters in both plotlines of a book composed of structurally experimental prose indicates that Bulgakov's literary intentions fall into Lognivoic's category of borderline poetics.

Techniques that Loginovic considers specific to such borderline poetics are present throughout *The Master and Margarita*, many of which also correspond to elements of *vertep*. These techniques include the use of a compositional structure that transcends the boundaries of literary genres, which can easily be seen in Bulgakov's dual plotlines and his strong reference to the theater.[46] A strong sense of unreality is derived from *The Master and Margarita*, which, according to Loginovic, reflects the "instability of the political and historical destiny of the region" of the writer.[47] This unreality is often reinforced by the presence of madness, dream sequences, or demonic possession within the work.[48] Madness marks nearly a dozen characters within the novel, notably the Master and Bezdomny, but additionally many minor characters that encounter Woland. Loginovic also notes that irony, parody, and black humor are also often present to aid in exorcising political horror.[49] The exploits of Woland's gang are obvious examples of such techniques, which create parodied views of life under Stalin in addition to providing a more slapstick genre of humor. Additionally, an uncertain sense of closure is usually found in borderline poetics, as the author has to avoid replacing the ideology of the state with another.[50] Such uncertainty is certainly reflected in the criticism of *The Master and Margarita*, which

continues to develop and is often quite contradictory. Critics cannot even agree who is the hero in this novel. The uncertain fates of the characters at the novel's close also exemplify this characteristic.

In light of its relation to other works composed by Slavs under totalitarian regimes, it is clear that Bulgakov was trying to explore his own reality of Soviet Moscow in relation to that of other times and places. As Proffer stated in her biography of Bulgakov, "for all his imaginative powers, Bulgakov is a writer wedded to social and political reality. . . what interests Bulgakov is the way in which [other social and political realities] are like his own."[51] Without the contrast of Yershalaim and the events that take place there, *The Master and Margarita* is left teetering precariously between the fanciful and the absurd. Only through its juxtaposition with the second storyline—which is historically based, and whose style and content offer a semblance of stability and reality to the reader—does Moscow becomes a real place. For an individual like Bulgakov, living through events just as disconcerting as a visit from the devil to Moscow, it is likely the world he envisioned of Pontius Pilate took on a reality of its own for him as well. Yershalaim offered this alternate "more realistic" reality to the characters within *The Master and Margarita*, for whom the story took on immense importance. Margarita became "absorbed" by the novel,[52] the Master looses his mind from troubles stemming from the novel's rejection, and Ivan Bezdomny is locked away in the asylum for his ravings about Pontius Pilate.

To say that Bulgakov was just writing about Soviet Russia, however, is to dismiss the breadth and scope of his literary accomplishment. Proffer claims:

> Bulgakov had a much greater ambition, namely to put the cataclysmic events of his life, the Russian Revolution, and the rule of Stalin, in historical perspective, and to reaffirm the connection of post-revolutionary Russian literature to the great traditions of the nineteenth century.[53]

This desire provides a reason why Bulgakov would choose to contrast Moscow with Yershalaim at "the moment when sacred and profane histories merge."[54]

To connect Moscow with a moment that took place in both secular and religious history, a moment so greatly examined and widely known, was not just an attempt to connect his present to all of history and literature. Such a comparison, and such a strong assertion that Jesus existed, struck down the atheistic beliefs espoused by Bulgakov's contemporary intellectuals and the Soviet state. This assertive juxtaposition made a reality in which Jesus existed appear more real than that of contemporary Moscow.

The inclusion of the Yershalaim storyline also provides a foundation for the novel's apocalyptic conclusion, which incorporates yet again a concept that is both secular and religious in function. As Weeks asserts:

> The notion of apocalypse (as seen in the final two chapters) and the notion of history are synonymous, inasmuch as the apocalypse is the symbol of the eschatological notion of history espoused by Christians. The very meaning of apocalypse implies a conflict between history as men experience it and see themselves participating in it, and History as directed by some external authority.[55]

By employing the concept of apocalypse in the Moscow storyline, Bulgakov yet again emphasizes the unreality of his own times by placing the biblical concept not in the "gospel" storyline, but in the storyline of "reality." The notion of apocalypse also implies the judgment Bulgakov passes on contemporary Moscow: if things continue in the present vein, Moscow will lead to its own destruction at the hands of History.

Ellendea Proffer derives a further conclusion from the similarities between the two plotlines:

> The point of the Master's novel is also the point of Bulgakov's novel: there exist moral absolutes, concepts which are unaffected by revolutions or tyrannies, and one ignores them at one's peril. [Bulgakov's] focus is on those who know better, those who hear the voice of conscience but stifle it in the name of political expediency.[56]

Who better to exemplify the Muscovites' stifling of the conscience than he who stifled his own conscious for political reasons, and ended by crucifying Christ? This side of Pilate is well explored, if not as widely known, and

Anna E. Shoemaker

a convincing argument can be made from even the canonical gospel texts alone that Pilate was sympathetic to the plight of Jesus of Nazareth[57]. Bulgakov's conclusion, in which he rewards Pilate's suffering and remorse for his actions with "light" while leaving the Master in a state of "peace," seems to imply that the Muscovites are more guilty in their actions than even Pilate. With such examples of cowardice as Pilate's before them, how could they still behave themselves in this manner? The juxtaposition of the two storylines thus serves to condemn Bulgakov's Soviet comrades for their cowardly behavior, and asserts that despite the current situation, there are moral absolutes to which one must adhere.

The Soviet government forbid *Vertep* for its inclusion of both religious references and social commentary.[58] Just as *vertep* communicated moral lessons and social criticisms through the juxtaposition of scenes from the life of Christ and humorous scenes from daily life, the two plotlines of Bulgakov's equally forbidden novel relate his moral, political, and historical message to his readers. This similarity was noted in 1987 by a Russian director-and-designer team, who adapted the novel to the stage in a production that recalls that of Wyspianski's *Wesele*.

> The director of this four-hour show divided his characters into three groups: manipulators, puppets and humans. There were representatives of the supernatural world, influencing human destiny alongside the NKWD (political police) officers: these were the manipulators. The puppets represented the human characters under the control of the supernatural (devilish) or political powers. The human actors—such as those playing the Master and Margarita—were characters defending their independence. Only one character changed his status in the course of the story: Ivan Bezdomny (Ivan the Homeless) was a puppet at the beginning of the story, but because of his growing humanity, became human by the end.[59]

As a native of the Ukraine in the early 1900s, it is highly probably that Bulgakov would have been familiar with *vertep*. But even if Bulgakov did not have *vertep* in mind when he was composing *The Master and Margarita*, it has proved a viable medium for adapting the novel theatrically, especially as a means for examining the development of the main characters. A discussion of *The*

Master and Margarita in the context of *vertep* and its influence also allows for a greater understanding of the Slavic traditions upon which Bulgakov drew in this novel: traditions of social criticism through comic theatrical performance, and above all, through the use of metaphor. As Henryk Jurkowski wrote, "Speaking through metaphors is the only way to remain within the frontiers of art, avoiding the descent into political manifestation and complaint."[60] The Soviets found the metaphors in Bulgakov's writing and in *vertep* theater too powerful to permit into the public sphere until the stranglehold of the State slightly loosened from the late nineteen-sixties onward. From that time on, such metaphors have spoken volumes of the Slavic experience under totalitarian regimes. A greater familiarization with the means by which these metaphors are communicated will ensure not just a more comprehensive understanding of the Slavic experience, but will also affirm that literature and art created under totalitarian systems must be addressed in a plurality of models to reveal the scope of their content. The art created under totalitarian regimes becomes a hybrid of what came before, during, and after the regimes. *The Master and Margarita* is but one example of the complexity of this genre, but as an example, it amply reveals the variety of tools necessary to adequately examine it in its entirety.
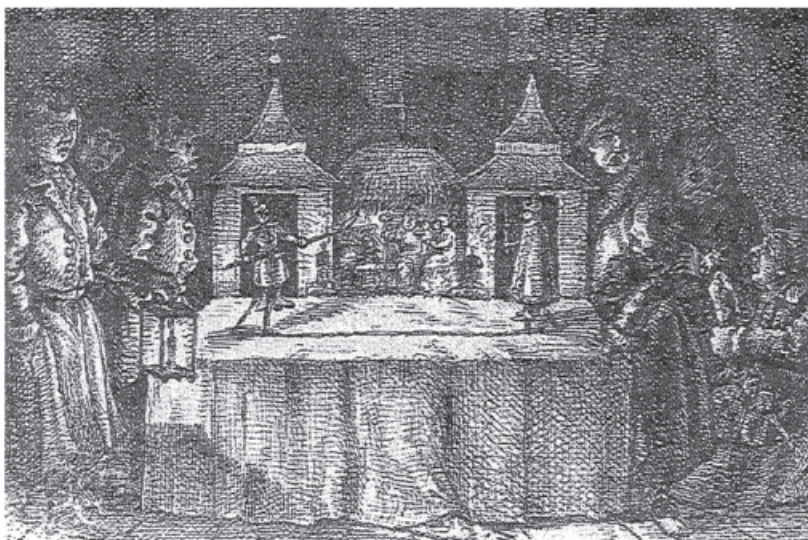
46. "Szopka" from Medyka, by Kajetan W. Kielsinski, etching, 1837. The oldest icongraphic document of a Polish szopka.
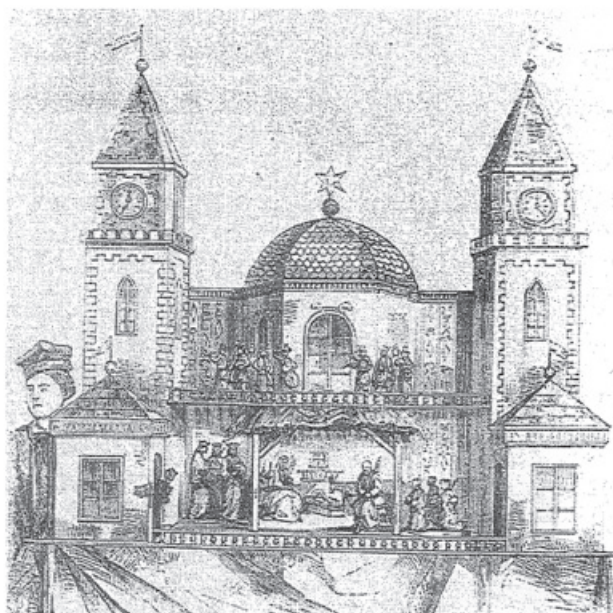
47. The Krakow "szopka".
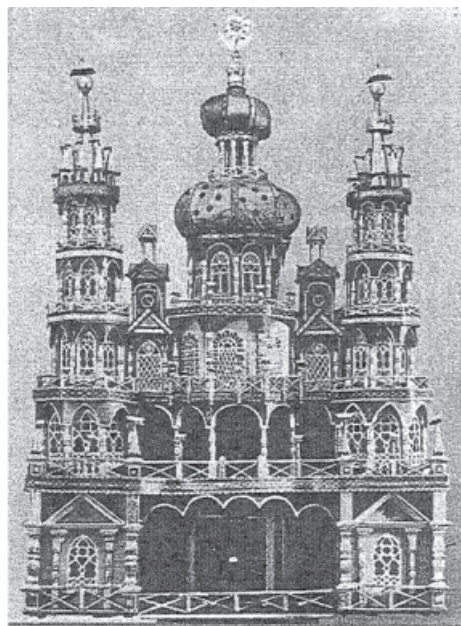
48. The Krakow "szopka" of the Ezenekier family.

49. Detail from the stage of the "szopka" of the Ezenekier family.

50. Three Magi from a Bielorussian "batleyka".
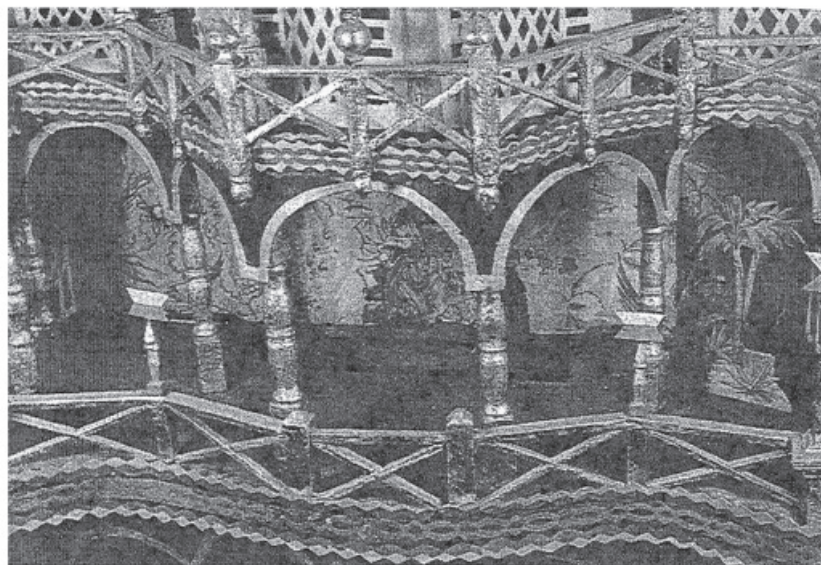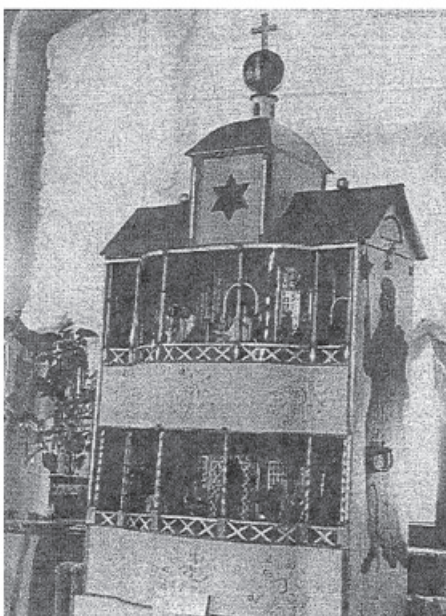(Photo: Juliusz Wolski).

51. Bielorussian "batleyka".

Anna E. Shoemaker

Endnotes

1 Loginovic, Tomislav Z. *Borderline Culture: The Politics of Identity in Four Twentieth-Century Slavic Novels.* Fayetteville: The University of Arkansas Press, 1993. 38.

2 Kelly, Catriona. *Petrushka: Russian Carnival Puppet Theater.* Cambridge: Cambridge University Press, 1990. 54.

3 Jurkowski, Henryk. *A History of European Puppetry.* Vol 1. Lewiston: Edwin Mellen Press, 1996. 298.

4 Jurkowski, Henryk. *A History of European Puppetry.* Vol 1. 66.

5 Jurkowski, Henryk. *A History of European Puppetry.* Vol 1. 67-68.

6 Jurkowski, Henryk. *A History of European Puppetry.* Vol 1. 232.

7 Jurkowski, Henryk. *A History of European Puppetry.* Vol 2. Lewiston: Edwin Mellen Press, 1998. 286.

8 Jurkowski, Henryk. A History of European Puppetry. Vol 1. 2.

9 Jurkowski, Henryk. *A History of European Puppetry.* Vol 2. 287.

10 Jurkowski, Henryk. *A History of European Puppetry.* Vol 1. 298.

11 Jurkowski, Henryk. *A History of European Puppetry.* Vol 1. 298.

12 Jurkowski, Henryk. *A History of European Puppetry.* Vol 1. 234.

13 Jurkowski, Henryk. *A History of European Puppetry.* Vol 1. 234.

14 Jurkowski, Henryk. *A History of European Puppetry.* Vol 1. 235.

15 Jurkowski, Henryk. *A History of European Puppetry.* Vol 1. 237.

16 Jurkowski, Henryk. *A History of European Puppetry.* Vol 1. 298.

17 Vertep (Christmas Puppet Theater). Embassy of Ukraine. *www.ukraineinfo.us/culture/christmas.html#3.* 14 November 2005.

18 Jurkowski, Henryk. *A History of European Puppetry.* Vol 1. 299.

19 Revutsky, Valerian. Encyclopedia of Ukraine. "Vertep." *www.encyclopediaofukraine.com.* 14 November 2005.

20 Vertep (Christmas Puppet Theater).

21 Jurkowski, Henryk. *A History of European Puppetry.* Vol 2. 61.

22 Jurkowski, Henryk. *A History of European Puppetry.* Vol 2. 182.

23 Jurkowski, Henryk. *A History of European Puppetry.* Vol 2. 183-184.

24 Jurkowski, Henryk. *A History of European Puppetry.* Vol 1. 299.

25 Revutsky, Valerian. Encyclopedia of Ukraine. "Vertep." *www.encyclopediaofukraine.com.* 14 November 2005.

26 Loginovic, Tomislav Z. 39.

27 Weeks, Laura D. *The Master and Margarita: A Critical Companion.* Evanston: Northwestern University Press, 1996. 11.

28 "The Russian Orthodox Church: Not Spreading the Faith." *The Economist.* Vol 357, Issue 8202. 21 December 2000. 65-67.

29 Weeks, Laura D. 3.

30 Weeks, Laura D. 11.

31 Proffer, Ellendea. *Bulgakov: Life and Work.* Ann Arbor: Ardis, 1984. 538.

32 Loginovic, Tomislav Z. 39.

33 Bulgakov, *The Master and Margarita.* New York: Penguin Books, 1966, 1997. 384, 396.

34 Bulgakov, Mikhail. 288.

35 Loginovic, Tomislav Z. 43.

36 Weeks, Laura D. 28-29.

37 Bulgakov, Mikhail. 385

38 Bulgakov, Mikhail. 393.

39 Bulgakov, Mikhail. 393.

40 Bulgakov, Mikhail. 305.

41 Bulgakov, Mikhail. 319.

42 Pittman, Riita H. *The Writer's Divided Self in Bulgakov's "The Master and Margarita."* New York: St. Martin's, 1991. 17.

43 Bulgakov, Mikhail. 361.

44 Loginovic, Tomislav Z. 23.

45 Loginovic, Tomislav Z. 28.

46 Loginovic, Tomislav Z. 27.

47 Loginovic, Tomislav Z. 26.

48 Loginovic, Tomislav Z. 27.

49 Loginovic, Tomislav Z. 27.

50 Loginovic, Tomislav Z. 23.

51 Proffer, Ellendea. 249.

52 Bulgakov, Mikhail. 142.

53 Proffer, Ellendea. 564-565.

54 Weeks, Laura D. 49.

55 Weeks, Laura D. 46-47.

56 Proffer, Ellendea. 565.

57 Wroe, Ann. *Pontius Pilate.* New York: Random House1999. 230-235.

58 Jurkowski, Henryk. *A History of European Puppetry.* Vol 2. 407.

59 Jurkowski, Henryk. *A History of European Puppetry.* Vol 2. 410.

60 Jurkowski, Henryk. *A History of European Puppetry.* Vol 2. 411.

Handheld Device for Remotely Measuring Brain Function

Adam Wang—*Electrical Engineering*

Abstract

Infants, and especially premature infants, are carefully monitored in their incubators, and their brain health is of great concern. While the currently preferred procedure is to attach a device to a patient's forehead to monitor pulse rate and oxygen levels in the brain, the development of a remote handheld system would make it possible to spot check patients from a distance, without having to deal with the obtrusiveness of attaching probes. Additionally, past studies have shown that measuring changes in blood volume and oxygen levels in the forebrain can be used to study brain function. Current work suggests that affordable and safe handheld devices for contact systems can be built from inexpensive components, such as LEDs and photodiodes, to measure these parameters.

The goals of this project were not only to assess the feasibility of remotely sensing brain function with a handheld device based on simple near-infrared technology, but also to suggest a design for such a device. As part of the project, experiments have been designed to advance the development of a handheld remote sensing device that would measure indicators of brain health and function, such as the arterial pulse and changes in blood volume and oxygen levels in the brain. In this study, remote sensing demonstrated promising results for use over small distances. In addition to describing this process, this paper includes suggestions for extending remote sensing over greater distances for use in more practical, real world situations.

Introduction

Approximately nine percent of babies are born prematurely, and many need special care in a Neonatal Intensive Care Unit (NICU) [1]. Since they are not fully developed, premature babies frequently suffer from health problems and are carefully monitored in their incubators. Common health concerns include apnea, anemia, respiratory distress syndrome, and patent ductus arteriosus, of which pulse rate, blood volume, and oxygen levels are generally key indicators [2]. Furthermore, several recent studies have suggested that premature babies are impacted negatively by noise, light, and activity in their NICUs. Thus, many hospitals take care to maintain a quiet environment, shield the babies from light, and handle them slowly and deliberately [1]. A handheld remote sensing device for measuring a baby's brain health parameters, such as blood volume and oxygen levels, would provide a more discreet and desirable method for assessing these health indicators.

In addition to monitoring the brain health of babies, such a device could also be used to evaluate other brain functions. Current devices use contact methods for monitoring brain function during tasks such as problem solving or lying. A probe placed on the subject's forehead measures changes in blood volume or oxygen levels in certain regions of the brain [3]. These changes correspond to brain activity in the region of the brain measured. If, for example, a certain region of the brain

displays increased blood volume after a subject tells a lie, monitoring this region for changes in blood volume can determine whether the subject tells a lie again. Moreover, determining which parts of the brain are more active, as determined by changes in blood volume or oxygen levels during certain tasks, can map out overall brain function. Past tests determined that certain areas of the forebrain show activity during lying or solving anagram puzzles. A handheld device for remotely measuring brain function would offer portability and convenience to both the user and the subject [3].

Background

Near-infrared (NIR) light has been used in recent years to non-invasively measure optical properties of tissue [4]. In particular, NIR light with a wavelength of 700-900 nm has optimal properties for measuring blood volume and oxygenation levels since most tissue—other than oxygenated and deoxygenated hemoglobin—absorb little light at these wavelengths. These qualities permit deep light penetration and backscattering from light sources such as white light, lasers, or LEDs, which can be measured by light detectors, such as photomultiplier tubes or photodiodes.

In addition to its noninvasiveness, the low-cost and convenience of NIR imaging has attracted much recent interest [5]. Some current NIR spectroscopy applications include brain functional imaging, breast cancer imaging, and muscle activity monitoring. For example, the finger pulse oximeter revolutionized hospital care with its ability to monitor arterial oxygen saturation and pulse rate while still being portable, noninvasive, and capable of providing continuous real-time monitoring [6].

Given the success of contact model NIR devices, some current studies are moving toward remote sensing using NIR light. The goal of these studies is to combine the functionality of contact model NIR devices with the versatility of remote sensing, which might eliminate the subject's awareness of being tested altogether. One such opportunity is brain oxygen monitoring in premature infants. The use of such a device could be extended by measuring brain function to monitor comatose or brain-injured patients.

Adam Wang

Depending on the need, various NIR systems—including continuous-wave systems and time-resolution systems—can be used. A time-resolution system can measure the absorption coefficient and reduced scattering coefficient of tissue to find the absolute concentrations of deoxygenated hemoglobin (Hb) and oxygenated hemoglobin (HbO$_2$). From these two values, the absolute blood volume and oxygenation level can be determined [5]. However, this method relies on accurate measurement of photon arrival times that are backscattered from the tissue [7]. A much simpler method for NIR spectroscopy is a continuous-wave system, which merely emits a constant light at tissue and measures the backscattered light. An important limitation of a continuous-wave system is that it can only measure changes in Hb and HbO$_2$. Thus, only changes in blood volume and changes in oxygen levels can be measured [5].

All NIR methods use a NIR light source to illuminate tissue. The tissue surface reflects some of the incident light, while the rest enters the tissue. Light inside the tissue is either absorbed or scattered about until some of it reemerges. A light detector then measures this backscattered light. In the case of a continuous-wave system, the change in backscattered light is measured. Since Hb and HbO$_2$ are the greatest sources of absorption of the NIR light used, only these two are considered in the Beer-Lambert Law

$$I = GI_0 e^{-(\varepsilon_{Hb}C_{Hb} + \varepsilon_{HbO_2}C_{HbO_2})L} \qquad (1)$$

where I is the light intensity after absorption and backscattering; G is a constant attenuation; I_O is the input light power; ε_{-Hb} and ε_{-HbO_2} are the molar extinction coefficients of deoxygenated and oxygenated hemoglobin, respectively, and determine the amount of light absorbed; C_{Hb} and C_{HbO_2} are the concentrations of deoxygenated and oxygenated hemoglobin, respectively; and L is the photon path length, which can be determined experimentally and varies based on the particular setup of the light source and detector [5]. Note that ε_{-Hb}, ε_{-HbO_2}, and L are all functions of the light wavelength used. Water and other tissues have absorption coefficients

that are orders of magnitude lower than Hb or HbO$_2$ and are not considered when dealing with NIR light [6].

A continuous-wave system needs at least two light sources of different wavelengths to operate. Let I_{-760} be the backscattered light intensity from a 760 nm light source at a predetermined "baseline state," and let C_{-Hb} and C_{-HbO_2} be the unknown concentrations of deoxygenated and oxygenated hemoglobin concentrations at the baseline state. Then, for a light source of 760 nm, let the optical density (OD) be

$$
\begin{aligned}
OD_{760} &= \ln\frac{I_{-760}}{I_{760}} = \ln\frac{GI_0 e^{-(\varepsilon_{Hb,760}C_{-Hb} + \varepsilon_{HbO_2,760}C_{-HbO_2})L_{760}}}{GI_0 e^{-(\varepsilon_{Hb,760}C_{Hb} + \varepsilon_{HbO_2,760}C_{HbO_2})L_{760}}} \\
&= \left(\varepsilon_{Hb,760}\left(C_{-Hb} - C_{Hb}\right) + \varepsilon_{HbO_2,760}\left(C_{-HbO_2} - C_{HbO_2}\right)\right)L_{760} \\
&= \left(\varepsilon_{Hb,760}\,\Delta C_{Hb} + \varepsilon_{HbO_2,760}\,\Delta C_{HbO_2}\right)L_{760}
\end{aligned} \qquad (2)
$$

where I_{760} is the backscattered light intensity from the 760 nm light source at some other state, and ΔC_{Hb} and ΔC_{HbO_2} are the changes in concentrations of deoxygenated and oxygenated hemoglobin, respectively, from the baseline state. Similarly,

$$OD_{830} = \ln\frac{I_{-830}}{I_{830}} = \left(\varepsilon_{Hb,830}\,\Delta C_{Hb} + \varepsilon_{HbO_2,830}\,\Delta C_{HbO_2}\right)L_{830} \qquad (3)$$

Thus, by using two wavelengths, the amount of change in Hb and HbO$_2$ from the baseline state can be determined:

$$\Delta C_{Hb} = \frac{\varepsilon_{HbO_2,830}\,OD_{760}/L_{760} - \varepsilon_{HbO_2,760}\,OD_{830}/L_{830}}{\varepsilon_{Hb,760}\varepsilon_{HbO_2,830} - \varepsilon_{HbO_2,760}\varepsilon_{Hb,830}} \qquad (4)$$

$$\Delta C_{HbO_2} = \frac{\varepsilon_{Hb,760}\,OD_{830}/L_{830} - \varepsilon_{Hb,830}\,OD_{760}/L_{760}}{\varepsilon_{Hb,760}\varepsilon_{HbO_2,830} - \varepsilon_{HbO_2,760}\varepsilon_{Hb,830}} \qquad (5)$$

Then, the change in blood volume, ΔBV, and the change in oxygenated blood, ΔOXY, can easily be computed:

$$\Delta BV = \Delta C_{Hb} + \Delta C_{HbO_2} \qquad (6)$$

$$\Delta OXY = \Delta C_{HbO_2} - \Delta C_{Hb} \qquad (7)$$

Adam Wang

Since the molar extinction coefficients of Hb and HbO_2 are the same for light with wavelength 800 nm [Fig. 1], if a third light source is used with a wavelength at this isosbestic point, then the change in blood volume, $\_BV$, could be validated by OD_{800} alone [5].

Project Goals

The goals of this project were not only to assess the feasibility of remotely sensing brain function with a handheld device based on simple near-infrared technology, but also to suggest a design for such a device.

Figure 1 | *Molar extinction coefficients of Hb and HbO$_2$, as a function of wavelength, plotted on a logarithmic scale [8].*

Molar Extinction Coefficients of Hb and HbO$_2$
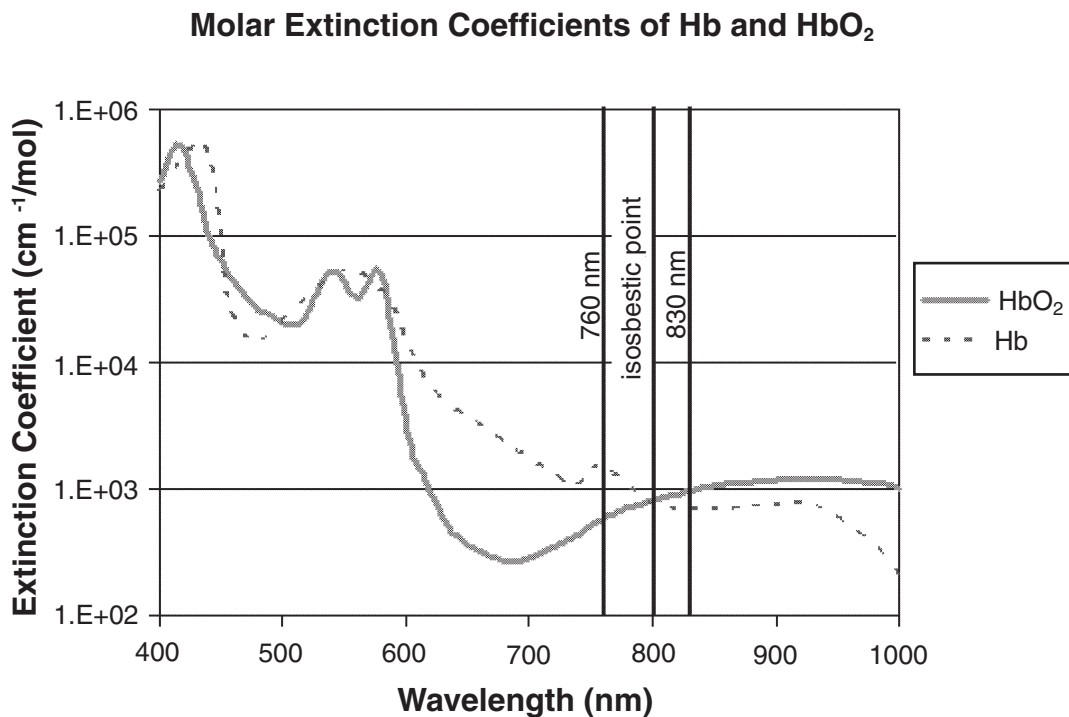
When monitoring brain health, a handheld remote sensing device needs a predetermined "healthy state" to use as the baseline state so that changes or deviations in blood volume or blood oxygenation levels can be tracked to gauge brain health. Depending on the case, a tolerable range of fluctuation or deviation should be allowed, but once the change in blood volume or oxygenation levels from the predetermined healthy state exceeds these bounds, the device should alert the device operator. Similarly, when monitoring brain activity, a predetermined "resting state" could be used as the baseline state, and any changes in blood volume or blood oxygenation levels in a region of the brain may indicate variations in brain function.

While current procedure attaches a device to a patient's forehead for monitoring pulse rate and brain oxygen levels, the development of a remote handheld system would make it possible to spot check patients from a distance, without having to deal with the obtrusiveness of attaching and maintaining probes. Current work suggests that affordable and safe handheld devices for contact systems can be built from inexpensive components such as LEDs and photodiodes [3]. These components are small, safe, and operate at low voltage levels. A general system that can remotely measure brain function must include a light source and driver, a light detector, and a processing unit [Fig. 2].

Figure 2 | *Block diagram of handheld device for remotely sensing brain function.*
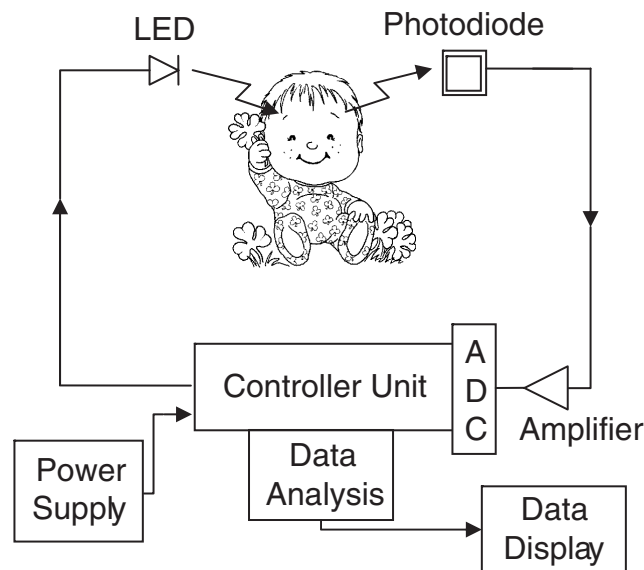
For this study, fundamental building blocks were selected to find the minimum requirements to build an effective, durable device at low cost. In addition to being small and very inexpensive sources of light, LEDs emit more diffuse light than lasers and, unlike lasers, are not subject to U.S. Food and Drug Administration power constraints when directed at humans [5]. Also, due to their ease of use and versatility, photodiodes were selected as light detectors for this study. Unlike photomultiplier tubes, they do not require high voltages and can operate at light levels that would easily damage highly sensitive photomultiplier tubes [9]. Finally, an ordinary programmable microcontroller was selected to control the LED, sample the photodiode response with its built-in analog-to-digital converter (ADC), and process the sampled data. The object of much of this study was to determine the efficacy of these components when used together.

Experimental Setup

As with any digital device, an analog signal must first be quantized before sampling. The continuous range of the analog signal is broken up into a finite set of discrete values, to which the analog signal is mapped. In the case of the ADC selected for this study, a 0-5 V signal was uniformly quantized into a 10-bit value (1024 quantization levels), with 0 corresponding to 0 V and 1023 corresponding to 5 V. Thus, the precision of this particular ADC is $5/2^{10}$, or approximately 5 mV.

If signal variation is too small, the chosen quantization levels may not detect such variations. In this situation, one solution is to extract the desired signal and amplify it to better fit the dynamic range of the ADC so that more accurate sampling can be achieved [Fig. 3].

Much of the circuit built to perform this exact function was constructed from operational amplifiers. The first key step to taking advantage of the ADC's full dynamic range was to remove the DC component so that only the component of the signal of interest was amplified. This was accomplished by subtracting a variable DC offset that ranged from 0 V to +5 V and could be adjusted by a potentiometer. Next, another op-amp configuration amplified the signal with a gain up to 100, adjusted by a 100 k_ potentiometer [Fig. 5].

A Motorola 9S12C32 microcontroller was programmed to sample at certain frequencies and to drive two LEDs (760 and 830 nm). For simplicity, the sampled data was stored and analyzed offline. As the

Adam Wang

| Figure 3 | The original small signal's DC component is first reduced, and then the signal is amplified to better match the ADC's full dynamic range. |

Original Small Signal

DC Offset Signal

DC Offset Signal

Amplified Signal
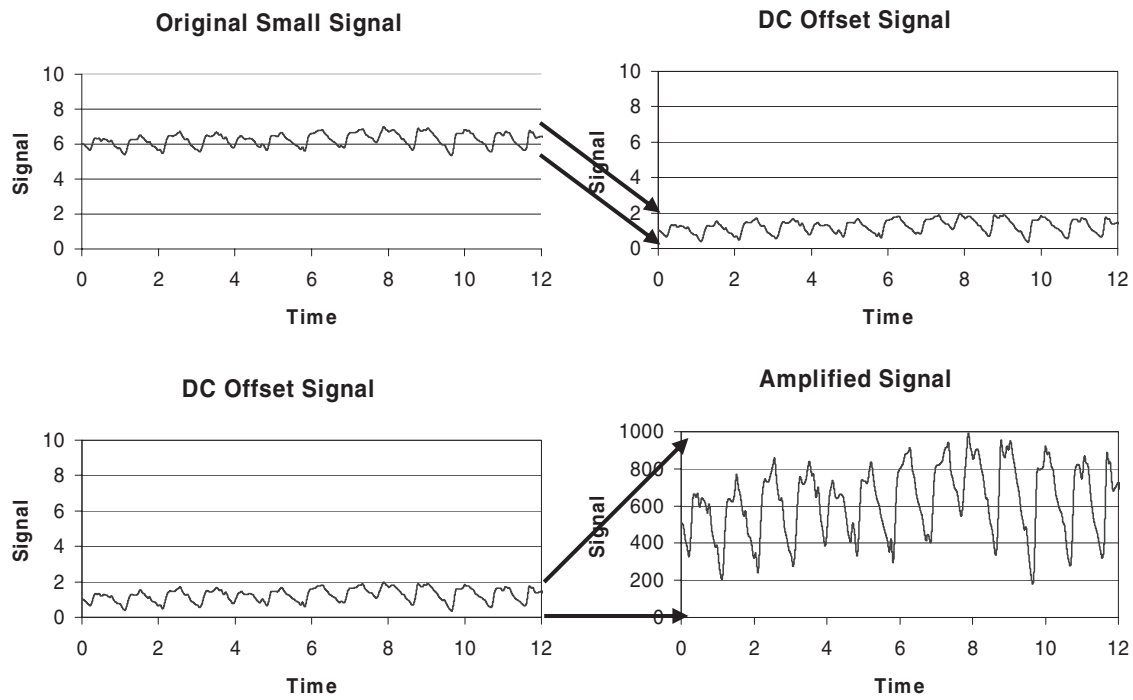
| Figure 4 | Wratten filter #89B's percent transmission curve as a function of wavelength [10]. The filter minimally affects the two LEDs used while shorter wavelengths (visible light) are blocked. |

Wratten Filter 89B

Adam Wang
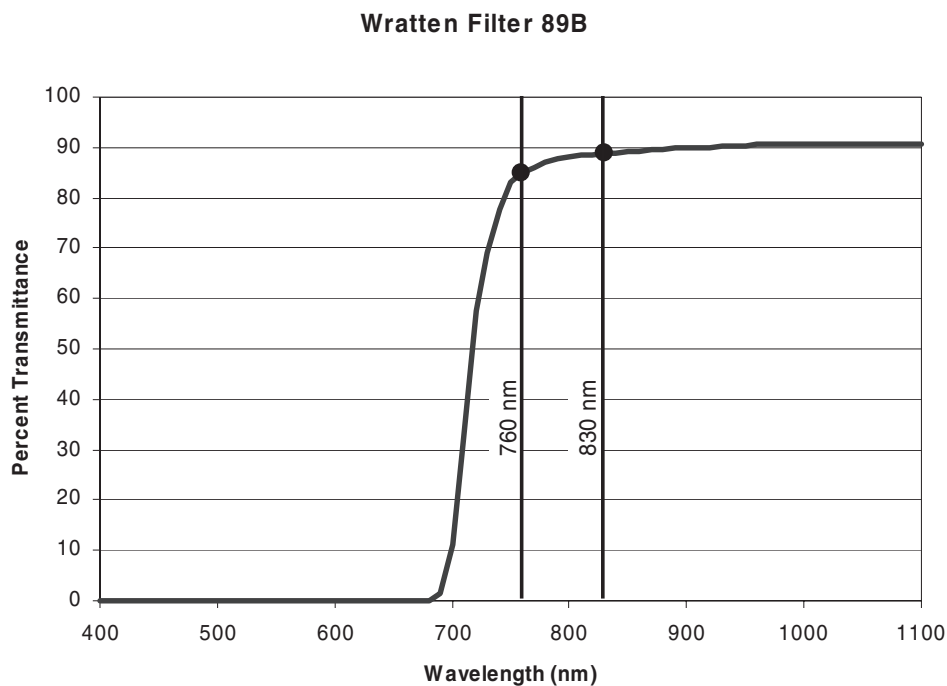
project progressed, the necessity of using higher quality components became apparent. A Wratten gelatin light filter #89B placed over the photodiode filtered out undesired wavelengths [Fig. 4] and greatly reduced noise (since photodiodes respond to a wide range of wavelengths).

Higher precision, lower noise OP27 op-amps replaced the more standard op-amps originally used (80 nV compared to 1000 nV peak-to-peak noise in the 0.1 to 10 Hz range), and the OPT101 photodiode was replaced by the FDS1010 Si photodiode (9.7 ◊ 9.7 mm), which has 18 times the surface area for collecting light. When the passive FDS1010 Si photodiode was inserted, the leads were fed through a difference amplifier with a pre-amp gain of 10 [Fig. 5]. Furthermore, a first order RC low-pass filter was added to reduce high frequency noise in the signal, such as the 60 Hz AC component in power from wall outlets and room light.

Experimental Results

Two aspects of remote sensing were explored: (1) measurement of the arterial pulse and (2) tracking of changes in an ink test's light absorption. The former sought to determine the human subject's pulse rate, while the latter sought to simulate changes in concentrations of Hb and HbO_2.

Arterial Pulse

Starting with the contact model, where both the LED and photodiode were in contact with the human subject (in this case the author), the arterial pulse was easily extracted. An LED shined light through the subject's thumbnail, and a photodiode measured the transmitted light through the thumb. The amount of blood at the tip of the thumb fluctuated in relation to the arterial pulse. This rhythmic fluctuation was the basis for determining the pulse rate. As the blood volume increased, the amount of light transmitted decreased due to the greater absorption of NIR light by Hb and HbO_2. The raw data was run through a digital fourth-order Butterworth low-pass filter to clean up the signal. Using a Discrete Fourier Transform, the pulse rate of the subject was determined over a sampling duration of 30 seconds to be 74 beats per minute, within 2 beats per minute [Fig. 6]. This was validated by counting the pulse rate via the subject's wrist. In this setup, the presence of room light was not a problem since it merely contributed to the transmitted light and was still modulated by the arterial pulse.

After showing that the contact model could be reproduced, a small-scale setup for remote sensing [Fig. 7] was built to test the feasibility of remote sensing. Neither the LEDs nor the photodiode were in

| **Figure 5** | *Circuit diagram showing op-amps (±12 V power supply) used to extract signal throughout this project.* |

Adam Wang
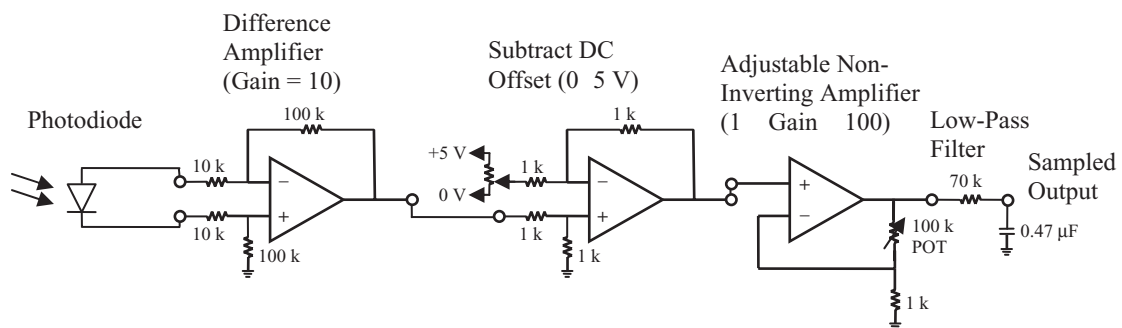
Figure 6 | *Measured transmitted light (left) through thumb with a contact setup, and its frequency spectrum (right), showing a pulse rate of approximately 74 beats per minute.*

Arterial Pulse of Thumb

DFT of Arterial Pulse
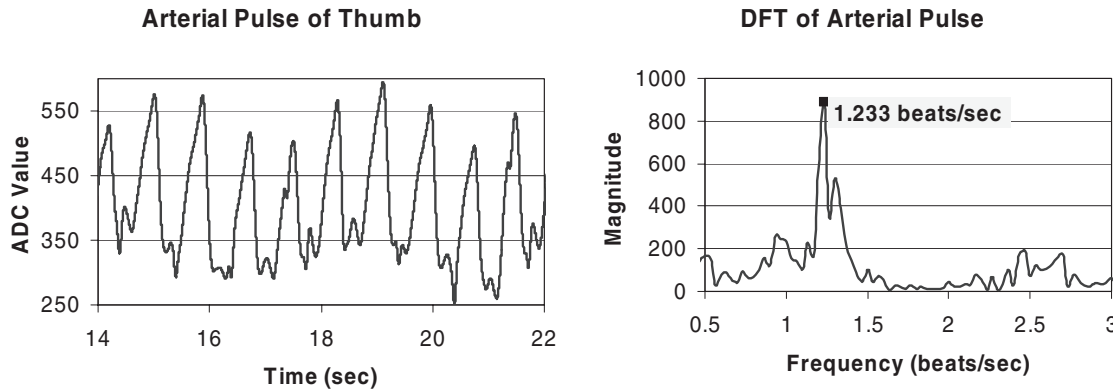
direct contact with the subject's hand tissue; instead, they were fixed a small distance from the tissue, as a step toward achieving distances more likely for remote sensing. A barrier prevented light from the LED from directly illuminating the photodiode and reduced reflected light from the surface of the tissue. Over the course of hundreds of tests, it was determined that extracting the arterial pulse from the measured backscattered light required: (1) higher quality op-amps, (2) low-noise OP27 chips, (3) the addition of a light filter, (4) the change to a greater surface area FDS1010 Si photodiode, and (5) the reduction of incident room light. Even with these enhancements, extracting the arterial pulse from this scaled down setup of remote sensing proved difficult because of the small signal and noise.

Figure 7 | *Small-scale remote sensing setup with neither the LED nor the photodiode in contact with the tissue.*

Small-Scale Remote Sensing
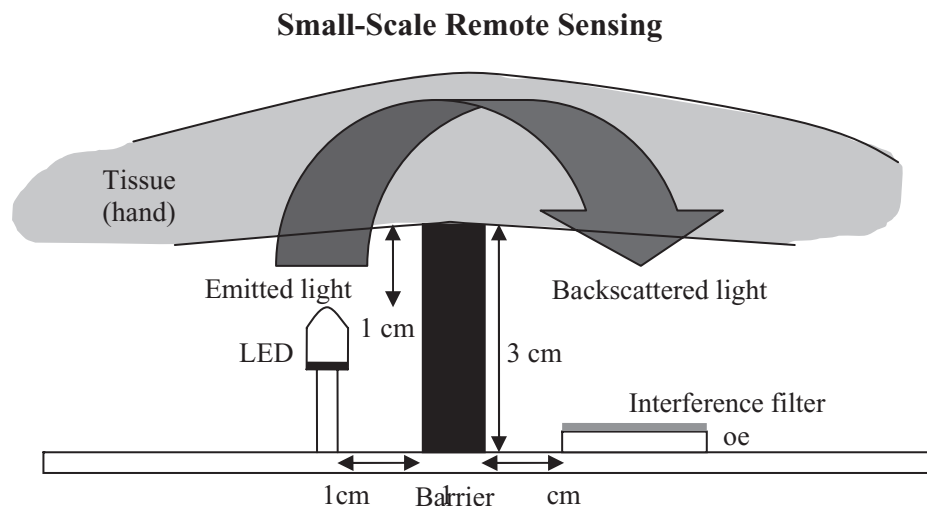
In order to utilize Equations 2 through 7, which can determine the change in blood volume or oxygen levels in tissue, two LEDs are necessary. To monitor the change in the intensity of backscattered light from each wavelength, the two LEDs cannot emit light simultaneously; instead, they must be time-shared and alternately emit light at a rate much greater than the rate of change in the parameter of interest. For finding the arterial pulse, the microcontroller flashed a 760 nm and an 830 nm LED alternately for a 100 ms duration each. The average backscattered light intensity was different for each LED, but the arterial pulse modulated the backscattered light for both wavelengths [Fig. 8].

tween the beaker and the light barrier, with both the 760 and 830 nm LEDs, and with and without the presence of room light.

The photodiode sampling for the ink tests differed from the arterial pulse tests since no specific frequencies were being tested for; only the general trend over a long time period was sought. For example, with the beaker 10 cm from the 760 nm LED and no room light, the measured backscattered light clearly decreased as 1 mL of ink was added approximately every minute [Fig. 10] to the 1 L of intralipid solution. The signal stabilized and showed little change before more ink was added to the solution, which was recorded by pressing a button

Figure 8 | *Measured backscattered light (left) from the author's hand with the small-scale remote sensing setup and two time-s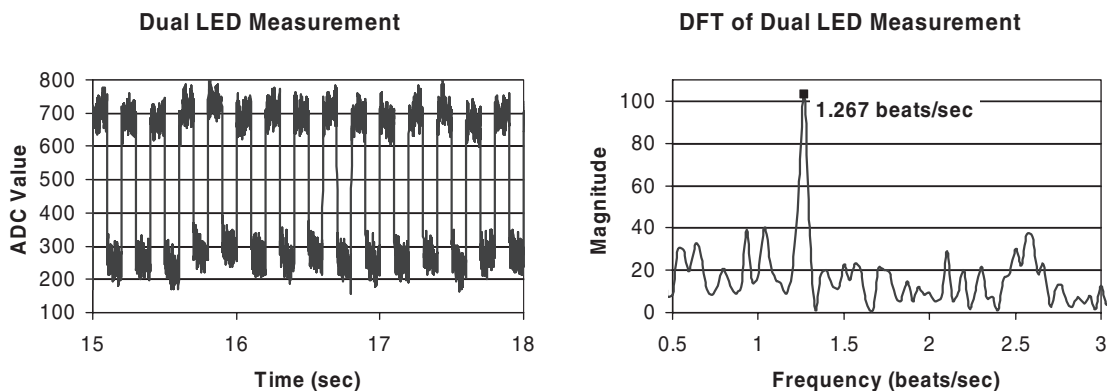hared LEDs; the frequency spectrum of the backscattered light (right), showing a pulse rate of approximately 76 beats per minute modulating the backscattered light.*

Dual LED Measurement

DFT of Dual LED Measurement

1.267 beats/sec

Ink Tests

In addition to remotely sensing a subject's pulse rate, another goal for the handheld device was to monitor changes in blood volume and oxygen levels in the brain. To achieve this, the handheld device must be sensitive to changes in concentrations of Hb and HbO_2, which can be simulated using an ink test. An intravenous fat emulsion solution in a large plastic beaker provided a scattering medium that allowed incident light to be backscattered, which a photodiode then collected [Fig. 9]. As ink was slowly added, light absorption increased. Various setups were tested, with varying distances be-

on the microcontroller's docking module and is displayed as vertical lines in Figure 10.

While the 760 nm and 830 nm LEDs showed similar effectiveness in the different ink test configurations, the increased distance between the beaker and LED and presence of room light greatly affected the system's ability to determine drops in backscattered light. Adding room light to the configuration used to obtain Figure 10 injected significantly more noise to the raw data and required twice the gain to extract a similar signal, even with the Wratten light filter [Fig. 11]. When the distance between the beaker and the LED was increased

Adam Wang

| *Figure 9* | Remote sensing ink test setup with handheld device components. |

| *Figure 10* | Ink Test with a 760 nm LED at a distance of 10 cm from the beaker with no room light. The backscattered light intensity drops with each addition of ink, marked by the red vertical lines. |
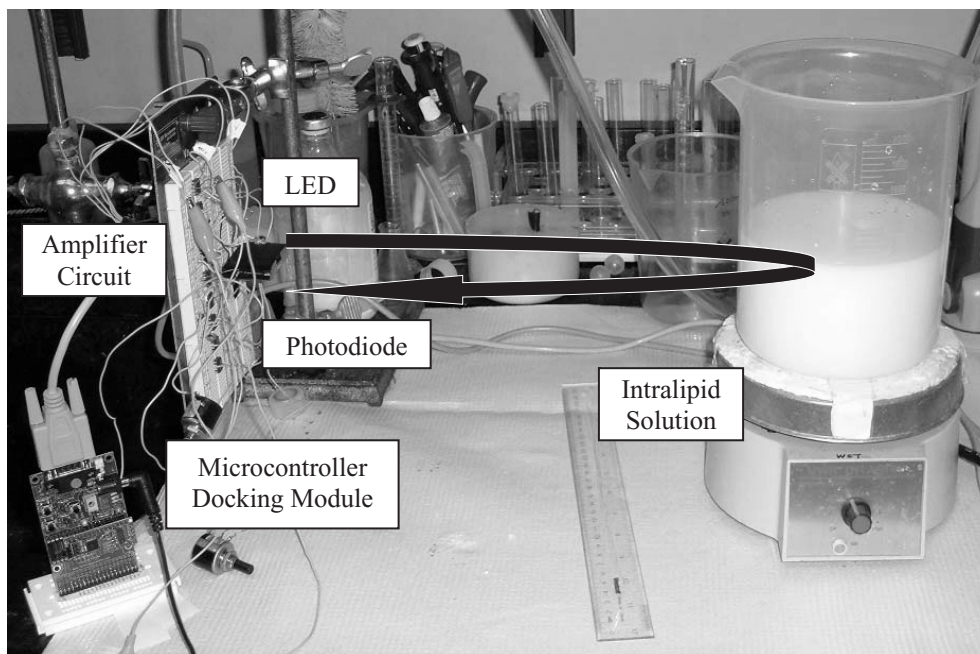
Ink Test (760 nm LED, 10 cm Distance, No Room Light)
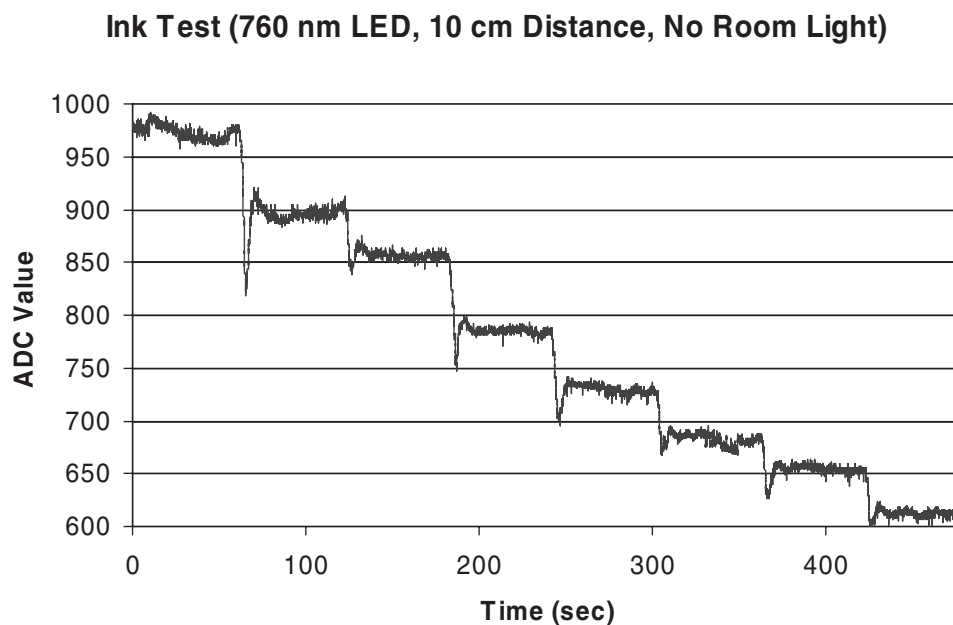
Figure 11 | *Ink test with room light present (left) and ink test at a greater distance (right) show a general degradation of the signal but still demonstrate a response to ink being added.*

Ink Test (760 nm LED, 10 cm Distance, Room Light)

Ink Test (760 nm LED, 30 cm Distance, No Room Light)
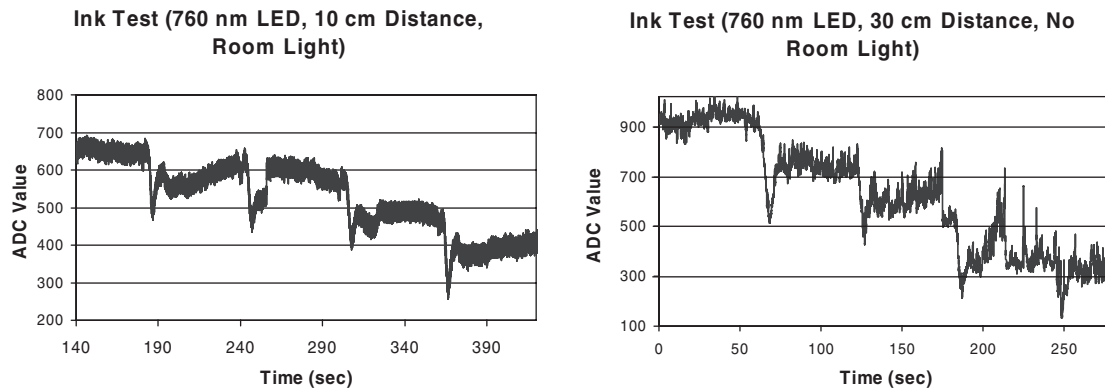
to 30 cm, the downward trend in backscattered light due to the addition of ink could no longer be measured. Even without room light, a distance of 30 cm significantly weakened the signal and required five times the gain to extract [Fig. 11]. When the distance was increased to 60 cm, the signal was lost.

Discussion

While the basic components of this study indicate some success in remote sensing of pulse rate and changes in absorption values, several improvements are needed. For example, even small-scale remote sensing detection of the arterial pulse required careful shielding from room light. The ink tests also showed that the distance between the subject and device, along with the presence of room light, greatly affected the ability to accurately track changes in light absorption. More realistic remote sensing scenarios for such a device might require distances of up to several feet between the subject and the device in the presence of room light. Weak signals exacerbate this problem of high sensitivity, requiring an extraordinarily stationary subject and operator. Furthermore, room light injects significantly greater noise, including a large 60 Hz component that must be diminished by a low-pass filter.

When two LEDs of different wavelengths are time-shared—as they must be to monitor changes in blood volume or oxygen levels—the backscattered light intensity from the two different wavelengths will differ. Therefore, if the same gain circuit amplifies the signal from both LEDs, the difference in backscattered light intensity between the two LEDs will be amplified, and both signals may not fit in the ADC's range. To solve this problem, the gain can be lowered to accommodate the different backscattered light intensities. However, this change causes the signal from each LED to be smaller too. Either the LED power should be adjusted so that the backscattered light intensities from the two LEDs are very close in value, or, better yet, each wavelength should have its own DC offset and gain circuit that the microcontroller could sample at the appropriate time.

Since the variable volume of blood due to the arterial pulse only accounts for 5% of the total blood volume in tissue [6], finding the arterial pulse from backscattered light would be much more challenging than tracking general changes in the overall trend of blood volume and oxygen levels in the brain. The ink tests demonstrated the *qualitative* ability of the handheld device components to track changes in absorption values; however, a *quantitative* indicator is needed. To accomplish this, effective blood tests must closely simulate the scattering and absorption coefficients of human tissue and quantitatively calibrate a system with

Adam Wang

known changes in the absorption coefficients at the wavelengths used.

Toward the end of the experiments, the issue of bandwidth and shot noise arose. Inherent to photodiodes, shot noise is proportionally related to the square root of the bandwidth of the photodiode [5]. Therefore, in order to limit shot noise, the bandwidth of the entire system should be limited to the signal of interest, which, at its greatest frequency, would probably not exceed a few hertz. Some tests showed that simply adding a capacitor in parallel to the negative feedback resistor of the difference amplifier could significantly reduce shot noise. The system originally built did not take shot noise into consideration and had a bandwidth of approximately 600 Hz. Adding a 47 μF capacitor would reduce the bandwidth to approximately 1 Hz but would still allow for the detection of the arterial pulse while reducing shot noise by almost a factor of 25 [Fig. 12]. Further tests are required to determine the effects of shot noise.

Recommendations

Due to the weak signal and the high level of noise inherently present in remote sensing, sources of noise should be targeted aggressively—clean power sources and low-noise components are essential. The Wratten gelatin filter used was a good start for reducing unwanted room light, but a better, more expensive solution would be a narrow bandpass lens coating that would exclusively pass the wavelengths of the LEDs. Furthermore, the gelatin filter is sensitive to cleaning, and a handheld remote sensing device should be durable and hardy, considering the more rigorous uses intended outside of the lab.

A weak signal entails the need for a high gain circuit, which also unfortunately amplifies sampled noise. Thus, various measures could be enacted to obtain a stronger signal, such as reducing ambient lighting, increasing the LEDs' power, minimizing the distance between the device and the subject, and maximizing the collecting area of the backscattered light. An array of LEDs or photodiodes could be explored as a possibility to increase the effective illumination intensity or effective collecting area of the handheld device. A simple way to increase the effective collecting area of the photodiode would be to use a large lens to collect light and focus it onto the photodiode.

Since the change in blood volume or oxygen level in the brain is determined by the change of measured backscattered light intensity, care must be taken to limit any movement of the subject or the operator of the handheld remote sensing device. Slight angular movements of either the device or the subject could change the incident light intensity from the LED onto the subject or backscattered light intensity from the

Figure 12 | System built for this study (left) with a bandwidth of about 600 Hz that can be improved by filtering noise early on by limiting the bandwidth.
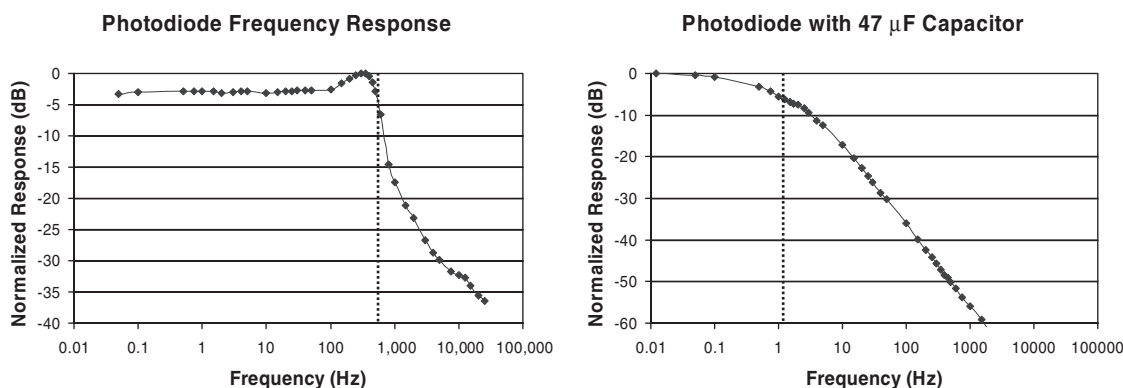
System built for this study (left) with a bandwidth of about 600 Hz that can be improved by filtering noise early on by limiting the bandwidth.
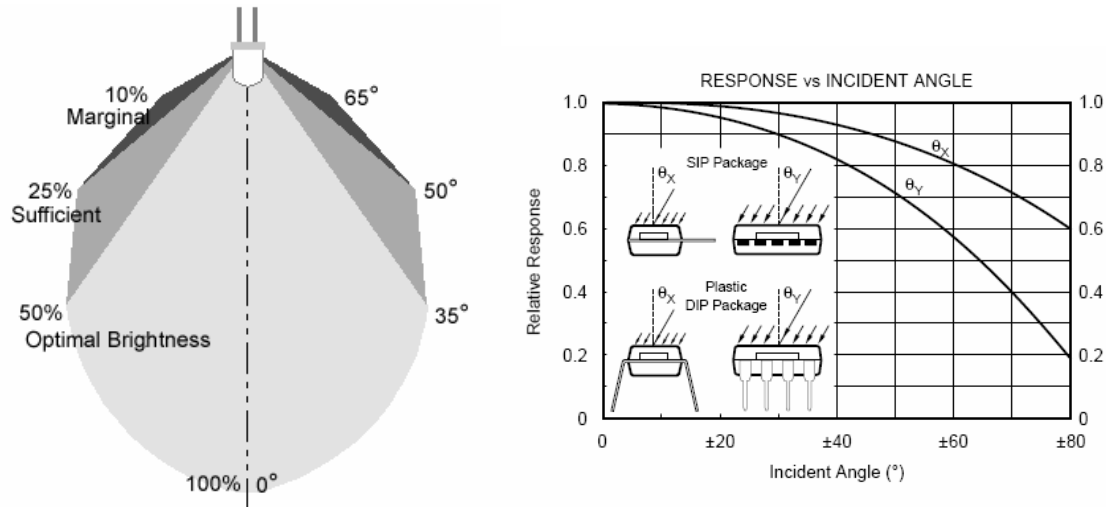
subject onto the photodiode [Fig. 13]. Additionally, if the LED is considered as a point source, the light intensity varies inversely with the square of the distance from the source to the subject, so changes in the distance between the LED and the subject would also affect the measured backscattered light intensity. In the presence of room light, even a passerby's shadow could dramatically change the readings due to the high gain necessary for accurate sampling. Change in a subject's skin condition, such as perspiration or body temperature, is yet another factor that may affect backscattered light measurements.

The microcontroller selected for this study would not have any trouble performing the calculations demanded by Equations 2 through 7 to monitor changes in blood volume or oxygen levels in the brain. Its I/O ports could digitally display real-time data on an LCD, light up various indicator LEDs as to the subject's health, read user inputs, and sound audible warning alarms. Moreover, instead of the analog potentiometer knobs used for this study, the microcontroller could control more precise digital potentiometers; the gain and DC offset could be adjusted automatically to keep the signal in the ADC's range and to take advantage of the ADC's full dynamic range. All changes to the potentiometers would be recorded and taken into consideration in the microcontroller's calculations. Nonetheless, a Digital Signal Processor may be desired for its generally greater computational power, ADC range, and quantization precision.

Conclusions

This study demonstrated the feasibility, challenges, and limitations of remote sensing with an LED, photodiode, and microcontroller. Together, these components offer the opportunity for a compact, durable, and effective handheld device if a consistent and stable signal can be extracted. Small-scale remote sensing tests, under highly controlled conditions, have demonstrated the ability to determine not only the pulse rate of a subject, but also changes in blood volume or oxygen levels in experiments designed to simulate brain function measurements, which could allow for remote monitoring of NICU infants' brain health or remote measuring of brain activity. Further study that implements several of the recommendations made here could make remote sensing of brain function a reality. Even so, the exciting possibility of remote sensing with a handheld device requires much additional work before it can be an effective technology.

Adam Wang

Acknowledgements

References

[1] FamilyFun. *FamilyFun: Health Dictionary: Prematurity.* <http://familyfun.go.com/parenting/child/health/childhealth/dony89enc_prem/> (24 July 2005).

[2] Nemours Foundation. *A Primer on Preemies.* July 2004. <http://kidshealth.org/parent/growth/growing/preemies.html>(24 July 2005).

[3] B. Chance, S. Nioka, and Y. Chen, Shining new light on brain function, *SPIE's OEMagazine,* 3 (2003) 16-19.

[4] Z. Zhao, X.C. Wang, and B. Chance, Remote sensing of prefrontal cortex function with diffusive light, *SPIE,* 5616 (2004) 103-111.

[5] Y. Lin, G. Lech, S. Nioka, X. Intes, and B. Chance, Noninvasive, low-noise, fast imaging of blood volume and deoxygenation changes in muscles using light-emitting diode continuous-wave imager, *Rev. Sci. Instrum.,* 73 (2002) 3065-3074.

[6] A. Zourabian, A. Siegel, B. Chance, N. Ramanujan, M. Rode, and D. Boas, Trans-abdominal monitoring of fetal arterial blood oxygenation using pulse oximetry, *J. Biomed. Opt.,* 5 (2000) 391-405.

[7] D. A. Boas, D. H. Brooks, E. L. Miller, C. A. DiMarzio, M. Kilmer, R. J. Gaudette, and Q. Zhang, Imaging the body with diffuse optical tomography, *IEEE Sig. Proc. Mag.,* 18 (2001) 57-74.

[8] Prahl, Scott. *Tabulated Molar Extinction Coefficient for Hemoglobin in Water.* 4 March 1998. <http://omlc.ogi.edu/spectra/hemoglobin/summary.html> (24 July 2005).

[9] Wikipedia. *Photomultiplier.* 6 July 2005. <http://en.wikipedia.org/wiki/Photomultiplier_Tube > (24 July 2005).

[10] Molitor, Andrew. *Wratten Filters for Infrared- & UV-Photography.* <http://www.a1.nl/phomepag/markerink/irfilter.htm#89B> (24 July 2005).

[11] Grandwell Industries Inc. *LED Sign Viewing Angle and Brightness.* 4 July 2005. <http://www.grandwell.com/vw_angle.htm> (24 July 2005).

[12] Burr-Brown Products from Texas Instruments. *OPT101: Monolithic Photodiode and Single-Supply Transimpedance Amplifier.* 23 July 2003. <http://www-s.ti.com/sc/ds/opt101.pdf> (24 July 2005).

On Writing Illness: How Virginia Woolf Finds a New Language for Illness in The Voyage Out, Mrs.Dalloway, and The Waves

Adam Critchfield

Abstract

This article is an examination of three of Virginia Woolf's characters through the lens of her essay, *On Being Ill* (1930), in which she ponders the "undiscovered countries" that are disclosed when we fall ill. Woolf writes that the transition from a state of health to one of illness can simultaneously bring to view "wastes and deserts of the soul'" and "precipices and lawns sprinkled with bright flowers." Given the power of such experiences and their commonality, Woolf claims that, traditionally, illness has rarely treated as the subject of literature. She calls for a 'new language' of illness to be found. I explore how she developed this new language of illness for herself by looking at three characters in three of her novels: Rachel from *The Voyage Out* (1915), Septimus from *Mrs.*

Dalloway (1925), and Rhoda from *The Waves* (1931). Woolf's essay work and autobiographical writing on illness, I argue, provide a key to understanding the coherence of imagery, experimentation with form, and the roles played by these characters in her novels. Each character I examine can be viewed as a "seer" who demands that the reader question the implications of illness and the social aspects of its construction. Illness, even as it may cloister us physically, opens up new realms of the psychological and allows us to understand the world in new ways. Woolf's work reflects a profound understanding of this surprisingly functional aspect of illness.

Whenever we fall ill—whatever the illness—we fall out of our habitual routines, unable to play our usual part in the world. It is not uncommon to feel useless or helpless as these negative aspects of illness accumulate and sometimes debilitate. While we are not putting one step properly in front of the other—while we are not rushing from item to item on a to-do list or fretting about the clock—we at least have the opportunity to be recumbent and to reflect. Maybe our reflections are constrained to the unusual aches and pains of the body, but even so we have been given a reprieve from our routine experience. Suddenly our own left thigh or right hand seems larger and more alien than a mountain range, foreign and outside the realm of our control. Illness forces us to listen to our body. We suddenly remember that our mind is chained, sometimes even ancillary, to the body; we feel less like our own masters and more like we are being driven by forces that are beyond us. Virginia Woolf argues that this inability to ignore the body and its sensations during illness is not so unlike the transports of passion or the wrenching away of personal control that accompanies falling in love. Given the abundance of literature devoted to love, why is there not an equally abundant supply of illness literature? This is exactly the question that Woolf poses in her 1930 essay *On Being Ill*.

By 1930, Woolf had already begun exploring and would continue to explore this question of illness in her life and literary career. As someone who suffered frequent bouts of physical and mental illness (the nature of which has been famously and frequently mulled-over

by scholars), she recognizes the capacity of illness to simultaneously bring to view "wastes and deserts of the soul" and "precipices and lawns sprinkled with bright flowers" (*On Being Ill* 3). She also understands illness as an experience common to everyone, functioning through tropes of the body as a simultaneous reminder of both our individuality and our common nature. "Considering how common illness is," begins Woolf, "how tremendous the spiritual change that it brings, how astonishing, when the lights of health go down, the undiscovered countries that are then disclosed. . . . what ancient and obdurate oaks are uprooted in us by the act of sickness, how we go down into the pit of death and feel the waters of annihilation close above our heads," she continues with a characteristically lyric flourish, "it becomes strange indeed that illness has not taken its place with love and battle and jealousy among the prime themes of literature" (*On Being Ill* 3-4). I am interested in showing how Woolf *has* found a new language for illness in her novels; how she has attempted to correct this lack of its address in the literary canon. It was this idea of illness leading to "undiscovered countries," to new lands of the physical and psychological, that allowed me to see a fascinating coherence of imagery in Woolf's novels. In particular I examine how Woolf writes illness through an imagistic web of nature and madness that brings together the characters of Rachel in *The Voyage Out* (1915), Septimus in *Mrs Dalloway* (1925) and Rhoda in *The Waves* (1931).

It is crucial to recognize that in using *On Being Ill* as a lens through which to retrospectively view these works and their relation to illness, we can put Woolf's own words and experience back into focus rather than reductively reading her work as evidence of a diagnosable illness. Much scholarship in the past has focused on biographical accounts of her illness, in turn understanding her work either psychoanalytically as "evidence of her preoedipal attachment to her mother, her unconscious guilt regarding her mother's death, and her father's overwhelming patriarchal presence" (Coates 244) or, by contrast, understanding her writing through the biological lens alone as direct manifestation of manic-depressive disorder (and therefore reading her words as

symptomology). As Kimberly Coates acknowledges in her own scholarship on illness and Woolf's aesthetic vision: "not one of these works addresses, much less acknowledges, 'On Being Ill.' If we attend to Woolf's own discussion of illness in 'On Being Ill,' we pay heed to what she valued most highly—her art" (Coates 245). Furthermore, examining her words as symptomology assumes a kind of direct, unconscious, or unfiltered transcription of her illness into her work, and doing so is to run the risk of slighting her as an artist. Mining her personal experience of illness was no easy task, as it necessitated a brutal confrontation with her periods of 'madness.'[1] In the confrontation of that personal experience she is able to productively recover the sensations and experiences of illness and refashion what would have otherwise been lost, or devalued as evidence of incapacity or insanity, into startling works of art. Iain Bamforth offers a potential critique of *On Being Ill*, posited in his introduction of the essay as it is included in his anthology *The Body in the Library*, that "pain, especially severe and protracted pain, actually destroys our ability to express it and, a fortiori, the imagination's attempt to shift it into the realm of fiction or myth." But this is exactly why the intersection of writing and illness is so important: as we shall see later, these periods of 'debilitation' were envisioned by Woolf as a chrysalis of sorts that, through writing, *could* be recovered and the experience of the body transformed into literature.

Woolf first addressed illness in her 1915 debut as a novelist, *The Voyage Out*. It is a coming-of-age story whose heroine, Rachel, moves further and further from the constraints and expectations of a patriarchal father whose late nineteenth-century mores and views of women would condemn her to a cloistered existence. She is a naïve but searching girl with artistic drive and great potential. The sea-vessel *Euphrosyne* takes her to South America, into a lush realm where the primitive is allowed to persist unmitigated by the English will. Gardens are wild and untrimmed (the gardener of the villa where Rachel stays has been too busy clipping flowers for his lover to properly tend the garden), and many of the vacationers seek to re-create English society abroad (the imposition of their social time over natural time,

the corralling of the wild into tame gardens) rather than appreciate what the land has to offer. Rachel, however, far from seeing this new landscape as "indecent" as Mrs. Chailey finds it, is eager to explore it and learn about life. The reader finds this verdant atmosphere expressive of Rachel's blossoming as a young woman, and a perfect backdrop for her first love. *The Voyage Out* then makes a very peculiar turn. The narrative begins to show a quirkiness that readers of Woolf's later fiction will recognize. Her language takes an angular, fractured turn, and we see a hint of Woolf's future modernist explorations. We can beneficially compare a scene from the beginning of *Voyage Out*, in the style of the traditional sitting-room drama, to a later passage that is angular and abstract:

> "Oh, Rachel, how d'you do," she said, shaking hands.
> "How are you, dear," said Mr. Ambrose, inclining his forehead to be kissed. His niece instinctively liked his thin angular body, and the big head with its sweeping features, and the acute, innocent eyes. (15)

> They began therefore to describe how this felt and that felt, how like it was and yet how different; for they were very different. Voices crying behind them never reached through the waters in which they were now sunk. . . . The grasses and breezes sounding and murmuring all round them, they never noticed that the swishing of the grasses grew louder and louder, and did not cease with the lapse of the breeze. A hand dropped abrupt as iron on Rachel's shoulder. . . . She fell beneath it, and the grass whipped across her eyes and filled her mouth and ears. Rolled this way and that, now seeing only forests of green, and now the high blue heaven, she was speechless and almost without sense. (283)

The first passage features clear dialogue, expository language, and a familiar narrator reporting on the traits and appreciations of the characters. The second is more exemplary of Woolf's abstract, surreal experimentation. It is completely unclear what can be taken literally from this passage. There is a surprising fracture in the plot as well, as the expected culmination of this coming-of-age story in Rachel's impending marriage is tragically interrupted. She contracts an exotic fever and dies.

Adam Critchfield

Returning briefly to Woolf's own complex experience with illness will help us understand *The Voyage Out*. By the time she wrote her first novel, the author had suffered illnesses ranging from mental breakdown to influenza that had left her bed-ridden, sometimes for weeks, with fevers, headaches, and hallucinations; often these were completely debilitating. In writing, though, Woolf manages to complicate the understanding of illness beyond its traditional conception as entirely destructive or detrimental. By reading her diaries we see that in these moments of sickness lies the germination of some of her best, most original and modern work:

> If I could stay in bed another fortnight I believe I should see the whole of the Waves. I believe these illnesses are in my case—how shall I express it? Partly mystical. Something happens in my mind. . . ; it shuts itself up. It becomes chrysalis. I lie quite torpid, often with acute physical pain—then suddenly something springs. (Bell 287, 16 February 1930)

It is not surprising that in *Orlando* (1928) the title character makes the dramatic transition from male to female after just such a chrysalis-like experience that resembles yet resists the label 'illness.'

Woolf is clearly interested in the relationship between illness and mystic experience. Her mystic view of the incomprehensibility of life leads her to say that it is this quality that we instinctively seek when we are sick:

> In illness words seem to possess a mystic quality. We grasp what is beyond their surface meaning, gather instinctively this, that and the other—a sound, a colour, here a stress, there a pause—which the poet, knowing words to be meager in comparison with ideas, has strewn about his page to evoke, when collected, a state of mind which neither words can express nor the reason explain. Incomprehensibility has an enormous power over us in illness. . . . In health meaning has encroached upon sound. Our intelligence dominates over our senses. But in illness, with the police off duty, we creep behind some obscure poems by Mallarmé or Donne, some phrase in Latin or Greek, and the words give out their scent and distil their flavour, and then, if at last we grasp the meaning, it is all the richer for having come to us sensually first. . . . (*On Being Ill* 21-22)

It is the unchaining of the mind as rational chariot-driver that is necessary for such mystical, transcendental experiences. Illness is one way in; Woolf's own hallucination of birds singing in Greek found itself written into *Mrs Dalloway* as one of Septimus' experiences of "disproportion."[2]

Returning to Rachel's abrupt end: how does the relationship between illness and incomprehensibility help us understand the final third of *The Voyage Out*? As previously stated, the language begins to feel more modern, more atmospheric, and more evocative rather than expository. A key passage, and a perfect example of this incomprehensibility, is that impressionistic excerpt already referred to in which Rachel is traveling with Terence in the jungle (and when she most likely contracts her fatal illness). It is especially important to pay attention to the images of "voices crying. . . through the waters in which they were now sunk" and "the grasses and breezes sounding and murmuring all round them" (*Voyage Out* 283). The invasion of nature into Rachel's body (the grass that "filled her mouth and ears," her vision being "only forests of green, and now the high blue heaven, she was speechless and almost without sense") occurs at approximately the same point in the narrative when she is falling sick and falling in love. Woolf explicitly makes this connection between love and sickness in *On Being Ill*:

> The public would say that a novel devoted to influenza lacked plot; they would complain that there was no love in it—wrongly however, for illness often takes on the disguise of love, and plays the same odd tricks. (6)

Woolf is giving Nature the supreme power in *The Voyage Out*, which ties love and illness together and makes Rachel an exploration of both. There is a reason why we use the expression "to *fall* in love" or say that we are "*madly* in love" with someone; both entail a sense of helplessness and irrationality. The decorous, calculating Englishmen and Englishwomen in the novel (who stay in a proper British-style hotel as contrasted with the Ambroses' unobtrusive villa) have a difficult time accepting such losses of control. They would rather label

their surroundings as obscene and place blame for Rachel's death on her exploration of them: "She probably ran risk a dozen times a day that might have given her the illness. It's absurd to say she caught it with us," says Wilfrid to Mrs. Thornbury (*The Voyage Out* 359).

We can understand Rachel's illness as her being taken under by the forces of nature with which she is so intimately connected. Rachel's strongest connection to an element of nature is with water: an image that will bind her to Woolf's other "seer" characters, Rhoda and Septimus. Rhoda is something of a reprise of Rachel, especially in her fluidity as a character; she is someone who can never be pinned down. The sea, for both of them, stands for boundlessness and a lack of limits as opposed to the arbitrary labels which well-to-dos like Mr. and Mrs. Dalloway are so keen on using. The ship is another important figure that binds both Rachel and Rhoda. It is the solitary vessel that takes Rachel out of England into new realms of possibility. The ship is also the symbol onto which Rhoda seizes:

> Alone, I rock my basins; I am mistress of my fleet of ships. But here, twisting the tassels of this brocaded curtain in my hostess's window, I am broken into separate pieces; I am no longer one. . . The wave breaks. I am the foam that sweeps and fills the uttermost rims of the rocks with whiteness; I am also a girl, here in this room. (*The Waves* 106-107)

This passage expresses the tension between the desire to be one with nature and the realities of being an individual that seems to be Rhoda's primary source of anguish. Woolf recognizes this same tension as one of the hallmarks of being ill: an acute awareness of one's individual, particular nature.

The fact of being alone with our illnesses is one of the reasons why her characters Rhoda, Septimus, and Rachel all feel like they are on a precipice of sorts, exposed to the elements. In *The Waves*, Rhoda feels just such an exposure when she interacts with people; as she understands herself, she "has no face." Rhoda's desire to escape herself is one of her predominant features, allowing us to understand her comments about not desiring to eat and the shock of meeting people. In her

monologue during the meal towards the end of the book, she expresses anguish at the obligation of eating:

> But since these rolls of bread and wine bottles are needed by me, and your faces with their hollows and prominences are beautiful, and the table-cloth and its yellow stains, far from being allowed to spread in wide and wide circles of understanding that may at last (so I dream, falling off the edge of the earth at night when my bed floats suspended) embrace the entire world, I must go through the antics of the individual. (*The Waves* 224)

Her reluctant acceptance of embodiment ("these rolls of bread and wine bottles are needed by me") is underscored by her desire for escape, for disembodiment. Rhoda feels chained to her body. On the way to meet her five childhood friends as an adult, she wants to hide herself in the trees, but for once she walks "straight up to you instead of circling round to avoid the shock of sensation as I used" (222). It is Rhoda who most acutely feels the tension between being in the moment, having to absorb the shock of sensation, and the desire to "spread wide and wide circles of understanding." Rhoda is constantly in danger of "falling off the earth" for this very reason. She wants to be alone in the darkness, as if by escaping her individual nature she can slide into the universal; she resolves this tension at the end of the book by committing suicide. Although her suffering (and it is clear that Rhoda suffers) and its result are no less tragic, there is an understanding on the part of the reader that Rhoda has finally "returned" to her primordial roots. And she has hardly left the narrative. As the novel draws to a close, the majority of the text's narrative has become Bernard's monologue, and Rhoda appears to slip quietly and unobtrusively out of the narrative when Bernard recounts Rhoda's suicide to the reader in a simple, straightforward manner. Two pages later, however, it is Rhoda's voice that seeps into Bernard's monologue when Bernard sees himself in a mirror:

> Oh, but there is your face. I catch your eye. I, who had been thinking myself so vast, a temple, a church, a whole universe, unconfined and capable of being everywhere on the verge of things and here too, am now nothing but what you see. . . .

Adam Critchfield

Lord, how unutterably disgusting life is! What dirty tricks it plays us, one moment free; the next, this. . . . Always there is the enemy; eyes meeting ours; fingers twitching ours; the effort waiting. (*The Waves* 292)

This hardly sounds like the Bernard who thrives on the presence of others, as the reader has come to know him. Instead, the eyes that meet ours are "the enemy." The preoccupations sound much more like Rhoda's, as we see perhaps even more strongly on the next page:

Strange that the face of a person, whom I scarcely know save that I think we met once on the gangway of a ship bound for Africa—the mere adumbration of eyes, cheeks, nostrils—should have power to inflict this insult. Yet this shadow which was sat by me for an hour or two, this mask from which peep two eyes, has power to drive me back, to pinion me down among all those other faces, to shut me in a hot room; to send me dashing like a moth from candle to candle. (*The Waves* 293)

The moth flitting about in the dark is a strong evocation of Rhoda's spirit. Rhoda, in the dark, can lose the boundaries and difficulties of individual nature; she can exist before the beginning of the story even—before time, before distinctions, before "the sun rises" when "the sea [is] indistinguishable from the sky" (*The Waves* 7).

Even though the male characters in *The Waves* each aspire to write poetry, Rhoda *is* poetry, embodying its qualities of searching, a desire for unity and transcendence of the mundane, and its function as a vehicle for mystic incomprehensibility. Early in the book, as she reads in the library, Rhoda's voice becomes that of the poem "The Question" by Percy Bysshe Shelley; Woolf purposely eliminates any punctuation that would distinguish between the boundaries of Rhoda and the poem:

I will wander down [the hedge] and pick flowers, green cowbind and the moonlight coloured may, wild roses and ivy serpentine. . . . I will sit by the river's trembling edge and look at the water-lilies, broad and bright, which lit the oak that overhung the hedge with moonlight beams of their own watery light. (*The Waves* 56-57)

The imagery here, especially of the water's edge and the moonlight, is evocative of Rhoda's fluid, nocturnal nature.

There is a strong connection between Woolf and this unconventional "heroine" who provides the writer with a way of transforming her own experience of illness into literature. She says that one of the tests of a book's successfulness and usefulness to a writer is "if it makes a space in which, quite naturally, you can say what you want to say." Indeed, the morning that Woolf wrote that sentence in her diary she also wrote: "this morning I could say what Rhoda said. This proves that the book itself is alive: because it has not crushed the thing I wanted to say, but allowed me to slip it in, without any compression or alteration" (Bell 297-298, 17 March 1930).

Before writing *The Waves*, Woolf had already made another such 'space' for speaking through her creation of Septimus Warren Smith. Here again is a character with a mystical connection with nature, evoked through familiar images: "he remained high on his rock, like a drowned sailor on a rock. I leant over the edge of the boat and fell down, he thought. I went under the sea. I have been dead, and yet am now alive, but let me rest still, he begged" (*Mrs Dalloway* 75). The experience of this shell-shocked war veteran runs counter to that of Clarissa Dalloway during the day in June that has been immortalized in *Mrs Dalloway*. Septimus' seemingly unconnected saga of madness and suicide runs parallel to Clarissa's preparation for her party. The reader senses a strange complementarity of these two threads, one dark and the other light, and the inevitability of a convergence. Underneath the mundanities of every day and the preparations for the party, there is Septimus' dramatic inner conflict. Just as everybody, in time, goes through "the whole and unending process of changes, heat and cold, comfort and discomfort, hunger and satisfaction, health and illness, until there comes the inevitable catastrophe" (*On Being Ill* 5), every single moment of life possesses these same tensions as soon as we "enlarge the bubble," as Rhoda might say[3]. Clarissa and Septimus show us the coexistence, in the same moment, of health and illness, the present and the past, the party and the war.

Given the horrors of war, it is not surprising that Septimus is left without rational "proportion," the term that his doctors in *Mrs Dalloway* are so fond of using. Dr. Holmes conflates sanity, or proper rational proportion, with his own narrow views of truth, morality, and British society. The restoration of Holmes's notion of proportion is an attempt to restore Septimus to baseline normality, but just as a tranquilizer might stop suffering, it also numbs the patient, dulls him or her, and disconnects him from the source of the suffering. The source of the suffering doesn't disappear; merely, the communication is severed. The patient becomes unable to voice the causes of suffering, which are so often social and not solely psychological or physical. In restoring Septimus' proportion, Drs. Holmes and Bradshaw would eliminate his source of creativity, his mystic connection with Nature, and any unusual or progressive ideas he might have. It is the shell-shocked Septimus who is constantly scribbling poetry in a mad effort to be understood and to communicate what he understands as a result of his so-called disproportion. He hears birds chirping in Greek, opening him up to the intense beauty of the moment around him:

> To watch a leaf quivering in the rush of air was an exquisite joy. Up in the sky swallows swooping, swerving, flinging themselves in and out, round and round, yet always with perfect control as if elastics held them... and now and again some chime (it might be a motor horn) tinkling divinely on the grass stalks—all of this, calm and reasonable as it was, made out of ordinary things as it was, was the truth now; beauty, that was the truth now. Beauty was everywhere. (*Mrs Dalloway* 76)

Septimus sees a unity that underlies the mundane, as do Rhoda and Rachel. Just hearing the word "time" in a commonplace expression uttered by his wife ("'It is time,' said Rezia"[4]) elicits a profound meditation:

> The word 'time' split its husk; poured its riches over him; and from his lips fell like shells, like shavings from a plane, without his making them, hard white, imperishable, words, and flew to attach themselves to their places in an ode to Time; an immortal ode to Time. He sang, Evans answered from behind the tree. The dead were in Thessaly,

Evans sang, among the orchids. There they waited till the War was over, and now the dead, now Evans himself –. (*Mrs Dalloway* 76)

Septimus, while no longer able to function in a "normal," socially productive way, is clearly in touch with something deeper. This character makes the reader strongly question what it means to be sane or well. Given the nature of his experience of war he understandably challenges British imperialistic norms. The proportion that his doctors seek to return him to simply cannot exist after such harrowing experiences.

Woolf sympathizes with Septimus' frustration with his doctors; she herself knows how important and productive the moments spent at the extremities of our sanity can be. Her own experience of "madness" had important results, as she describes in her diary:

> I can't spin a sentence, & sit mumbling & turning & nothing flits by my brain which is as a blank window. So I shut my studio door, & go to bed, stuffing my ears with rubber; & there I lie a day or two. And what leagues I travel in the time! Such 'sensations' spread over my spine & head directly I give them the chance; such an exaggerated tiredness; such anguishes & despairs; & heavenly relief & rest; & then misery again. Never was anyone so tossed up & down by the body as I am I think. (Bell 174, 11 February 1928)

As soon as Woolf relinquishes control of the directing mind and "gives [the sensations of the body] the chance," she is "tossed up and down" by the experience, much like one of Rhoda's ships at sea. She has to accept what nature presents to her: just as she lets her body experience physical sensations, she lets her mind register impressions that fall like "an incessant shower of innumerable atoms" (*Modern Fiction* 287) onto her brain. How else could she have written her experience of illness without the poetic, fragmented, modernist play of form and stream-of-consciousness that she uses so beautifully? Rather than trying to subdue the mind's capriciousness in these moments and make an explicitly rational order of it, she lets the large, mysterious, and often confusing and ambiguous world speak for itself; she writes her complex sense of this world into her work.

Adam Critchfield

Woolf, then, is something of a medium for the mind's registration of sensation: "the mind is the most capricious of insects—flitting fluttering" (Bell 175).

The doctors that Woolf dealt with personally as family members fell ill and as she herself struggled with mental illness find themselves written into her fictional characters like Holmes and Bradshaw in *Mrs. Dalloway*. Woolf's critique of the doctors is one way of tackling the social implications of illness. Not only does the doctor have power over her as a patient seeking his help, but doctors also occupy a social position imbued with the power to dictate the lines between sanity and insanity, often with a dangerous conflation of madness with immorality. In Stephen Trombley's research into Woolf's experience with her doctors, he elaborates on 'moral insanity,' a concept developed by Woolf's own doctor George Henry Savage:

> We. . . encounter eminent doctors writing of concepts such as 'moral insanity' and 'aesthetic insanity,' concepts which cannot be rigorously defined, and which often find a spurious justification in social, political, class or sexual bias. (Trombley 2)

It becomes easy for political and sexual biases to take root and propagate themselves through the often-unquestioned social and moral eminence of doctors. We see this even in *The Voyage Out* with Rachel's doctor, Dr. Rodriguez. He is a moneyed, upper-class doctor whose power derives from his position within the dominant patriarchal order. This particular character is respected despite his bumbling incapability in treating Rachel, his shortcomings instead being easily brushed off with placating gestures towards Terence, who gives him the benefit of the doubt because of the social status of medicine and the blind trust that follows. In *The Voyage Out* and *Mrs. Dalloway*, then, Woolf situates the doctor within his socio-historical context and examines medicine as an institution that sometimes reinforces and promotes hegemonic understandings of the world. Stephen Trombley gives us just such a critique of Woolf's doctor, George Savage:

> The diagnosis of insanity as developed by Savage is a tool for enforcing personal and political beliefs, and social and moral expectations in an arbitrary and subjective fashion. (Trombley 123)

> Despite his arguments for biology and heredity, however, Savage also had psychological opinions of mental illness, though they are hardly more than the products of personal bias and culturally prescribed Victorian stereotype. He believed, for instance, that spoiled children were likely to develop unsound minds and that too much education was mentally harmful for the lower classes and for intelligent young women rebelling against their natural roles as wives and mothers. (Trombley 14)

Such views threaten to silence messages from the margins of society and the margins of the mind. Indeed one of Savage's treatments for Woolf was strict bed-rest for prolonged periods, involving no reading or writing[5]. This was torture for Woolf; such a stifling of her creativity, a would-be silencing of her critical and progressive voice, is analogous to the tranquilizers offered to Septimus, a man who would dare to speak out against war and empire.

Ultimately, it is our lot as human beings to be tossed about by the unforgiving and often indifferent realities of human experience; this is perhaps felt no more strongly than during our moments of illness and 'rational disproportion.' Our body undergoes an epic interior journey that is just as important, and as heroic, as the journeys of literature's traditional war heroes. Septimus can be seen as Woolf's new answer to the war hero, who is hardly the coward that Dr. Holmes would have us believe he is. It is Septimus who listens to the interior journey, to Time and to Nature; the body is our link, and we must listen.

> We must let the body speak, and Nature speak through it. Yet it is not only a new language that we need, more primitive, more sensual, more obscene, but a new hierarchy of the passions; love must be deposed in favor of a temperature of 104; jealousy give place to the pangs of sciatica; sleeplessness play the part of villain; and the hero become a white liquid with a sweet taste—that mighty Prince with the moths' eyes and the feathered feet. . . . (*On Being Ill* 7-8)

Virginia Woolf forges her new language of illness in Rachel's fever, Rhoda's eyes ("like pale flowers to which moths come in the evening"[6]), and Septimus' divine madness. And perhaps we can now understand that when Rhoda drops a twig into her basin as "a raft for a drowning sailor"[7], she is doing so for every Septimus Warren Smith who, even as his 'illness' has left him exposed on a lone precipice, will never truly be lost at sea.

Works Cited

Iain Bamforth, ed. *The Body in the Library: A Literary Anthology of Modern Medicine.* New York: Verso, 2003.

Bell, Anne Olivier, ed. *The Diary of Virginia Woolf. Volume Three 1925-1930.* New York: Harcourt Brace Jovanovich, 1980.

Coates, Kimberly Engdahl. "Exposing the 'Nerves of Language': Virginia Woolf, Charles Mauron, and the Affinity Between Aesthetics and Illness." *Literature and Medicine* 21, no. 2 (Fall 2002) 242-263.

Lee, Hermione. *The Novels of Virginia Woolf.* London: Methuen & Co Ltd, 1977.

Trombley, Stephen. *'All That Summer She was Mad': Virginia Woolf and Her Doctors.* London: Junction Books Ltd, 1981.

Woolf, Virginia. "Modern Fiction." 1919. *The Virginia Woolf Reader.* New York: Harcourt Brace Jovanovich, 1984.

———. *Mrs Dalloway.* 1925. London: Penguin Books, 2000.

———. *On Being Ill.* 1930. Ashfield, Massachusetts Paris Press, 2002.

———. *The Voyage Out.* 1915. New York: Harcourt, Inc. 1920.

———. *The Waves.* 1931. New York: Harcourt, Inc. 1931.

Endnotes

1. Lee 95.
2. *Mrs. Dalloway* 26
3. "Yet there are moments when the walls of the mind grow thin; when nothing is unabsorbed, and I could fancy that we might blow so vast a bubble that the sun might set and rise in it and we might take the blue of midday and the black of midnight and be cast off and escape from here and now." (*The Waves* 224)
4. *Mrs Dalloway* 76
5. Trombley 110
6. *The Waves* 16
7. *The Waves* 18

Adam Critchfield

Kiss and Tell:
Racial and Ethnic Variations in Adolescent Dating and Sexual Practices

Emily A. Fisher—*Population Research Center
and School of Biological Sciences*

Abstract

*O*bjective: In order to gain a better understanding of racial/ethnic discrepancies in fertility and marriage, indicators of adolescent sexual behaviors were analyzed. The relationship between dating frequency, sexual experience, and age at the onset of dating and sex for non-Hispanic whites, non-Hispanic blacks, U.S. born and foreign born Mexicans was analyzed. Gender differences were analyzed for each racial/ethnic group. *Methods:* Using the National Longitudinal Survey of Youth 1997, data were compiled on adolescent dating and sexual behaviors and the proportions of each population were compared in terms of dating and sexual experience. Logistic regression analysis was used to determine the statistical significance of differences that existed between the

populations. *Results:* An analysis of dating frequency found that non-Hispanic whites are much more likely to date than non-Hispanic blacks ages 13-16. U.S. born Mexican adolescents are less likely to report dating frequently compared to non-Hispanic whites, although they date occasionally at the same proportion as non-Hispanic white adolescents. Female Mexican adolescents born in the United States are significantly less likely to date than their male counterparts. The analysis of sexual experience found that non-Hispanic blacks are most likely to have had sex by ages 13-16. U.S. born Mexican adolescents are more likely to have had sex than non-Hispanic whites, but not as likely as non-Hispanic blacks by ages 13-16. U.S. born Mexican males are significantly more likely to have reported having sex than their female counterparts. The relationship between age at first sex and age at first date suggests that non-Hispanic blacks are the most likely to have had sex before they started dating. U.S. born Mexican adolescents and non-Hispanic whites have the same proportions that have sex before dating, to have started dating and had first sex at the same age, and to have sex after the onset of dating. Possible implications are discussed.

Adolescence is typically thought of as a temporary hell plagued by raging hormones, parental distrust, and a thirst for independence. Perhaps the most conspicuous of the "teenage issues" is the initiation of sexual behavior. Dating and sex are a preoccupation of not only teenagers, but society as a whole. Controversy surrounding birth control, abortion, and teenage pregnancy in current politics make the dating and sexual practices of teenagers especially relevant.

Much research has been done on the sexual behaviors of white and black teenagers and the disparity between those two groups. However, the second fastest growing minority group in the U.S. today is Hispanic, of which 58.5% is of Mexican origin (Census 2000). Little has been published about the dating and sexual practices of Mexican-American adolescents in the United States. This paper compares the dating and sexual practices of racial/ethnic groups in the U.S. with an emphasis on how the Mexican-American population compares to the other groups. In particular, questions regarding dating frequency and age at sexual initiation will be analyzed by race/ethnicity and gender. This paper contributes to the growing body of knowledge on adolescent romantic behaviors through two unique ways. First, the data set used in this analysis, the National Longitudinal Survey of Youth 1997 (NLSY97), is one of the first with a large cohort of Mexican adolescents. Second, dating behaviors as well as sexual behaviors are analyzed in order to better understand adolescent romantic behavior. In this paper, I address the following questions: How do the dating frequencies of non-Hispanic whites, non-Hispanic blacks, and Mexican-Americans compare across gender? How do Mexican-Americans, non-Hispanic blacks and non-Hispanic whites compare in terms of sexual experience across gender? Is the relationship between sexual intercourse and frequency of dating similar between Mexican-Americans, non-Hispanic blacks, and non-Hispanic whites? How does this relationship appear when comparing across gender for each racial/ethnic group? These questions are important in order to better understand one of the largest minority groups in the U.S.. The answers to these questions can aid in anticipating future fertility patterns and population structure, as well as have policy implications for sex education and structural needs for the Mexican-American population.

The NLSY97 survey used in this study provides new data on adolescents from the late 1990's, in contrast to earlier studies that used data from the late 1970's (NLSY1979). The questions posed will be answered using cross tabulations and logistic regression analysis to compare the proportions of each population that behave romantically and sexually similar.

Literature Review

There is extensive research on variations in marital behavior and fertility across racial/ethnic groups. Much of this research focuses on the underlying causes behind these differences which aides in understanding these relationships. After briefly reviewing some principal theories explaining racial/ethnic discrepancies, relevant research on dating and sexual behavior among adolescents is discussed.

Adam Critchfield

Among many studies on marital and fertility patterns of Mexican-Americans, Oropesa, Lichter, and Anderson (1994) found that Mexican-American women did not differ significantly from non-Hispanic white women in nuptial practices. However, they did note that this was misleading as it masked the generational differences of Mexican-American women. One of their main findings suggest that foreign born Mexican women faced the most obstacles and were much more likely than non-Hispanic white women or U.S. born Mexican-American women to have marriage chances affected by their local "marriage market" of eligible men.

Many other studies have focused on what has been called the Mexican-American nuptiality paradox. Although Mexican-Americans are more socio-economically disadvantaged than non-Hispanic whites, they are at least as likely to marry as non-Hispanic whites and are more likely to marry at younger ages. Socio-economically disadvantaged groups are usually less likely to marry than non-Hispanic whites, thus creating a paradox of the Mexican marriage norm (Oropesa, et al 1994). It is not uncommon for researchers to cite familism as the theoretical explanation for the high level of Mexican nuptiality. Familism is an assimilation theory that argues that cultural values will persist throughout multiple generations in the receiving country because individual immigrants have psychological and emotional ties to the parent culture. It is argued that Mexicans have a oriented culture that values marriage, family, and 'traditionally female' roles of chastity or motherhood (Wildsmith 2001). However, Raley, Durden, and Wildsmith (2004) argue that the Mexican marriage differences (from the expected characteristics) are not driven solely by a family oriented culture. Their argument is derived from data that shows that Mexican immigrants tend to marry earlier than Mexican nationals. The argument is also supported by the finding that when controlling for employment, school enrollment, and family background, Mexican-Americans were less likely to marry in any six month period than non-Hispanic whites.

Studies that include Mexican-American adolescent girls tend to focus on the struggles with traditional "feminidad" behavior (Dietrich 1998). Dietrich supported this idea of Mexican culture in her ethnographic study centered on a group of Mexican adolescent girls in a "varrio" in Los Angeles. Dietrich documents the girls' struggles with sexuality and dating as well as the fear of being labeled a "'ho" or its antithesis, a "schoolgirl." Dietrich does propose that as a result of the girls' limited opportunities due to their families' socio-economic status and the constraint of the passive role of females in their community, the girls make decisions that sometimes result in gang activity, teenage pregnancy, or dropping out of school (Dietrich 1998). Due to this body of information that suggests Mexican females tend to marry earlier than would be expected and tend to subscribe to a value system oriented around female chastity (regardless of whether this results from the cultural norms or from socio-economic opportunities), Mexican female adolescents are expected to report less sexual and dating activities than Mexican males and females of other racial/ethnic groups.

Furstenberg, Morgan, Moore, and Peterson (1987) found in a study of black and white adolescents that black adolescents were more likely to have had sexual intercourse than white adolescents at all ages. They also found that although age at first date and exclusive relationships were indicative of earlier sexual intercourse, these indicators did not explain the racial discrepancy. The purpose of their study was to explain differences between whites and blacks in pre-marital sexual relations. They concluded that the normative behaviors of the black population were more liberal towards and accepting of adolescent and pre-adolescent sexual relations. The authors do note that their analysis was based on data that is difficult to interpret due to substantial cultural and personal variation. The data suggests that we should expect black adolescents to report more sexual experience and more dating experience than non-Hispanic whites.

While much research exists on the black/white differences in adolescent sexual behavior, only recently have researchers began looking at how Mexican adolescents behave sexually. Cooksey, Mott, and Neubauer (2002) studied early adolescent friendship patterns and their relationship to adolescent sexual initiation and dating

behaviors. The same indicators for dating frequency and for the relationship between age at sexual initiation and age at the onset of dating were used for this study. Cooksey, Mott, and Neubauer conclude that pre-adolescent friendships are strong indicators of adolescent relationships and sexual experiences. Unfortunately, the study only examined racial/ethnic differences in behavior between black adolescents and non-black adolescents. The current study was also able to look at the racial differences through the lens of gender, lending greater insight into how the various racial/ethnic groups behave.

A study conducted by Upchurch et al. found that black males had the youngest median age (15.0) at first sex while white and Hispanic males, as well as black and white females, had a median age at first sex of about 16.5 years. Hispanic and Asian females were found to have first sex rates half that of white females at all ages (Upchurch et al. 1998). It concluded that socioeconomic controls could explain the disparity of age at first sex between all female groups. The gender disparity between Hispanic males and females was attributed to cultural factors. The study was unable to fully account for the large difference for black males although it did reference the Furstenberg study. (Furstenberg concluded that in the black community, premarital sexual relations were more acceptable.) A related study found that Mexican-American female adolescents in Los Angeles were most similar to non-Hispanic white females in timing of first sexual intercourse, but that foreign born Mexican-American adolescent girls were older at first sexual intercourse (Aneshensel 1989). This body of data requires one of the aforementioned hypotheses to be adjusted. It is expected that black male adolescents report the most sexual and dating experience. This data also suggests that we should anticipate Mexicans to resemble non-Hispanic whites in sexual experience, although female Mexicans should report lower percentages of sexual experience than non-Hispanic white females.

In light of these previous studies, our hypotheses can be refined as follows: 1) Non-Hispanic blacks will have higher rates of sexual experience than non-Hispanic whites, especially non-Hispanic black males.

2) Non-Hispanic blacks will date more than non-Hispanic whites. However, the Cooksey, Mott, and Neubauer study indicates that at older ages, black adolescents date less frequently than non-black adolescents. It might also be expected that Mexican adolescents have dating frequencies most similar to non-Hispanic whites since the Oropesa, Lichter, and Anderson study indicate that Mexican-American marriage tendencies are not significantly different from non-Hispanic whites (1998). Also, the Aneshensel (1989) study suggests the hypothesis that U.S. born Mexican adolescents will have similar sexual experiences to non-Hispanic whites. At the same time, we also expect a gender difference, with female Mexican adolescents reporting less dating and sexual activity than their male counterparts.

Data

The data used in this paper come from the 1997 National Longitudinal Survey of Youth (NLSY97). The 1997-2001 Codebook for the NLSY97 describes the parameters of the study and how the data were collected. The NLSY97 is one of a set of surveys sponsored by the U.S. Department of Labor, Bureau of Labor Statistics. This survey was conducted in order to gather information on adolescents between the ages of 12 and 16. The data set includes five consecutive years (1997-2001) of data collection on this cohort. Respondents and respondent's parents participated. The primary respondent, the adolescent, was asked questions concerning participation in the labor market, education, and social interactions. The parents answered questions concerning the community in which the youth lives, the household income, and parent's socio-economic indicators. The first wave of the survey in 1997 includes responses from 8984 adolescents. During the initial round of surveying, interviewers went to randomly selected households and asked if any inhabitants of the house were between the ages 12 and 16. All adolescents between these ages were eligible for the survey. The respondents came from 147 different non-overlapping primary sampling units composed of 6,819 different households yielded the 8,984 primary respondents for the first wave of sampling. Because the data collection was based on

Adam Critchfield

households, there are 1,862 households that have more than one respondent (i.e. siblings) (NLSY97 Codebook). All non-Hispanic black, non-Hispanic white and self-described Mexican respondents between the ages of 13 and 16 were included in the current study. Due to the small number of Asian respondents, they were also omitted. Other Hispanics and mixed race respondents were also omitted. With these exclusions, the final sample size for the study was 2070, although only 1596 respondents provided data on dating.

Measures and Methods

Data from the baseline 1997 survey was used to construct the race/ethnicity variable. The responding parent's race/ethnicity was used to indicate the adolescent's race/ethnicity. The parent's U.S. citizenship status and, if applicable, year of immigration were used to indicate the adolescent's place of birth. If the parent was a U.S. citizen, the adolescent was considered born in the U.S.. If the parent was not a U.S. citizen, the adolescent was considered born in the U.S. if the immigration year was the same or before the year the student was born.

The dating frequency variable was created using baseline data from the 1997 survey. Each respondent was asked one of the following questions: "How many dates you had in past year with boyfriend?" or "How many dates you had in past year with girlfriend?" The adolescent could indicate one of six responses ranging from "Never" to "Once a week or more (>50 per year)." "Married" adolescents were excluded. An adolescent who indicated 0 to 3 dates that year was put into the "Never to Few" group; 4 to 25 times per year were put in the "Less than once a month to 2-3 times per month" group; anything greater than 50 were put into the "Once a week or More" group.

Data from the 1997 to 2001 waves of the survey were used to compile the dependent variables measuring timing of first sex and the relationship between dating and sex. Each year a respondent was asked at what age they had first had sex ("Have you ever had intercourse, that is, had sex, made love, or gone all the way with a person of the opposite sex?" How old were you the first time you had sex with a person of the opposite sex?) If the

respondent had answered the question in a previous year, the question was not asked again. "Sexually experienced" was defined as anyone who answered this question. The question was summed across all survey years so that each respondent had either an age value or no data available for this question. Adolescents who had ever responded to this question were considered experienced while all others were considered not sexually experienced. Similarly, the dependent variable that described the relationship between sex and dating was created by pooling together all respondents who reported having their first sexual intercourse between the ages of 13 and 16. From this pool, the respondents were classified into one of three categories based on the relationship between the age of their first date (which was retrospectively summed like the age at first sex variable) and the age at first sex.

Analysis was conducted using cross-tabulations of the dependent variable by the independent (usually race/ethnicity). Logistic regression analysis was used to determine the significance of any differences between the populations, compared to non-Hispanic white adolescents. All data were weighted; however for the regression analysis the weights were normalized to a mean of one to maintain statistical validity.

Results

The results of dating frequency by gender and race/ethnicity can be seen in Tables 1 and 2 and Figure 2. Table 1 shows that at younger ages all racial/ethnic groups are significantly less likely to be dating frequently. As expected, the percentage of U.S. born Mexican females reporting never or rarely dating (60.3%) was more than twice the percentage of non-Hispanic white females reporting the same thing (26.6%; $p < 0.05$; see Table 2). U.S. born Mexican males were almost three times as likely to report dating often than their female counterparts (25.1% for males vs. 8.8% for females; $p < 0.05$). With respect to racial/ethnic discrepancies, U.S. born Mexicans and non-Hispanic blacks differed significantly from non-Hispanic whites. Overall, both groups were less likely to report dating frequently than non-Hispanic whites. However, U.S. born Mexican males and non-Hispanic black females were not

Table 1 | *Percent Distribution of Dating Frequency for 13-16 Year Old Adolescents by Race/Ethnicity and Age.*
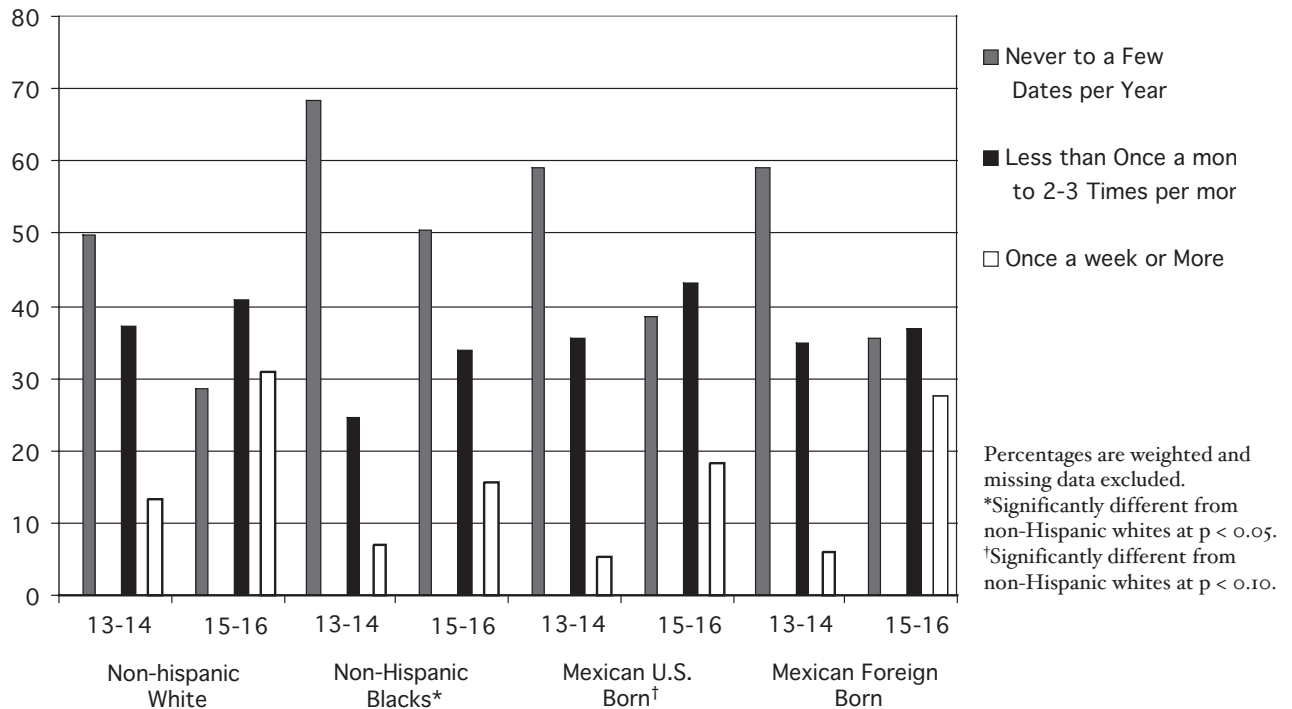
Percentages are weighted and missing data excluded.
*Significantly different from non-Hispanic whites at p < 0.05.
†Significantly different from non-Hispanic whites at p < 0.10.

Figure 1 | *Percent Distribution of Dating Frequency for 13-16 Year Old Adolescents by Race/Ethnicity and Gender.*

Percentages are weighted and missing data excluded.
*Significantly different from non-Hispanic whites at p < 0.05.
†Significantly different from males of the same racial/ethnic group at p < 0.05.

Adam Critchfield
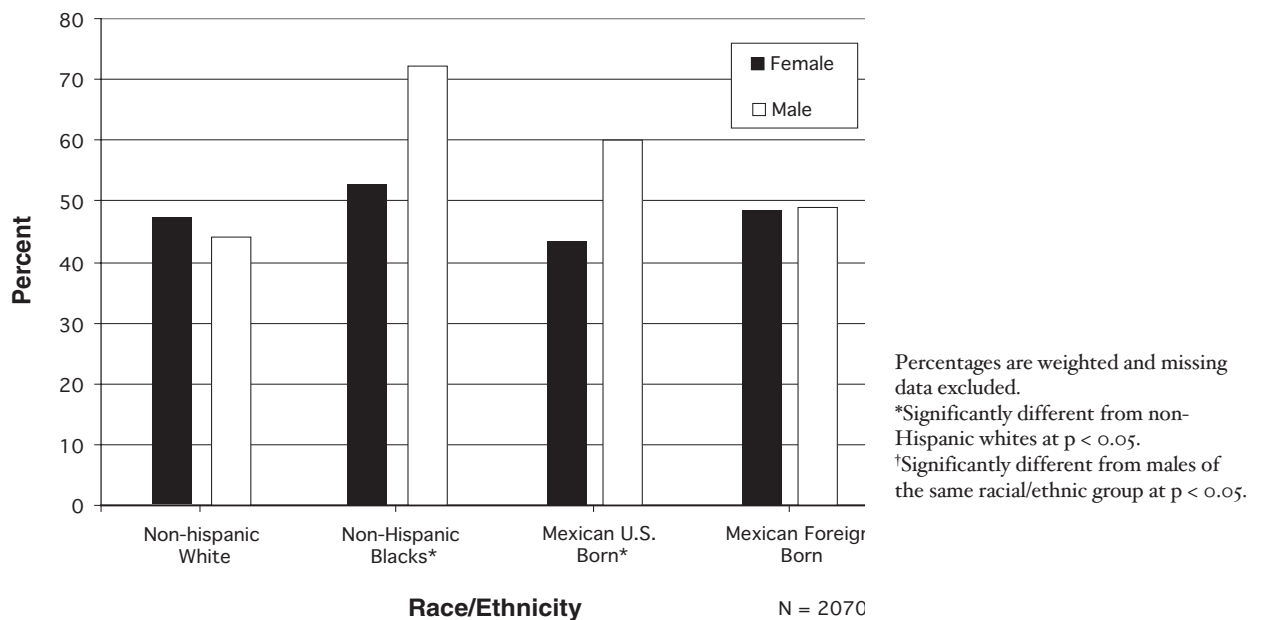
Table 2 | *Percent Distribution of Dating Frequency for 13-16 Year Old Adolescents by Race/Ethnicity and Gender.*

significantly different from non-Hispanic whites, but their peers of the opposite sex made up for the difference when analyzing the entire racial/ethnic group. As shown in Table 2, black males are significantly less likely to report dating frequently (once a week or more) (9.6%; p< 0.05) than their non-Hispanic whites peers (25.6%; p<0.05). Foreign born Mexicans tend to follow the pattern of U.S. born Mexicans in that females were less 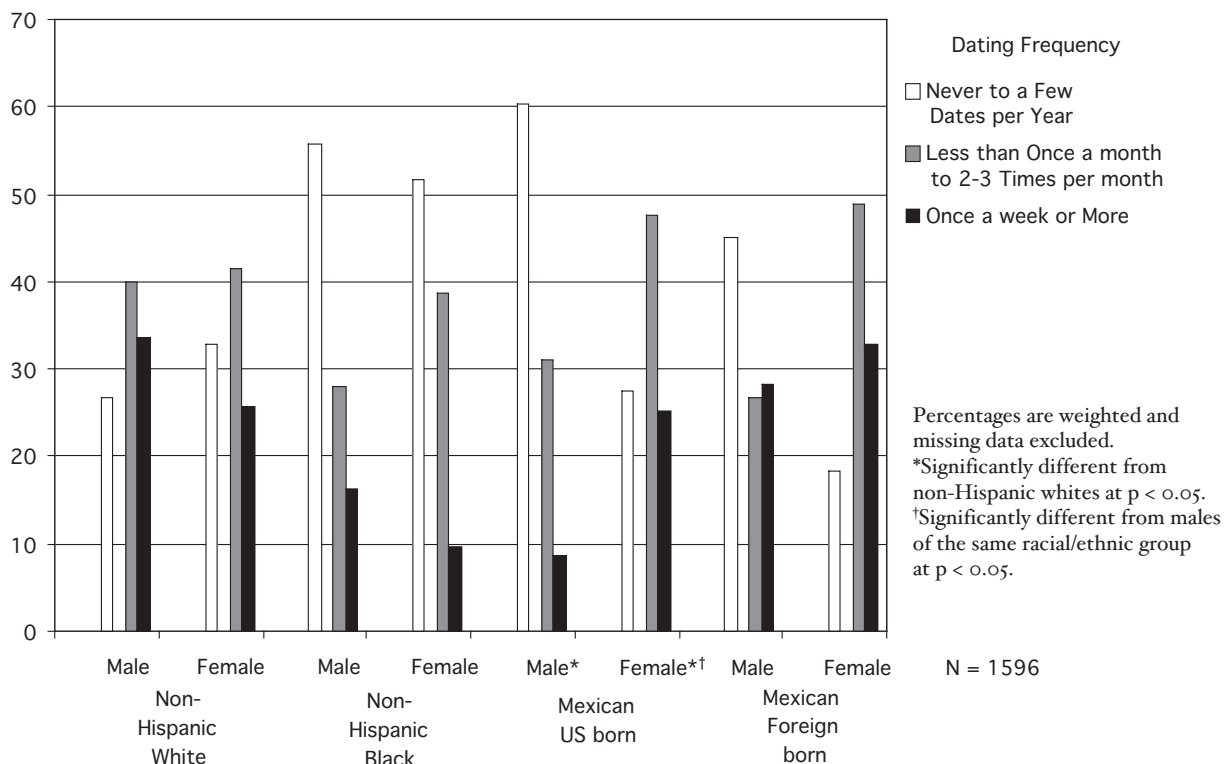likely to report dating (45.0% reported never dating or rarely dating) while foreign born Mexican males were more likely to do so (18.5%). These results were not statistically significant, most likely due to the relatively small sample size.

The results of the sexual experience analysis (Table 3 and Figure 1) showed that non-Hispanic black males are significantly more likely to report sexual experience (73%) than black females (53.1%; p < 0.05) and than any other racial/ethnic group. The U.S. born Mexicans follow a similar pattern where the males were more likely to report having had sex (55.3%)compared to non-Hispanic whites

Adam Critchfield

(45.2%; p < 0.05) while U.S. born Mexican females were less likely to report having had sex than non-Hispanic whites (47.7%; p < 0.05). Foreign born Mexicans followed a similar pattern; males (49.2%) were more likely than females (48.1%) to report having had sex. However, due to their small sample size, the information on foreign born Mexicans is only suggestive and not conclusive. The racial/ethnic differences show that both non-Hispanic blacks and U.S. born Mexicans have significantly different percentages of sexually experienced youth at ages 13-16. The gender difference in the non-Hispanic black and U.S. born Mexican groups is significant as shown in Figure 1. In both groups, females are significantly less likely to report having had sex than their male counterparts.

The results of the first two analyses are interesting to consider together. Although black males are the most likely to report sexual experience, they are the least likely to report dating with any frequency. U.S. born Mexican females are the least likely to report dating or sexual experience. Also, males and females in the same racial

Table 3 | *Percent Distribution of Dating Frequency for 13-16 Year Old Adolescents by Race/Ethnicity and Gender.*
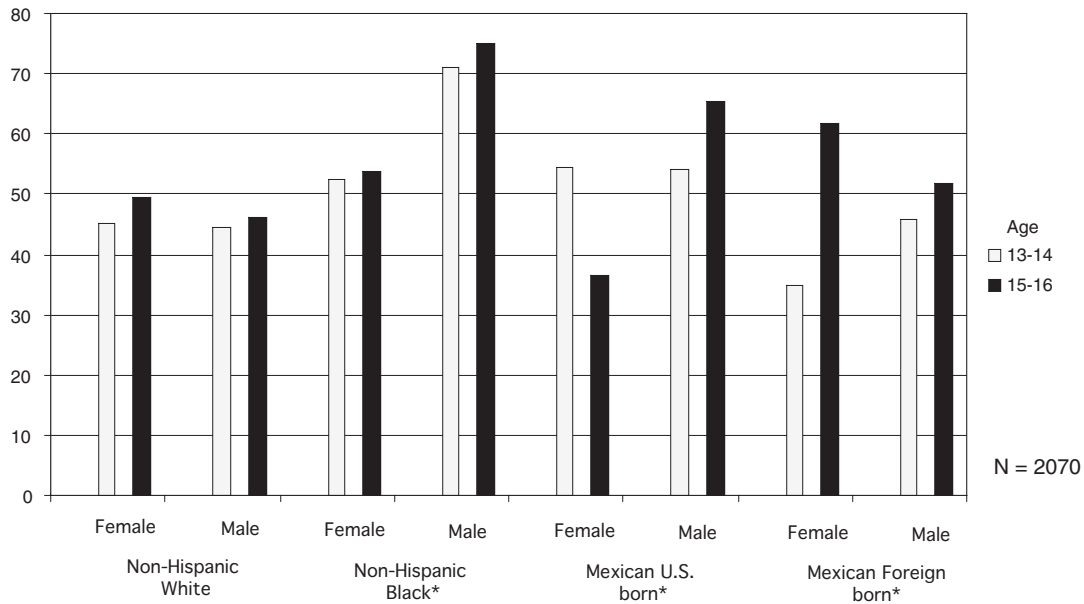
Percentages are weighted and missing data are excluded.
*Significantly different from non-Hispanic whites at p < 0.05.
†Significantly different from males in the same racial/ethnic group controlling for age and race/ethnicity at p < 0.05.

Table 4 | *Percent Distribution of Dating Frequency for 13-16 Year Old Adolescents by Race/Ethnicity and Gender.*

Percentages are weighted and missing data excluded.
*Significantly different from non-Hispanic whites at p < 0.05.
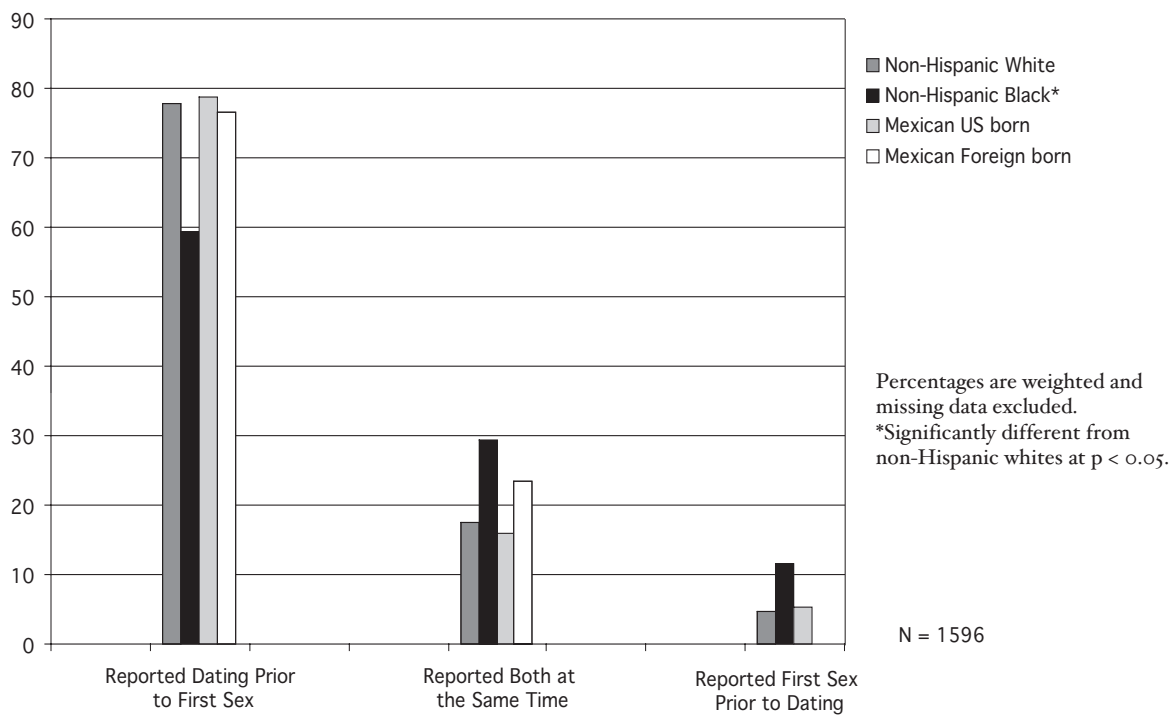
Adam Critchfield

Table 5

Relationship between Age at First Sexual Intercourse and Age at First Date for Adolescents Aged 13-16 who Have Had Sexual Intercourse by Race/Ethnicity and Gender

Legend:
- Reported Dating Prior to First Sex
- Reported Both at the Same Time
- Reported First Sex Prior to Dating

Percentages are weighted and missing data excluded.
*Significantly different from non-Hispanic whites at p < 0.05.

N = 1596
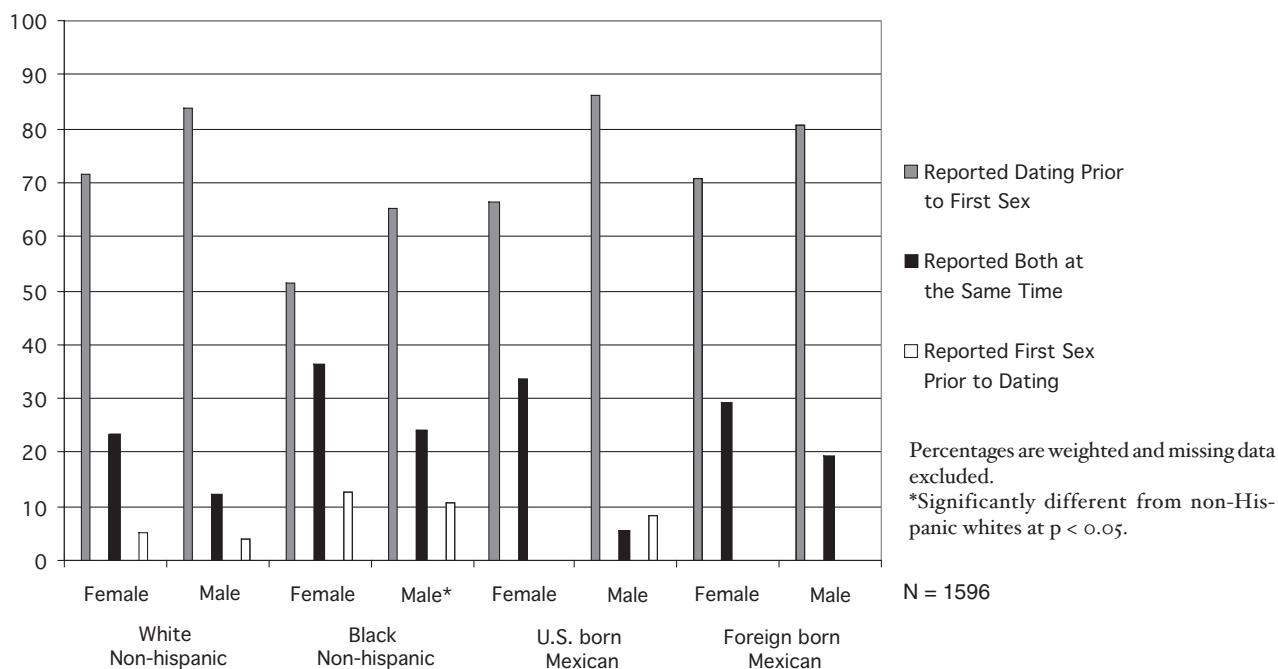
group tend to behave differently. Black males tend to report more sexual experience and less dating experience while their female counterparts report more dating experience and less sexual experience overall. U.S. born Mexican females report little dating or sexual activity while their male counterparts report the opposite. The final analysis looks at the reported relationship between age at first sex and age at the onset of dating.

Table 4 shows a significant difference between non-Hispanic blacks and non-Hispanic whites. Non-Hispanic blacks reported more than twice the proportion of first sex occurring prior to dating (11.5%) than both non-Hispanic whites (4.6%) and U.S. born Mexicans (5.2%; p < 0.05). Most of this difference is due to the non-Hispanic black males and not the females, as evidenced by Table 5. Perhaps the most surprising result from this analysis is with respect to gender. Females are less likely to report their first sex happening after dating than males. Table 5 indicates that females are more likely to report sex occurring at the same age as dating

rather than sex before dating. There are no significant differences between the genders in each racial/ethnic group, although the trends are at odds as can be expected from the results of the previous two analyses. U.S. born Mexican males are more likely to report sex after dating (86.2%) while their female counterparts are more likely to report both occurring at the same age (33.5%) than males. None of the foreign born Mexicans reported sex occurring before dating.

Discussion

This paper set out to better understand the relationship between sex and dating in adolescents and how that relationship appeared when analyzed by gender and race/ethnicity. Data were gathered retrospectively on teens' age at first date, age at first sex, and frequency of dating. The results of the dating frequency analysis found that non-Hispanic whites are much more likely to date than non-Hispanic blacks at ages 13-16. U.S. born Mexican adolescents are less likely to report dating frequently

Adam Critchfield

Figure 2 | *Percent Distribution of Dating Frequency for 13-16 Year Old Adolescents by Race/Ethnicity and Gender.*
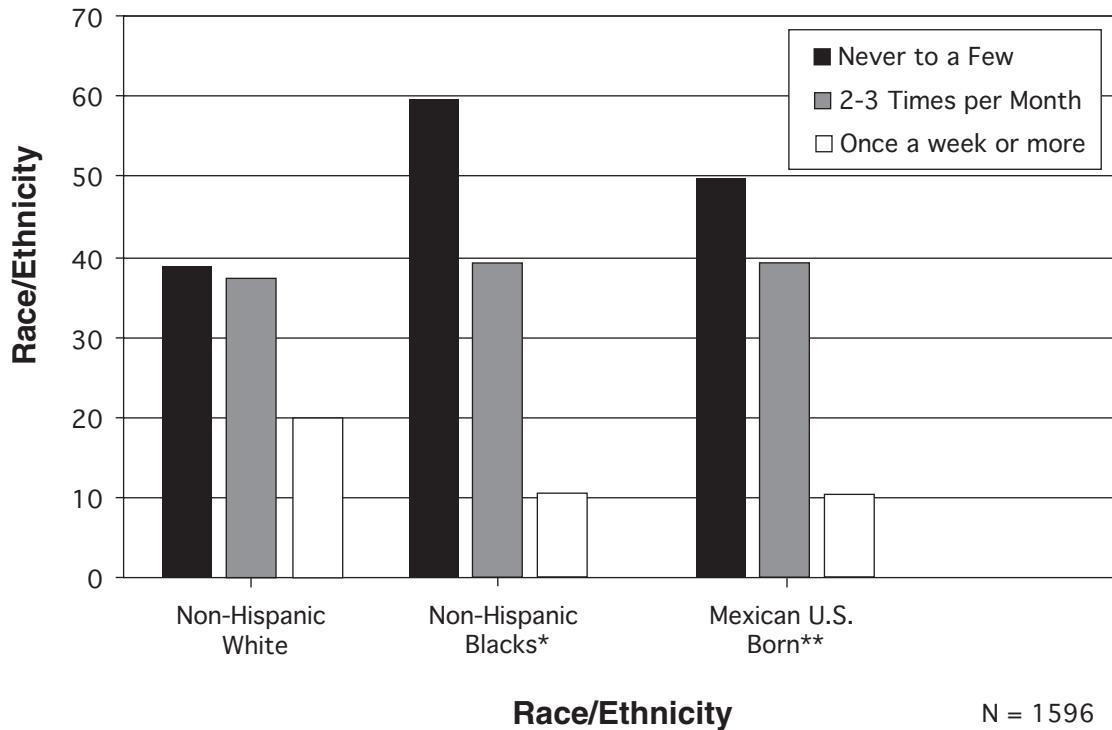
compared to non-Hispanic whites, although they date occasionally at the same frequency as non-Hispanic white adolescents. Female Mexican adolescents born in the United States are significantly less likely to date than their male counterparts. The results of the sexual experience analysis found that non-Hispanic blacks are most likely to have had sex by ages 13-16. U.S. born Mexican adolescents are more likely to have had sex than non-Hispanic whites, but not as likely as non-Hispanic blacks by ages 13-16. Black males are more likely to have had sex by ages 13-16 while black females are significantly less likely to have had sex. The proportion of U.S. born Mexican males reported having sex is more than three times that of their female counterparts. The results of the relationship between age at first sex and age at first date found that non-Hispanic blacks are the most likely to have had sex before they started dating. U.S. born Mexican adolescents and non-Hispanic Whites have approximately the same proportions that have sex before dating, to have started dating and have first sex at the

same age, and to have sex after the onset of dating. However, the only significant differences in this relationship are the following: 1) black males are significantly more likely to report having first sex before dating and 2) females are significantly less likely to report having sex after dating than males.

Conclusion

The results of this study correspond well with previous research. As expected, non-Hispanic black adolescents were more likely to report sexual activity at younger ages while non-Hispanic white adolescents and U.S. born Mexican adolescents reported more similar dating and sexual experiences. Mexican female adolescents were less likely to report dating or sexual activities than their male counterparts as ethnographic studies of Mexican-Americans suggested.

Limitations of this study include problems of retrospective self-reporting and the classification of dating and sexual activities. It is possible that due to

Adam Critchfield

social expectations or accepted norms, adolescents are not accurately reporting their sexual or dating experiences. The questions used to create the dating variable asked the adolescent about dates with a girlfriend or boyfriend which may have caused some respondents who do not date exclusively to not report their dating activities. Different perceptions of dating contribute to many problems in a research question such as this. The term "dating" seems to go in and out of use throughout generations. Perhaps some adolescents think of dating as something older people do and consider their outings as "hanging out," whereas a third party would classify it as dating.

Second, the stereotypical dating is only feasibly done by the middle to upper-middle classes. Stereotypical dating includes going out in a car, perhaps to the movies or dinner. However, many families do not have the luxury of allowing the adolescent to use the car for his or her dating purposes. Through this lens, dating is only an option for people with enough income to have a car which the adolescent has access to. This could cause an under-reporting of adolescent dating relationships, among those who do not have the financial means to "date" but are nonetheless in a romantic relationship.

Another interesting confounding factor is the meaning of the word dating in the different socio-economic or racial/ethnic groups. There were definite differences between the racial/ethnic groups in the reporting of dating. However, it is not clear whether dating means the same thing in each community. Perhaps future research could look into how different racial/ethnic groups or socio-economic classes perceive stages in romantic relationships. Also, a larger cohort of foreign born Mexican adolescents would be interesting to analyze with respect to these variables. More information on how foreign born Mexican adolescents behave could help understand the acculturation and assimilation processes that Mexican immigrants undergo when moving to the United States. This information could prove important to better understand one of the largest minority groups in the U.S..

Answers to questions such as the ones addressed in this paper can aid in anticipating future fertility patterns and population structure as well as have policy implications for sex education and structural needs for the Mexican-American population. Better sexual education, more available child care, neighborhood structure and city planning could provide disadvantaged youth with safe "hang out" places or dating opportunities.

Finally, how other factors affect the relationships discussed in this paper would be an excellent next step to this research. For example, do these early relationships persist in relation to the parent's socio-economic class, the adolescent's neighborhood, or the adolescent's educational status? Racial/ethnic discrepancies including teen pregnancy rates, average age at first marriage, or highest educational level completed are societal issues that could benefit from analysis found in future studies of this nature. Better understanding of how and why youth behave in romantic relationships can only improve sexual education and our attempts at reducing some of these racial/ethnic discrepancies.

Works Cited

1) Aneshensel, C.S., Fielder, E.P., and Becerra, R.M.. "Fertility and Fertility related behavior among Mexican-American and non-Hispanic white female adolescents." *Journal of Health and Social Behavior*. 30(1). 1989: 56-76.

2) United States Census Bureau. Census 2000. *Census 2000 Brief: The Hispanic Population. http://www.census.gov/prod/2001pubs/c2kbr01-3.pdf*

3) The Ohio State University. Center for Human Resource Research. NATIONAL LONGITUDINAL SURVEY OF YOUTH, 1997-2001 [Computer file]. ICPSR version. Columbus, OH: The Ohio State University, Center for Human Resource Research/Washington, DC: U.S. Dept. of Labor, Bureau of Labor Statistics [producers], 2003. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2004.

4) Cooksey, E.C., Mott, F.L., and Neubauer, S.A.. "Friendships and Early Relationships: Links to Sexual Initiation among American Adolescents Born to Young Mothers." *Perspectives on Sexual and Reproductive Health*. 34(3). May-June 2002: 118-126.

5) Dietrich, Lisa C. *Chicana Adolescents: Bitches, 'Ho's, and Schoolgirls.* Praeger: Westport, Connecticut, 1998.

6) Furstenberg, F.F., Morgan, S.P., Moore, K.A., and Peterson, J.L.. "Race Differences in the Timing of Adolescent Intercourse." *American Sociological Review.* 52(4). Aug 1987: 511-518.

7) Oropesa, R.S., Lichter, D.T., and Anderson, R.N.. "Marriage Markets and the Paradox of Mexican American Nuptiality." *Journal of Marriage and the Family.* Vol. 56. Nov 1994: 889-907.

8) Raley, R.K., Durden, T.E., and Wildsmith, E.. "Understanding Mexican-American Marriage Patterns Using a Life-Course Approach." *Social Science Quarterly.* 85(4). Dec 2004: 872-890.

9) Upchurch, D.M., Levy-Storms, L., Sucoff, C.A., and Aneshensel, C.S.. "Gender and Ethnic Differences in the Timing of First Sexual Intercourse." *Family Planning Perspectives.* 30(3). May-June 1998: 121-127.

10) Wildsmith, E. M.. "Female Headship: Testing Theories of Linear Assimilation, Segmented Assimilation, and Familism among Mexican Origin Women." Population Research Center, Working Paper Series. The University of Texas at Austin. 2001. *http://www.prc.utexas.edu/working_papers/past_series/series_00-01.html*

Adam Critchfield

The Struggle for Identity
Internal Tensions of Mexico's
Partido Acción Nacional

Carlo Daniel Guerra, *Department of History*

Abstract

Mexico's political system underwent a transition in 2000 after the *Partido Acci—n Nacional* (PAN) broke the *Partido Revolucionario Industrial*'s (PRI) 71 year rule over the country's government. The PAN's remarkable electoral success in the previous two decades came only after being a marginalized political party since its founding in 1939. Perceived as a conservative and closely-knit group, the PAN has struggled to maintain its unified front during much of its tenure.

In analyzing the PAN, I explore several questions. The first question is how internal conflicts actually assisted the PAN find a suitable platform to advance its ideology. Internal competition in the party focused on members who pledged to adhere only to a lofty

ideological platform, and those who wanted to moderate the party for better electoral results. The party's inner tensions culminated in its failure to nominate a presidential candidate in 1976, which forced the PRI to reform the election system in hopes of maintaining its pseudo-democratic control of the country. The PAN, despite its electoral success, continues to search for the most appropriate platform for its principles.

I also explore the importance of PAN's historical development in Mexican politics, and how it changed the role opposition parties played in the government. The PAN's success came only after its members accepted changes in key aspects of the party's creed, but maintained reverence for the original principles. Though the party continues to evolve, its main strength still lies in the solid ideological foundation it has promoted since its founding.

Introduction:
Historical Signifiance of *Acción Nacional*

On December 1, 2000, diplomats from all over of the world, ranging from U.S Secretary of State Madeline Albright to Cuban president Fidel Castro, convened in Mexico City to witness the presidential inauguration of Vicente Fox Quesada. The nominee for the *Partido Acción Nacional* (PAN) and former Coca Cola executive had achieved what was once considered an impossible feat—political victory in Mexico as an opposition candidate. Fox's victory ended the seventy one year rule of the *Partido Revolucionario Industrial* (PRI), a party often associated with corruption and electoral fraud.[1] The new president rose to the podium and acknowledged his four children first, instead of the customary nod to the legislative body. He outlined his "*Nuevo Mexico*" plan, which detailed twenty seven policy programs aimed at revitalizing the country's economy and eliminating government ineptness. Fox then personally administered an "Oath of Honor" to his cabinet and top advisors, a sign that re-affirmed his goal to forever change Mexican politics.[2]

President Fox experienced the typical political mud slinging characteristic to any high profile election. Surprisingly, not all of it came from the PRI's political

machine. The "doctrinal" *panistas,* tied to the conservative philosophy of the party's founders, argued Fox had compromised the staunch PAN ideals for the sake of victory. Fox, a product of the private sector and epitome of the "political entrepreneurs" who had recently joined the PAN, relied more on his *bravado* image rather than ideology to achieve a greater level of appeal. Often wearing jeans and cowboy boots and speaking with a populist flare, Fox was a far cry from the introspective philosopher often associated with the PAN. Several older *panistas* saw Fox's alliances with other political interests, including the Mexican Green Party, as naked political ambition at the expense of the PAN's venerable ideological platform.

This study argues that a variety of internal tensions sustained the PAN in a hostile political environment, eventually and eventually led to electoral success. Since its founding in 1939, the PAN's different sectors have tried to steer the party in what each thought was the best direction. As the country's most prominent conservative party, the PAN's initial groups—intellectuals, Catholic activists, and businessmen— were all connected by rightist views and their opposition to the PRI's dominance. But changing social conditions in the country and growing dissatisfaction towards the government started to attract more pragmatic voters and members to the PAN, which gradually led to more electoral gains. These transitions of internal power were significant and, as one will see, beneficial for the survival of the PAN.

Panistas claimed they represented the true ideals of the Revolution of 1910, which were corrupted by the radical factions of ensuing governments. Unlike the PRI, which juggled numerous factions and periodically shifted the application of its principles to serve whatever social conditions were prevalent at the time, the PAN prided itself on preserving a clear ideology that did not waver just to appease changing political conditions.[3] PAN leaders believed it was their duty to gradually educate the masses about democracy's complexities, and not resort to primitive political strategies just to attain power. Admittedly, the PRI's complete control had much to do with the PAN having no other choice than to stay true to its original ideology. However, the PAN pledged

Carlo Daniel Guerra

that its vision, once laid out to all of Mexico's citizens, would be the true cure for the country's social ills.

Antecedent Conditions, 1928-1939

To better understand the PAN, one must examine the political landscape in which it was founded. The PAN's creation resulted from a disenfranchised right wing sector subjugated by an emerging authoritarian regime. This new group eventually succeeded in becoming an opposition right-wing party by tempering its characteristically fiery conservative rhetoric. The struggle between conservatives and liberals dated back to the country's founding, and as with other Latin American countries, centered on the role of the Church. As an increasingly anticlerical regime emerged in the early twentieth century, an equally militant pro-Church uprising resulted in violent episodes between the two groups, followed by the systematic subjugation of conservative thought. The ruling government then crafted a political system where major social groups were collectivized and placed under its authority. This era in Mexico's political history is significant to the PAN because the party's founding fathers, concerned over the growing power vacuum and pseudo-socialist agenda, wanted to establish a new type of conservative party. These men maintained the anti-socialist rhetoric of their conservative predecessors, but explicitly stated that violence and conflict were not to be used. This new philosophy, brought about by the violent history in Mexico, would be the foundation for the PAN.

Emergence of the 'Revolutionary Family'

In his last State of the Union address in September 1928, Mexican president Plutarco Elias Calles called for the creation of an official government party. Calles wanted to keep his influence over the presidency, but the Constitution's "no re-election" clause forbade him from being re-elected. Calles selected Emilio Portes Gil as interim president, the first of several leaders Calles controlled. A convention held in March 1929 officially established the *Partido Nacional Revolucionario* (PNR). The "party of the government" placed rival politicians

from smaller parties and organizations under a single banner. Though the PNR experienced changes in both name and platform, its absolute control over elections and policy making became the main characteristic of the Mexican political process.

The PNR's domination of the election process solidified its position as the only possible avenue for political gain. The party's hierarchal structure left local and state politics dependent on the central authority for resources and rewards. PAN leaders later called for greater local autonomy in political affairs and identified *caciques* (local political machines) as the major obstacle to grassroots opposition. The party's internal rules absorbed any opposition groups, which further expanded the government's influence. The concept of "no re-election," originally placed as a stopgap to autocratic control, had been taken advantage of through careful political manipulation.

Gil's appeasement effects towards Catholic sympathizers prompted Calles to hold a special election in 1929 in hopes of finding a more malleable protégé. The PNR nominated Pascual Ortiz Rubio, a former Michoacán governor and Congressional deputy. José Vasconcelos, a former Obregón cabinet official, ran against Rubio in opposition to the PNR's emerging political monopoly. Conservative intellectuals, including PAN founder Manuel Gómez Morín, campaigned on behalf of Vasconcelos. But when Rubio defeated Vasconcelos with 93.6 percent of the vote, this victory not only demonstrated the rampant fraud in the country but also alienatied of the conservatives from the "revolutionary family." At the same time, the Vasconcelos episode demonstrated that that conservatives needed to create a viable permanent political party to challenge the existing government.[4]

Rubio's victory revealed the party's authoritarian nature, but did not give the victor the corresponding powers to rule the country. Rubio, like his ousted predecessor, tried to govern without the approval of Calles, but ran into political and congressional roadblocks. The "*Jefe Máximo*" continued pressing his agenda despite poor health and his absence from Mexico City. Competition grew between the two men, and Calles's influence

left Rubio as an outsider looking in. On September 2, 1932, Rubio turned in his resignation to Congress, which was accepted without hesitation or discussion. As per Calles's instruction, Congress named Abelardo Rodriguez as interim president to finish Rubio's term.

The Reign of Lázaro Cárdenas

With the 1934 presidential election looming ahead, the PNR sensed rifts within its ranks as the leftist factions clamored for more social reforms than what Calles had proposed. Calles selected Lázaro Cárdenas as the presidential candidate for the 1934 election, and spurned much of the party's rightist sector. In contrast to Calles's increasing conservatism, Cárdenas championed socialist causes and associated himself with the more radical faction of the party. Along with his ties to the military, Cárdenas promoted agrarian reform movement and advocated for a more worker friendly government. With a Cárdenas's victory inevitable and no sign of Calles loosening his grip on the government, observers on both the right and the left felt the "*Jefe Máximo's*" reign on the Mexican government would only continue.

Cárdenas, however, traveled throughout the countryside and made a name for himself as someone who would not bow down to the status quo. His constant campaigning and promises solidified enough popular support to oust Calles from the party in 1935. Cárdenas, now in full control, implemented his controversial Six Year Plan and surrounded himself with Marxist intellectuals and placed them in key economic and educational positions. One of his first orders of business amended the Constitution's Article 3 to say that the state supported a socialist curriculum in public education. Large tracts of land were broken up and given to *campesinos* (peasant farmers), who then collectively farmed the land. Cárdenas also organized and armed the peasants to defend their new acquisitions against potential attacks from large landholders. The *ejido,* as these large collective farms were known, became the basis of rural society and economic development.

Cárdenas further enhanced the party's dominance by developing a relationship between the government and select institutions composed of the country's largest groups, most notably the rural and working population. The government recognized these *confederaciones* as official associations for each respective social sector, in return for the organization's complete political allegiance to the party. The corporatist structure, as it became known, made the state the all powerful force in society and allowed it to better organize and control the country's main social groups.[5] The *Confederacion Nacional Campesino* (CNC) placed all the farmers and rural population under one organization, whose main goal promoted the interests of those workers in the countryside. The CNC gave Cárdenas support from a large, but politically weak, sector of the population. The structure of the *ejido* program also facilitated political control, as the farming population became increasingly dependent on the state for supplies and procection. Nonetheless, the CNC's creation allowed for a sector of the population to have viable political clout for the first time in the country's existence.

Cárdenas made it very clear that a strong labor organization would be a pillar of his administration, and carefully wove a web of support between rival unions and pushed Congress to approve the new organization's creation. The government formed the *Confederacion de Trabajadores Mexicanos* (CTM) which received significant power when Cárdenas declared that in any labor dispute where the employer was clearly not right, the government would side with the CTM. In March 1938, Cárdenas renamed the party the *Partido Revolucionario Mexicano* (PRM). Later in the 1940s, the party organized the popular and middle class sectors in much the same fashion of the CNC and CTM.

Controlling Mechanisms of the Ruling Party

Similar to Calles's PNR, politically ambitious individuals had to join the PRM in order to advance in politics. The party, and therefore the government, became so inclusive its leaders adopted the national colors to be the party's colors while forbidding other parties from using the *tri-color.* The prescribed path to power in Mexican politics was to follow certain "informal rules" within the party, stressing loyalty over policy initiatives. By rising through the ranks of organizations

Carlo Daniel Guerra

such as the CTM or CNC, future politicians learned the bureaucratic process and policy making. The "informal rules" stressed knowing one's place in the party structure and avoiding anything that would upset one's superiors. In return for his loyalty, a public official could advance to a higher legislative or executive post during each election cycle.[6]

The corporatist structure solidified the PRM as the dominant force in Mexican politics, and gave the government the option to switch its ideology at various times in response to the country's given social conditions. Cárdenas's socialist agenda, for instance, was tempered during the more private enterprise friendly regime of his successor, Manuel Avila Camacho. Adolfo López Mateos (1958-1964) and Luis Echeverría (1970-1976) pursued progressive policies more akin with those implemented by Cárdenas. What tied all the political leaders, on both sides of the ideological spectrum, was their commitment to "revolutionary ideals" of the party. These ideals, however, were open to interpretation by each administration.[7]

The irony of Mexico's political system was that though it operated like an authoritarian regime, it actually appeared to be a democracy. The institutionalization of an "official party" shifted all political resources and rewards to its supporters only. Opposition parties, such as the PAN, relied on small dedicated membership, usually with ties to ousted social groups like the Church. The government held elections only to appear as though a legitimate democracy actually existed in the country. Truth be told, opposition parties were bought off or simply repressed. Several upstart parties or organizations took advantage of the government's willingness to co-opt them, as it was the most effective way to express their grievances. Faced with such overwhelming power, PAN founders could do little more than focus on creating an internal structure that would not allow for the party's members to be easily seduced by the government's offers. The PAN saw itself as having a sound and ideology that distinguished it from the ruling government's political appeasement methods. The PAN still faced the challenge of accepting token representation at the expense of "playing ball" with the

Carlo Daniel Guerra

PRI, or of boycotting all elections until they had an actual chance at substantial victory.

The PAN: Ideology and Founders

On September 14th, 1939, approximately one thousand delegates met in Mexico City to form the *Partido Acción Nacional*. The new party expressed concern over the continuation of Cárdenas's policies following the election of a new president in 1940. The PAN's founders offered a political alternative to overcome the discrimination it felt had become statutory through the Constitution. The party constructed a coherent conservative ideology with a well-developed internal structure whose political methods were different than previous opposition parties. The construction of a rigid structure gave the party much needed guidance during its formative years and allowed it to create a more unified front. Although the party founders all shared anti-government leanings, disagreement arose over how the PAN should implement its platform and whether to participate in elections. Manuel Gómez Morín, a respected academic, argued that participation in the elections was only way to gain any success. Efraín González Luna, a prominent Catholic activist, countered that the PAN's focus should be on defending its ideology and that participation only legitimized the government's fraudulent practices. These differences signified the beginning of internal tensions that would eventually assist the PAN into becoming a more productive and dynamic party.

Manuel Gómez Morín

Born in Chihuahua of a Spanish father and Mexican mother, Morín received his law degree from UNAM in 1918, where he was part of the "Seven Sages of Greece," a group of gifted students that included several of his intellectual rivals.[8] His interest in policy and philosophy grew out of the group's long meetings and discussions, all which occurred against the backdrop of the Revolution. Under the Calles administration, Morín drafted several major economic laws, including the creation of the Bank of Mexico and the country's first income tax system. During the Cárdenas administration, Morín served as rector at UNAM, where he developed

close relationships with leading academics that later joined him in PAN efforts. As a private investor and corporate attorney, Morín became well respected within the inner circles of the country's banking and financing community. Morín relied heavily on the business sector in the PAN's formative stages, only to see it exit the party the following decade.

Morín used his academic position to criticize the government on its economic and political policies that favored certain groups at the expense of others. He believed that the expansion of the state into economic matters, such as the nationalization of the oil industry, and the incorporation of the masses into an official government party, restricted individual liberties. The government's control over the economy made it more vulnerable to external factors, such as price drops in the country's natural resources.[9] Morín also believed the *ejido* program made land distribution impossible and was counterproductive to the nation's agricultural economy. The country, Morín argued, did not have the resources for the *ejidos* to be implemented successfully. The legal battle between the state and the disenfranchised landowners created a social rift and further alienated an influential sector of Mexican society.

'Civic Education': The Ideology of the PAN

The PAN's business leaders opposed the revolutionary government's labor and market stances. No worker should be exploited for any political means, either by forcing him to join part-affiliated unions, such as those under the CTM, or by disciplining him for not wanting to take part of governmental affairs. The party also regarded private enterprise as a legitimate source of social development, as long as it conformed to the nation's interest and to the Common Good. The government had authority to regulate the national economy, but could not monopolize it nor exclude any class from participation. The idea of limited state involvement in economic matters pleased both the business sector of the party and their clerical colleagues; the business sector wanted to have little governmental interference in business matters, and the Catholic activists saw the role of government assisting, not directing, business matters.[10]

The PAN argued private property provided the best means for ensuring national production. Private ownership guaranteed both individual dignity and strengthened fundamental human communities, such as families and civic organizations. Large scale collectivization, such as the *ejido* program, inhibited civic participation because the masses could not question a government that controlled most means of production. The PAN supported a limited communal land program where peasants could own small parcels of land that would generate for greater overall production. The party also argued that private land ownership encouraged people to participate in the political process more, since they had something tangible at stake.[11]

Despite the heavy Catholic activist spirit in the party, the PAN initially took a non-religious stance, largely because of Morín's influence. Morín saw that the PAN's initial role was to encourage civic duty, and not become an ecclesiastical political movement. Mexican citizens had not been trained to be responsible citizens; instead they were conditioned to think that the Mexican political process was true democracy. An independent political party, not tied to the Church or the State, was essential if a pluralist society were to exist in the country. The party leaders, including several devout Catholics, decided not align closely with the Church for fear of anticlerical repercussions. Also, several members were uninterested in pursuing clerical ends. The party could not afford to lose any supporters from its small base. The repercussions of being identified as "a party of the Church" were too great, especially since the PAN's goal was to be seen as an alternative to the revolutionary government.

According to the PAN, the state should serve its people in matters that society could not accomplish on its own, and should promote solidarity between the government and its people. The country could not restrict religious or educational liberties, but did have the obligation to promote opportunities for the "cultural improvement of the Nation."[12] The government should provide materials and resources for public education, but it was the family's responsibility to choose their child's proper curriculum. Several members claimed that

Carlo Daniel Guerra

only religion could offer a true scientific concept of life and the world and should, thus, be an option to children. The state held a monopoly on education under Article 3, despite no actual law stating it had the right to do so. [13]

The party called for more local autonomy to subdue the primitive *caudillismo* that still existed in town halls and city councils. The PRI had created a highly centralized government, with the lower levels of government, such as the municipalities and state legislatures, having no control over their initiatives or budgets. Juan H. Alvarez, PAN presidential candidate in 1958, stressed the need to "recognize the city's capacity to dictate its own stature within the frame of the Constitution."[14] Alvarez criticized the PRI's *caciquismo* methods for obtaining votes, and the need to "carefully secure" the authenticity of city council elections.[15]

Catholicism: The Spiritual Essence of the PAN

Catholic activists, the most prominent group of PAN supporters, became so influential that the party is still characterized as a "Catholic party." The business sector's eventual exit from the PAN strengthened the Catholic element's position, where it further defined the party's ideology and internal structure. Catholics opposed the Constitution's anticlerical stances and Cárdenas's socialist leanings. Many *panistas* were students or young professionals who had sided with Morín during the Article 3 struggle. Numerous members served in other Catholic movements before joining the PAN. The Catholic activists were committed to keeping state and Church affairs separate, but to still support a strong social welfare state. Cárdenas's Six Year Plan revealed the government's intention of implementing various radical interpretations of the Constitution. Since Article 130 forbade church participation in political matters, Catholic activists believed a political party, within the formal rules of political participation, provided a more effective avenue for expressing their interests.

Efraín González Luna, who became the second most influential PAN thinker behind Morín, served as a leader in ACJM, learning valuable administrative lessons and making important political and academic contacts. If

Morín was the architect for the PAN, then arguably Efraín González Luna became the party's inner voice. González Luna, born to a devout Catholic family in the state of Jalisco in 1898, was educated in Catholic schools before receiving his law degree from UNAM. As president of the Jalisco chapter of the ACJM, González Luna was sympathetic to the *Cristero* rebellion, but disagreed on using violence as a tactic. González Luna became a prominent lawyer in Guadalajara and held several civic positions in the city. His primary interest remained, however, on the role of religion and his leadership in Mexican Catholic circles enabled him to recruit heavily for the PAN.

González Luna envisioned a Mexico built on small properties maintained by individual families. He viewed people in the countryside as the most loyal and closest to God, and as providing all the essential resources for the rest of nation.[16] It was the nation's responsibility to protect, not antagonize, its members.[17] Man could not sustain himself without the help of others and the state. González Luna often referred to the "Common Good," which was the end all people strived for, but could only be obtained through the help of a nurturing society. All institutions, such as government, religious institutions, and government were created to assist people in achieving the Common Good.[18]

According to González Luna, the PRI gave way to the "Mexican paradox," where an anti-Catholic minority controlled a country built upon mission work.[19] González Luna saw the Church as the most important social force in the country, despite its subjugation by the government.[20] The government contradicted itself when it relied upon the Constitution to promote freedom of belief, but then concluded that "the fight 'against fanatics and persecutors'" signified a fight against Catholicism."[21] The PAN, according to González Luna, satisfied the Catholic doctrinal requirements and fought for the all religious and social freedoms. He stated that:

> "[the PAN] places extraordinary value in its efforts. . . of a harmonious coexistence and strong collaboration for a country secularly torn by exclusive factionalism."[22]

Guided by these principles and by its acknowledgment that political prominence was not a short term goal, the party's central mission was insert a sense of morality into politics. The country's citizens could only obtain the Common Good by educating itself on civic matters and laying out a set of principles that would guide the political process. As González Luna stated, "we do not lack citizens. . . what we lack, gentlemen, is citizenry."[23]

In Search of a Platform, 1939-1961

The PAN's dilemma on how to properly deliver its message in Mexico's political environment overshadowed the party's unified ideology and structure. On the surface, the PAN appeared to be a united party whose members and leaders all believed the conservative goals it initially outlined. The party, however, experimented with different roles— including zealous conservatism, Catholic militancy, and progressive Christian Democracy—over the 1940s and 1950s. Even the PAN's steadfast leaders like Morín and González Luna had differing views on the PAN's purpose, especially on whether to participate in elections. Examining the motives of the PAN's different tactics across several elections lets us see internal tensions that at first hindered its development. Those tensions, however, helped the party adapt with the country's changing social conditions. It is important to analyze these early identity struggles because it was through them that the party was able to overcome its initial mistakes and eventually become more competitive.

Initial Electoral Experiences

The party faced over two choices for the upcoming 1940 election: participating in them and supporting a candidate for the presidential race, or abstaining and concentrating on member recruitment and civic education. The PRM's selection of Manuel Avila Camacho had already caused some PAN members to reconsider their allegiance in favor of the private enterprise friendly Camacho. Others claimed their initial activism influenced Cárdenas to appease the right and select Camacho, but it would have been detrimental to support the "official candidate" after solidifying an opposition party.

General Juan Andrew Almazán, a wealthy military officer from northern Mexico, emerged as the only viable candidate the young party could support. Almazán's military record appealed to conservatives and moderates and was favored among the rightists of the PRM. Almazán had deep ties to Mexico's industrial community and campaigned on the promise to reverse the radical initiatives of the Cárdenas administration.

Morín and his intellectual colleagues in the party advocated for participation in the elections. Morín stressed that participation was the only way the party could grow. He was, however, realistic in his outlook and knew that initial victory was impossible. His statement of "Let there be no illusions so that you don't become disillusioned" summed up his feelings on the tremendous uphill battle the party faced in the electoral arena.[24] Almazán left an unfavorable impression on Morín following an initial meeting. Morín opened the debate to the entire assembly, knowing that his reservations would divide the party. The rank-and-file members, some eager for instant political gains, voted overwhelmingly to endorse Almazán.

González Luna and other prominent Catholic *panistas* opposed Almazán as well, a clear sign that the founding members did not support the upstart general. They described Almazán as "primitive. . . like Santa Anna. . . he would have dragged the county through an electoral struggle based on personality and armed conflict."[25] Participation in fraudulent elections, when they already knew the outcome, only legitimized the PRM's dominance. The Catholics argued for abstention based on the party's "civic education" clause. They chose to promote party ideals and educate as many people as it could in hopes that someday the country would awake from its political slumber and oust the autocratic government. This stance also implied not compromising on its ideals for the sake of gaining outside support.

'Loyal Opposition'

Relations between the PAN and the Camacho administration were quite amicable, relative to past relationships between the PRI and conservative groups. PAN rhetoric regarding the government "still main-

Carlo Daniel Guerra

tained accusations of 'anti-democracy' and corruption, but were softened."[26] Several PAN members were offered or appointed government posts, most notably the placement of a *panista* on the Supreme Court bench.[27] Recognition of the PAN allowed it to become prominent in public affairs, even if that recognition was only tacit in nature. The relatively friendly relationship with the government backfired in one aspect when PAN's financial backers realized that Camacho was pro-business, which prompted many PAN supporters to switch their allegiance to the PRI.

The 1943 federal elections did not fare any better for the PAN. The party once again split between those for participation and those for abstention. González Luna reiterated his abstention stance but to no avail; the party decided to participate after an assembly vote with 61.2 percent in favor.[28] The 1943 elections marked the first time in almost two decades that an opposition party participated in legislative races.[29] Hopes of using its new position within politics and establishing a full ticket were dashed as the PAN recruited only twenty-one federal deputy candidates and officially won 21,000 votes.[30] With only 5% of the total vote cast, the PAN failed to win any seats in the Congress's Chamber of Deputies, and claimed victory over only one municipality.

The Beginnings of Political Success

The PAN's "National Unity" strategy for the 1946 elections finally gave the party its first legislative gains. The party supported the campaigns of several independent conservative candidates and formed coalitions with other groups. These additions to the PAN support base led to the election of four *panista* federal deputies, but only after the government conceded those races following cries of fraud. In a surprising move, even Morín ran for office in his home state of Chihuahua. The PAN's president constructed a high profile and effective campaign, only to be denied because his father was a Spanish national.[31] Despite the Morín episode, the PAN members who win worked on implementing the party's legislative program. The deputies proposed the creation of an independent federal election organization and for more autonomy for municipalities. The PAN also advocated agrarian reform,

including a 1948 legislative package to extend credit to rural population, grant private ownership, and build irrigation infrastructure.[32]

The 1946 election cycle also encouraged the PAN to nominate a presidential candidate to run against Miguel Alemán, the PRI's candidate.[33] Alemán, like Camacho, proved to be a moderate, but his association with the government forbade the PAN from openly supporting him. The party again had to deal with the ever-present abstention supporters, who rightfully believed electoral fraud would only continue. The PAN's convention met in February and stated they would offer an alternative to Alemán who could unite the country and bring prestige to the party. The party tried to nominate Luis Cabrera, an old *politico* who had first worked under Carranza.

Despite its meager success resulted in the party's continued to participate in subsequent elections. The party won its first state legislature seat, in Michoacán, the following year. In 1949, the party assembly unanimously voted to participate in elections. The party had reached national proportions with 26 regional committees spread across the country. The government recognized four of the 69 PAN candidates as victors, and declared that the PAN received 31 percent of the total votes in Mexico City.[34]

In 1949, Morín stepped down from the PAN presidency, satisfied with the party's role as the official opposition and its recent election victories in the Chamber of Deputies and various municipalities. In his last speech as party president he congratulated the PAN on ushering in a "new era" in its political life. To Morín, the PAN had successfully developed its doctrine and programs, allowing it to withstand negative attacks. The new era "created new possibilities and responsibilities which would require new capacities and new strategies."[35]

Morín's departure, along with the business sector's earlier exit, meant that the party now relied on its last bastion of support—the Catholic activists.

The PAN solidified its image as a Catholic party when it selected González Luna as its presidential candidate for the 1952 election. González Luna, long a proponent of abstention, now found himself as the primary

opposition candidate for the presidency. Like his PAN colleagues, he ran under the assumption of defeat, and stressed that he only served a higher political calling. On the campaign trail he stated that "there is no mystery that the regime will construct a new fraud…we have sowed the harvest of hope, which we all know someday, will produce a magnificent crop of victory."[36]

González Luna received support from several Catholic organizations and unions, and held several rallies in conjunction with Church leaders. As militant *Sinarquistas* further infiltrated the PAN as party candidates for Congressional and local seats, González Luna's platform based on "progressive Catholic thought" convinced many Mexicans that the PAN was becoming a political front for the Church.

Increased Tensions within the PAN

With the Catholic activists solidifying their position as the party's steering committee, the PAN focused on promoting its most confessional and conservative platform to date. During the 1958 campaign, the party's increased clericalism showed when PAN presidential nominee Luis Alvarez openly condemned Article 130 of the Constitution when he stated it "attacks orderly human activity and its natural forms of social expression."[37] The campaign was marred by frequent acts of violence between PAN and PRI supporters, and even led to Alvarez being thrown in jail on one occasion. PAN members claimed that authorities opened fire on them during a rally in Michoacán and one *panista* was assassinated in city of Juárez.[38]

The PAN's national assembly in July 1958 reignited the argument between those advocating participation and those supporting abstention. Fraud was considered so widespread that despite the PRI's offer of a recount, pro-participation leaders like Morín advocated for the PAN to withdraw from the entire process. He stated that the PAN should have nothing to do with a corrupt regime's actions. *Panistas* that supported the recount process, on the other hand, included the handful of former deputies elected in previous elections. The debate between the two camps became so heated that two assembly meetings were held, but led to no consensus.

The PAN's internal struggles continued after the government conceded four federal deputy seats to *panistas* in late August 1958. Naturally, those that favored abstention saw this as merely a political handout, and a way for the PRI to continue its pretence of democratic rule. PAN leaders ordered the four winners to not to accept their seat, but the desire to federal deputies was too great to pass up. Two of the winners, Jaime Haro and Felipe Gómez Mont, were expelled from the party for disobeying the PAN leadership.

New Approaches for *Acción Nacional,* 1962-1978

The PAN's continuing struggle to identify itself and to promote its ideology continued into the first part of the 1960s. The rise of reform-minded *panistas* marked a new transitive period in the party's history as new electoral doors were finally opened to them. This transition was facilitated by several factors, including social tensions in the country and the re-introduction of more secular members to the PAN. These eventual successes did not come without a struggle, as both the PRI and the older doctrinal PAN members sought to limit such new initiatives. The culmination of the PAN's internal struggles came when the party failed to nominate a presidential candidate in 1976, which ironically forced the PRI to reform election practices in order to have some degree of opposition in the country. As the pragmatists' numbers rose within the PAN, the challenge became one of blending the party's new external appearance interacted with its traditional rules and operations. The pragmatists' rise signified a new era for both the party and Mexican politics, in which political parties had a chance to be competitive.

Reforms under Adolfo Christlieb Ibarrola

Under the guidance of Adolfo Christlieb Ibarrola (1962-1968), the PAN evolved from a primarily doctrinal party to a more open and pragmatic organization. Ironically, Ibarrola was not considered a reformist before his ascension to the party leadership. He, like other PAN leaders, was a devout Catholic, but remained conscious of the separation of Church and State. Recruited into

Carlo Daniel Guerra

the party when he was a student, Ibarrola joined the Federal Electoral Commission as the token PAN representative, a position that established his relationship with future Mexican president Gustavo Díaz Ordaz.[39]

Ibarrola maintained the PAN's anti-government rhetoric of corruption and the need for more openness towards democracy. He expressed concern over the PRI's control of all forms of communication—radio, the press, even cinema.[40] Protecting small private property in the countryside, and the limiting the *ejido* program became Ibarrola's two main points on agrarian reform.[41] Despite his unusually close relationship with the PRI elite, he stated that "there cannot be any democracy in Mexico while there is an official party that, for it's own ends, utilizes...the decision making process and the internal control of the government, including. . . the electoral process. . . and the methods of diffusion and information."[42]

He was convinced that the PAN could not survive under PRI's electoral system, despite the moderate success it had obtained. Ibarrola was no longer satisfied to play the role of "loyal opposition." The PRI reformed the country's electoral laws in 1963 after Ibarrola threatened to pull the PAN from the upcoming elections. The PRI knew that without an effective opposition, it could not claim legitimacy in the country.

Continued Doctrinal Opposition

Despite the pragmatic approach taken by Ibarrola, several elements within the PAN still pushed the party's revered ideology and more importantly, its clerical platform. González Torres had stepped down from the party's leadership in 1962, but managed to secure the party's presidential bid for the 1964 elections. The doctrinaires, led by Efraín González Morfín, the son of González Luna, argued that the pragmatic wing threatened to undermine the party's integrity and to shift the party's ideology for the worse. The 1970 presidential nomination process turned into a heated debate, with Morfín declared the winner after a third ballot. Morfín's nomination made it clear that the PAN, though satisfied over Ibarrola's progressive gains, was still committed to its original ideology and goals.

When Morfín received the 1970 nomination, the party was in a state of crisis. Ironically, Morfín and other influential leaders initially voted against participation in the upcoming elections. Other PAN leaders voted for partial participation (i.e., no deputy and municipal candidates) and focused on using the presidential race as the best way to spread PAN ideas and to attack the government. The party once again had troubles finding candidate slots for the Congressional seats. Several state delegations, including *panista* bulwarks Chihuahua and Baja California, voted for total abstention. In the end, full participation prevailed by a six-vote majority. By the time Morfín hit the campaign road, his PRI opponent, Luis Echeverría, already had a two month head start.

The PRI soundly defeated the PAN's nominee, but did recognize that the PAN received 14 percent of the popular vote, its largest gain at that time. The PAN also received the maximum 20 allotted opposition seats in Congress. Morfín's doctrinal approach succeeded because it revitalized the conservative faction, but also satisfied the progressive *panistas*. Elections and campaigning tend to unite the different factions within the party and the 1970 PAN campaign was no exception. Nonetheless, there were still high tensions between the doctrinal and pragmatic wings of the party, both of whom believed that the party gained momentum under the respective leaderships of Morfín and Ibarrola. During the second part of the decade, the PAN faced its biggest internal rift as both sides jockeyed for control of the party.

The PAN's Internal Crisis

José Angel Conchello, elected as the party's president in 1972, significantly overhauled the PAN's electoral and ideological strategy during his three-year reign. As a vocal opponent of President Luis Echiverría's socialist-leaning initiatives, Conchello led an influx of entrepreneurs and businessmen dissatisfied with the PRI's economic and political rhetoric into the PAN. Conchello believed the PAN's overly doctrinal position hindered its development, and further pushed the pragmatist approaches first taken by Ibarrola. Conchello's "open doors" policy gained the

PAN notable success in the 1973 elections win it won four deputy seats in areas the PAN had never had any foothold in.[43] These new approaches included building alliances with other marginalized groups, placing elections as the main focus of the party, and to "simplify [the PAN's] truths, and express the doctrine in terms more understandable and easier to comprehend by the general public."[44]

The doctrinal camp, alarmed at Conchello's philosophy, successfully secured the election of Morfín as the party president in 1975. Morfín reversed Conchello's initiatives and stressed the party's commitments to its original ideology and structure. He called the PAN's tenure under Conchello "destructive activism"that had stifled the party's intellectual morality. Morfín did not want the PAN to participate in elections if it led to compromises with people that did not believe in its ideological positions. The doctrinal wing outlined its stance when it stated that "*Acción Nacional* reiterates its fundamental convictions, extracting from them the proper methods of solution to those main problems in Mexico...the difficulties which the Party finds itself in multiple occasions should only confirm the urgency to maintain those ends."[45]

Conchello's supporters still remained committed to the pragmatic approach and held considerable influence over the party's internal operations. Morfín even accused Conchello of creating a parallel party with its own structure and funding.[46] When the time arrived to prepare for the 1976 elections, the party was so immobilized by internal bickering that the PAN failed to nominate a presidential for the first time in two decades.

Electoral Reform

Though the PAN's failure to nominate a presidential candidate in 1976 had more to do with its own internal shortcomings rather than any governmental repression, the PRI became concerned that only one name appeared on the presidential ballot. Upon taking office, José Lopez Portillo introduced a series of electoral reforms known as the *Ley Federal de Organizaciones Políticas y Procesos Electorales* (LFOPPE). LFOPPE's provisions included more public funding for opposition parties and gave full autonomy to the Federal Electoral Commission by separating it from the president-controlled Secretary of Government. LFOPPE also created a proportional representation system, allowing opposition parties to have easier access to Congressional seating. Under this system, opposition parties were allotted up to one hundred seats based upon the percentage of votes it received.[47] The PRI knew the proportional representation system would boost opposition numbers in Congress, so it created another 200 majority-representation seats to ensure legislative control.

The PAN split on whether to accept the PRI's LFOPPE laws. The Conchello-led pragmatists advocated its adoption while Morfín-led doctrinists argued for a boycott of LFOPPE. The doctrinal wing stated that LFOPPE produced no real initiatives for the PAN, and still kept the PRI in power. Conchello solidified enough support in the party that the PAN agreed to the government's new terms. When Morfín realized the party's consensus disagreed with his own stance, he resigned from the PAN in 1978.

With the exodus of several influential doctrinal leaders like Morfín, the PAN suffered a significant loss of electoral support. In some regions of the countries, the PAN's organizational structure fell apart completely. Yet, the PAN managed to re-group and won an astounding [43] Congressional seats in the 1979 elections.[48] Though the PAN continued to be a marginal party going into the 1980s, the rise of pragmatic supporters within the party ranks steadily increased the PAN's appeal in all levels of politics.

The PAN has now evolved from a marginalized "opposition" group to a competitive political party in Mexico. Following the 1977 electoral reforms, the PAN steadily increased its presence in the country's political landscape. The party, which had operated more like a spirited club than a true political organization, readily adapted to the country's changing political climate. In 1982, President José Lopez Portillo nationalized the banks, angering the country's financial and business community. This action increased the return of entrepreneurs to the PAN. The re-introduction of the business sector successfully pushed the PAN's pragmatic scope

Carlo Daniel Guerra

even further.[49] PAN representation in the Federal Deputy chamber almost tripled between 1985 and 1988, rising from thirty-two deputies to 101.[50] The PAN won the governorship in Baja California in 1989, marking the first time an opposition party defeated the PRI in a state-wide race. The party's new techniques and vision culminated with Vicente Fox's presidential victory in 2000.

Conclusion:
Challenges to the PAN's Rise to Power

The PAN still faces challenges from within its party and from outside political pressures in light of its recent success. In fact, it seems more difficult for the PAN to operate today than when it was an opposition party. The PAN knew electoral success was a far-fetched dream, and therefore focused their operation on promoting its ideological platform. Even the pragmatic and pro-participation *panistas* knew they had little chance, but believed the government would gradually concede to their charges. When the government conceded to the PAN's demands, the party found itself in a unprecedented position. The party found it difficult implementing its agenda in states and cities where decades of PRI control had entrenched certain types of political practices.[51] The PAN's gain in power forced it to become accountable to supporters and foes alike.

The PAN's new role still begs the question—how does a party maintain internal unity while appealing to a greater audience in hopes of obtaining power? The older *panistas* believed strict interpretation of their ideology would appeal to greater masses, given time and a proper platform. The newer panistas saw campaigning and compromise as the keys to power. While both strategies benefited the PAN at different points in its history, there are serious flaws in both arguments. It is impossible for a political party to stand by its original ideals without ever adapting to new social conditions and challenges. This strategy will only attract a select group of supporters to the party, and its platform will be viewed as reactionary or antiquated. On the other hand, parties that constantly shift their position are seen as weak organizations that pander to current interests, with no regard to pas or future conditions. The PAN

criticized the PRI's "pendulum politics" strategy on just these grounds.

The PAN's dilemma must be solved on two different fronts: how to minimize internal party tensions and how to properly govern in Mexico's new democracy. Factional debates within the PAN have occurred ever since its founding, whether between the secularists during the González Luna reign or Catholic activists during the pragmatic era of 1970s. More recently, older *panistas* criticized Fox for distancing himself from the party during the campaign and appointing non-party members to key Cabinet posts.[52] These internal tensions are only exacerbated by the party's rigid structures and rules. Even though the party's progressive rank-and-file members provide valuable resources to the party, the more conservative leadership tend to appoint like-minded colleagues to replace them, as is evident in the rise of "adherent" members in the PAN in the late 1990s. These members are those who are not granted the full membership status or party decision privileges, but still support the PAN with finances and affiliation. The party must cultivate better relations with these members and bridge the gap between them and the "true members." Since membership requirements have not been dramatically changed since 1939, they still benefit those who believe in the PAN's old way of thinking. Therefore, the PAN needs to reform its membership and promotion requirements. Such changes will boost the party's political prospects, contemporize its legislative and executive agenda to better suit the country's overall conditions, and encourage party members from all parts of the ideological spectrum to obtain leadership positions. The PAN's current structure will ensure that an internal sense of order prevails and any adverse reforms can be stopped before it permanently harms the party.

One essential goal of any political party is to effectively promote a political platform that serves the greater good of a state's citizens. All things considered, how can the PAN continue to properly govern and compete in Mexico? The party must first realize that it will not substantially change seventy years worth of government sponsored corruption and tradition within one presi-

dential term. Fox has struggled to pass his legislative and executive reforms through a divided Congress. While one can argue that his proposals might not be ideal, there seems to be a far simpler explanation. Fox's presidential administration has been the first time in Mexico's history that one entity does not fully control the country. The country's government must get accustomed to certain democratic processes— negotiation, alliance-building, and compromise—which will take substantial amount of time to learn. The PAN must also realize that parties experience gains and losses in a true democracy. The other prominent Mexican political parties, a resurgent PRI and the left-center PRD, are both serious players and both will have political successes and failures in the future.

Sixty years of struggling to maintain a high degree of party unity and to overcome internal struggles might have destroyed many political parties, but the PAN has established itself as a permanent player in Mexico's political landscape. The main goal for the PAN is to remain true to its roots, but not by reverting to its old way of thinking and operation. The party must now interpret its venerable principles and history in a way that appeals to today's Mexico and identifies with the country's current issues. Future *panistas* appreciate the unwavering ideology of their founders, but also apply that steadfastness to the realistic approaches of its current initiatives. Such an approach will ultimately give the PAN the true political platform it has been seeking, and perhaps let it attain the goals it continues to seek.

Bibliography

Primary Resources:

La Nación

La Palabra

Calles, Plutarco Elías. *Declaraciones y Discursos Políticas.* (Mexico: Centro de Documentación Política, 1979)

Cárdenas, Lázaro. *Palabras y Documentos de Lázaro Cárdenas, 1928-1940.* vol. I. (Mexico: Siglos Veintiuno Editoriales, 1978)

Constitución Política de los Estados Unidos Mexicanos. (Mexico: Edición de la Cámara de Diputados, 1963)

Gómez Morín, Manuel. *1915 y Otros Ensayos.* (Mexico: Editorial Jus, 1973)

Gómez Morín, Manuel. *Diez An os de Mexico, 1939-1949.* (Mexico: Documentos del PAN, 1996)

González Luna, Efraín. *El Hombre y el Estado.* (Mexico: Documentos del PAN, 1940)

González Luna, Efraín. *Los Católicos y la Política en Mexico.* (Mexico: Editorial Jus, 1954)

González Luna, Efraín. *Política Humanismo.* (Mexico: EPESSA, 1991)

González Morfín, Efraín. *Discursos.* (Mexico: EPESSA, 1992)

Ibarrola, Adolfo Christieb. *La Oposición.* (Mexico: Documentos del PAN, 1965)

Partido Acción Nacional. *El Mexico de la Oposición I: 8 Plataformas Presidenciales.* (Mexico: EPESSA, 1986)

Partido Acción Nacional. *Estatutos de Acción Nacional.* (Mexico: Documentos del PAN, 1939)

Partido Acción Nacional. *Hacia Mejores Días: Informes de los Presidentes de Acción Nacional 1949-1970.* (Mexico: EPESSA, 1990)

Partido Acción Nacional. *Tres Esquemas.* (Mexico: Documentos del PAN, 1969)

Partido Acción Nacional. *Principios de Doctrina del PAN.* (Mexico: Documentos del PAN, 1939)

Partido Acción Nacional. *Principios de Doctrina—su Proyección en 1965.* (Mexico: Documentos del PAN, 1965)

Partido Acción Nacional. *Programa Mínimo de Acción Política.* (Mexico: Documentos del PAN, 1946)

Vega, Calderón A. *Memorias del PAN*, vol. I. (Mexico: EPESSA, 1992)

Vega, Calderón A. *Memorias del PAN,* vol. II. (Mexico: EPESSA, 1992)

Zapata, Emiliano. *El Plan de Ayala,* ed. Valentín López González. (Cuernavaca: Cuadernos Zapatistas, 1979)

Secondary Sources:

Bailey, David C. *Viva Cristo Rey: The Cristero Rebellion and the Church-State Conflict in Mexico.* (Austin: University of Texas Press, 1974)

Carlo Daniel Guerra

Camín, Hector A. and Meyer, Lorenzo. *In the Shadow of the Mexican Revolution: Contemporary Mexican History, 1910-1989.* (Austin: University of Texas Press, 1993)

Camp, Roderic A. *Politics in Mexico: Decline in Authoritarianism.* 3rd ed. (Oxford: Oxford University Press, 1996)

Casten eda, Jorge. *Perpetuating Power: How Mexican Presidents were Chosen.* (New York: The New Press, 2000)

Cornelius, Wayne A. and Craig, Ann L. *Politics in Mexico: An Introduction and Overview.* (San Diego: Center of U.S.-Mexican Studies, 1988)

Fuentes Díaz, Vicente. *La Democracia Cristiana en Mexico.* (Mexico: Editorial Altiplano, 1972)

Gálvez, Uriel J. and Romero Vadilla, Jorge. *Un PAN que no se come: Biografía de Acción Nacional.* (Mexico: Ediciones de Cultura Popular, 1985)

Grayson, George. *Mexico: The Changing of the Guard.* (New York: Foreign Policy Association, 2001)

Knight, Alan. *The Mexican Revolution: Porfirians, Liberals, and Peasants,* vol. I. (Cambridge: Cambridge University Press, 1986)

Loaeza, Soledad. "El Partido Acción Nacional: de la Oposición a la Impaciencia Electoral," *La Vida Política Mexicana en la Crisis,* ed. Soledad Loaeza and Rafael Segovia. (Mexico: El Colegio de Mexico, 1987)

Mabry, John. *Mexico's Acción Nacional: A Catholic Alternative to Revolution.* (Syracuse: Syracuse University Press, 1974)

Molinar, Juan. "Vicisitudes de una Reforma Electoral," *La Vida Política Mexicana en la Crisis,* ed. Soledad Loaeza and Rafael Segovia. (Mexico: El Colegio de Mexico, 1987)

Mizrahi, Yemile. "Entrepreneurs in Opposition," *Opposition Government in Mexico,* ed. Victoria E. Rodríguez and Peter M. Ward. (Albuquerque: University of New Mexico Press, 1995)

Mizrahi, Yemile. *From Martyrdom to Power: The Partido Acción Nacional in Mexico.* (Notre Dame: University of Notre Dame Press, 2003)

Shirk, David A. *Mexico's New Politics: The PAN and Democratic Change.* (London: Lynne Rienner Publishers, 2005)

Von Sauer, Franz. *The Alienated "Loyal" Opposition.* Albuquerque: University of New Mexico Press, 1974)

Endnotes

1. The ruling party has changed its name three times during its reign. The *Partido Nacional Revolucionario* (PRN) was founded by Plutarco Elias Calles in 1929. Lázaro Cárdenas, president from 1934-1940, reorganized the party and named it the *Partido Revolucionario Mexicano* (PRM). In 1946, under the auspices of President Miguel Alemán, the party took the name that it has used ever since.
2. George W. Grayson, *Mexico: The Changing of the Guard,* (New York: Foreign Policy Association, 2001), p. 3-5.
3. The theory that the PRI changed its ideological platform to coincide with the country's social conditions is known as "pendulum politics"—the government "swung" to the other side of the political spectrum when it deemed appropriate. See Wayne A. Cornelius and Ann L. Craig, *Politics in Mexico: An Introduction and Overview* (San Diego: Center for U.S.-Mexican Studies, 1988).
4. Von Sauer, *The Alienated "Loyal" Opposition,* p. 31-33.
5. Camp, *Politics in Mexico,* p. 11.
6. Ibid., p. 22-23.
7. Cornelius and Craig, *Politics in Mexico,* p. 20.
8. Vega, *Memorias del PAN*, p. 19.
9. Morín, *Diez An os*, p. 31.
10. Ibid., p. 7.
11. Partido Acción Nacional, *Principios de Doctrina,* p. 7.
12. Partido Acción Nacional, *Principios de Doctrina,* p. 8.
13. *La Nación* (Mexico), December 3, 1939.
14. Partido Acción Nacional, *El Mexico de la Oposición 1: 8 Plataformas Presidenciales.* (Mexico: EPESSA, 1986) p. 81.
15. Ibid., p. 81.
16. Efraín González Luna, *Político Humanismo,* (Mexico: EPESSA, 1991) p. 75-78.
17. Ibid., p. 101-104.
18. Efraín González Luna, *El Hombre y el Estado* (Mexico: Documentos del PAN, 1940).
19. Efraín González Luna, *Los Católicos y la Política en Mexico,* (Mexico: Editorial Jus, 1954), p. 8.
20. Ibid., p. 46.
21. Ibid., p. 49.
22. Ibid., p. 68.
23. González Luna, *Humanismo Político,* p. 44.
24. Mizrahi, *From Matyrdom to Power,* p. 59.
25. Ibid., p. 20.
26. Gálvez and Vadillo, *Un PAN que no se come,* p. 38.
27. Vega, *Memorias del PAN,* p. 51.
28. *La Nación* (Mexico) May 7-9, 1943.
29. Von Sauer, *The Alienated "Loyal" Opposition,* p. 104.
30. Gálvez and Vadillio. *Un PAN que no se come,* p. 38.
31. Vega, *Memorias del PAN,* p. 220.
32. Ibid., p. 236-237.
33. The PRM changed its name to the PRI before the 1946 elections.
34. Mabry, *Mexico's Acción Nacional,* p. 48.
35. Morín, *Diez An os,* p. 292.
36. *La Nación* (Mexico) May 19, 1952.
37. Partido Acción Nacional. *8 Plataformas Presidenciales*, p. 84.

38. Mabry, *Mexico's Acción Nacional*, p. 57.

39. Mabry, *Mexico's Acción Nacional,* p. 72.

40. Adolfo Christlieb Ibarrola. *La Oposición.* Ediciones de Acción Nacional, no. 4 (Mexico: Documentos del PAN, 1965) p. 41-48.

41. Ibid., p. 58

42. Partido Acción Nacional. *Hacia Mejores Días: Informes de los Presidentes de Acción Nacional 1949-1970* (Mexico: EPESSA, 1990) p. 216.

43. Soledad Loaeza, "El Partido Acción Nacional: de la Oposición a la Impaciencia Electoral," *La Vida Política Mexicana en la Crisis,* ed. Soledad Loaeza and Rafael Segovia, (Mexico: El Colegio de Mexico, 1987), p. 87.

44. *La Nación* (Mexico) November 15, 1974.

45. Partido Acción Nacional, *8 Plataformas Presidenciales,* p. 138-139.

46. Soledad Loaeza, "El Partido Acción Nacional: de la Leal Oposición Leal a la Impaciencia Electoral" *La Vida Política Mexicana en la Crisis*, p. 88.

47. Juan Molinar, *"*Vicisitudes de una Reforma Electoral," *La Vida Política Mexicana en la Crisis*, p. 32.

48. Mizrahi, *From Martyrdom to Power,* p. 78.

49. Yemile Mizrahi, "Entrepreneurs in the Opposition," *Opposition Government in Mexico*, ed. Victoria E. Rodríguez and Peter M. Ward (Albuquerque: University of New Mexico Press, 1995)

50. David A. Shirk. *Mexico's New Politics: The PAN and Democratic Change.* (London: Lynne Rienner Publishers, 2005), p. 110.

51. Ibid.*,* p. 177.

52. Ibid.*,* p. 233.