Journal of Public and International Affairs
2011

Association of Professional Schools of International Affairs (APSIA)

and

The Woodrow Wilson School of Public and International Affairs Princeton University

Robertson Hall
Princeton University
Princeton, NJ 08544-1013
http://www.princeton.edu/~jpia

The Journal of Public and International Affairs is a joint publication of the Association of Professional Schools of International Affairs (APSIA) and the Woodrow Wilson School of Public and International Affairs at Princeton University. It seeks to publish scholarly articles on a diverse range of subjects, covering the areas of international affairs, development studies, and domestic policy. JPIA is an annual publication that accepts submissions from any advanced student at an APSIA institution. Submissions are reviewed in a blind screening process by an editorial board comprised of students at Princeton University and contributing editors from other APSIA schools. Subscription or copy requests may be sent to jpia@princeton.edu or JPIA, Robertson Hall, Woodrow Wilson School, Princeton University, Princeton, NJ 08544-1013.

1. World Politics—Periodicals. 2. International Relations—Periodicals. 3. Policy sciences—Periodicals. I. Association of Professional Schools in International Affairs. II. Woodrow Wilson School of Public and International Affairs. III. Title: JPIA.

ISSN: 1070-521x

Contents

Letter from the Editors

We are pleased to present the 2011 edition of the *Journal of Public and International Affairs (JPIA)*. Now in its 22nd year, JPIA continues to publish exceptional work of graduate students in public and international affairs. It has provided a unique forum for young scholars to present original research, analysis, and commentary on issues of domestic and international concern.

JPIA also provides an opportunity for professional and intellectual exchange among the members of the Association of Professional Schools of International Affairs (APSIA). Each year, JPIA solicits contributions from graduate students at the APSIA member schools. Contributing editors from each school play a crucial role in screening articles and working with their school's authors to prepare articles for submission. Each spring, contributing editors from the United States, Europe, and Canada meet at Princeton University for a reading weekend to debate and discuss the merits of the submissions.

The eight articles in this volume were selected following a rigorous evaluation of sixty-two articles from twenty schools. The editorial team used electronic prescreening, scoring, and editing throughout the selection process, which allowed the editors to spend more time providing detailed substantive feedback to the contributors. Following the reading weekend, the Princeton editorial team worked closely with each of the authors to provide multiple rounds of substantive feedback and editing.

The aim of this year's journal is to showcase domestic and international policy issues that are currently being debated by policy makers. Christian Phillips examines the American Health Benefit Exchanges, which was created by the Patient Protection and Affordable Care Act (ACA) to improve the affordability of health insurance in the United States. Her paper evaluates whether the Exchanges have fulfilled the conditions for a competitive market for consumers and to what extent states can enhance the ACA regulations to improve affordability and coverage. Philips concludes that while the Exchanges are an improvement, states should consider additional measures to help protect consumers, particularly those facing health crises.

Two articles this year examine state behavior and its impact on foreign policy decisions. Ches Thurber investigates why some states choose to conduct explicit nuclear testing while others simply hedge their capabilities. He argues that understanding these motivations can help policy makers deal with the rapid spread of nuclear technology. The paper adopts a case study approach and draws several conclusions that are then applied to dealing with Iranian nuclear situation. Gregory Hudson's paper uses a game theory model to examine the "balance of financial terror" between the U.S. and China. His paper tests several scenarios to determine the optimal move for each actor. Even though China is a significant holder of U.S. debt, Hudson's game theory models conclude that neither actor can unilaterally change the current status quo without harming their own interests.

Moving to international law and regulation, Rebecca Perlman's article uses the framework of international law to evaluate the use of drones in targeted killings. Her article analyzes the policy from three perspectives and concludes that targeted killings are not supported under the current international regime. She recommends that the U.S. should reframe its legal arguments, tighten definitions governing "self-defense," and work with the international community to propose regula-

tions that reduce the legal uncertainties. Adam Ross discusses the role of private military and security companies (PMSCs), specifically identifying why states do not wish to regulate this industry. He argues that states and private citizens benefit from an unregulated industry, which diminishes the momentum needed to create international regulation. His article concludes by offering potential approaches to motivate states, particularly the United States to regulate PMSCs.

Wes Heinkel and Alexandra Mace study the issue of homegrown terrorism in the United States. They studied twenty-seven terrorist plots and drew conclusions about the frequency, profile, and intentions of the perpetrators. The authors conclude that despite the recent media and political coverage, there are few facts to support the idea that the homegrown terrorist threat and radicalization of Muslim-Americans has increased post-9/11.

Mary Yang uses econometric analysis to test the impact of exchange rate volatility on trade within the East African Community (EAC). Her paper is premised on the argument that by creating a monetary union, the EAC can increase intra-regional trade and economic development. Her findings show a negative relationship between trade and exchange rate volatility, leading to the conclusion that a common currency among EAC countries would positively impact trade.

Turning back to domestic politics, Katie Cristol examines the challenges facing the municipal market with a focus on dispelling the current media attention on short-term problems. She argues that short-term threats such as default and changes to macroeconomic policy are often overemphasized. Instead, it is the long-term issues such as pension and retiree health benefit liabilities, bankruptcy risk, and minimal safety nets for municipalities that threaten the viability of the municipal markets. She proposes that increased transparency and regulation should be incorporated into government policy as well as reviewing existing debt issuance practices.

We extend our sincere gratitude to the Woodrow Wilson School of Public and International Affairs and to APSIA, which make the publishing of JPIA possible each year. In particular, we would like to thank Melissa Lyles, the Woodrow Wilson School's Director of Graduate Programs, for her guidance and support and Leona Rosso-Dzugan of Princeton University Printing Services for her work on the layout design. The journal would not have been possible without the dedication of our exceptional editorial staff. We thank Ian Aucoin, Sophia Peters, Elina Sarkisova, and Mary Yang for their excellent editorial work and Sarah Sieloff, Nazir Harb, and Drew Shaver for their invaluable contributions as associate editors. We are especially grateful to Dan Fitchler and Will Wagner for their outstanding work as Princeton's contributing editors. Finally, a special thanks goes to the contributing editors from the APSIA schools for their efforts in soliciting, evaluating, and editing this year's contributions.

Sarita Vanka and Lynn von Koch-Liebert
JPIA Editors-in-Chief

1

Improving the American Health Benefit Exchanges' Capacity for Increasing Access to Affordable Health Insurance

Christian D. Phillips

The American Health Benefit Exchanges created by the Patient Protection and Affordable Care Act (ACA) are designed to improve the affordability of health insurance by creating a more competitive market for consumers and establishing a mechanism for directly helping individuals pay for the plans that they find. To assess whether the Exchanges are likely to achieve these goals, this paper evaluates their potential to fulfill three basic conditions of a competitive market, and analyzes how well the subsidy and credit systems will function if enrollees face a healthcare crisis. The Exchanges are likely to meet the basic conditions for a robust health insurance market to varying degrees, and for the average family in an average healthcare utilization year, direct assistance through the Exchanges significantly improves affordability. However, in the case of a healthcare crisis that drives utilization and costs up near the upper thresholds of subsidies, there is wide variance in the total impact healthcare costs may have on a family or individual budget. States can

Christian D. Phillips is a 2011 graduate of the Master in Public Affairs program at Princeton University's Woodrow Wilson School of Public and International Affairs. She can be reached at christiandphillips@gmail.com.

enhance several of the provisions laid out in the ACA such as stronger regulation of plan design in the Exchanges, electronic interfaces linking costs to income, healthcare navigators, and others to ensure that the credits and subsidies go far enough when families may need them most and that plans are more affordable to purchase. While federal action has paved the way for increased access to affordable health insurance, state action will determine whether the Exchanges actually fulfill that promise.

I. Introduction

A December 2010 poll by the Kaiser Family Foundation reports that 54 percent of Americans are delaying medical care or treatments due to cost, and that 36 percent of Americans living in households with an income below $40,000 are having trouble paying their medical bills (Kaiser Survey 2010, 5). As the unemployment rate lingers in the double digits, these numbers may worsen as families spend down their savings. For workers who do have steady jobs, the portion of employers offering their employees health insurance continues to decline (Broadus and Thompson 2010, Fig. 2).

The current state of the American economy, combined with declining availability of insurance through employer-sponsored plans, has contributed to a growing number of families struggling to manage healthcare costs. The American Health Benefit Exchanges (Exchanges) established in the 2010 Patient Protection and Affordable Care Act (ACA) could play a vital part in expanding access to affordable health insurance. The Exchanges have typically been assigned two roles in improving health insurance affordability for those who do not have access to employer- or government-sponsored plans. First, the Exchanges are intended to improve the competitive nature of the insurance market by increasing transparency, providing consumers with meaningful choices among plans, and effectively pooling risk. Second, the Exchanges will serve as a venue for the government to provide direct payment assistance to enrollees in the individual insurance market through a system of premium credits and cost-sharing subsidies.[1]

This paper will show that the Exchanges can fulfill both of these roles in increasing affordability, but only to a limited extent. States must augment the minimum measures and regulations laid out by federal law in order to create Exchanges that ensure affordability for individuals and families in the individual insurance market.

The first section of this paper will evaluate the strengths and weaknesses of the Exchanges as more competitive markets. The second section will assess whether the system of premium credits and subsidies is sufficient for individuals and families facing health crises, and discuss the merits and potential shortcomings of the Exchange as a mechanism for direct assistance. Finally, the paper will review the primary finding that while the Exchanges have the potential to represent a more affordable marketplace for insurance, states need to take additional steps to ensure that families in the individual market have access to truly affordable coverage.

II. Exchanges As More Competitive Markets

The American insurance market has become highly concentrated in some areas, prompting some lawmakers, political commentators, and academics to partially blame rising health insurance costs on lack of competition (Robinson 2004; Blumberg 2009; Emmons, Guardado, and Kane 2010). The section title of the ACA containing provisions for the Exchanges – Consumer Choices and Insurance Competition Through Health Benefit Exchanges; Section 1311: Affordable Choices of Health Benefit Plans – signals that lawmakers intended the Exchanges to be a primary remedy for the competitive challenges of the current insurance market.[2]

The underlying question of the Exchanges as a market is whether buying health insurance can be transformed from its current state to a consumer experience similar to buying a car or airplane tickets. The ACA's new regulations include increased transparency measures, incentives for insurers to participate in the market, and experience-gained measures to attempt to mitigate adverse selection. These measures have the potential to make the Exchanges more successful and competitive on price and value than their predecessors, but significant challenges remain in each of these areas.[3]

The Exchanges attempt to establish a more competitive market along the lines of what economist Alain Enthoven describes as "managed competition,"[4] which is preferable to traditional free market competition when dealing with the special nature of products in the health insurance market. For the purposes of this paper, it is assumed that a robustly competitive health insurance market meets the following basic conditions:

1. Consumers must be able to clearly compare plans and insurers in order to assess the best value for their needs.

2. There must be enough options available on the market for consumers to successfully switch plans if they perceive a better value.

3. Adverse selection and other risk pooling issues unique to the health insurance market must be managed effectively.

Recent literature assessing the viability of Exchanges to successfully serve as competitive markets is mixed, largely a result of high variation in the states' response and the low number of states that have completed substantial development of the Exchange structures (Jost 2010; Lueck 2010; Riemer and Enthoven 2009). Apart from the Health Care Connector in Massachusetts and the Dane County Exchange in Wisconsin, there are few successful Exchange models to draw from and determine best practices. However, based on past attempts and analysis of the market incentives that the pools will create, it is clear that while the Exchanges will be a step forward in creating a more competitive insurance market, their success is more likely with further intervention by the states. In crafting interventions, states must balance the consumer benefits of regulation against insurers' aversion to benefit design controls and error on the side of consumers. States will learn what type of regulation is necessary or harmful during the first few years of the Exchanges.

Competitive Market Condition #1: Consumers Able to Clearly Compare Plans

The ACA's new set of regulations for insurance plans offered both within and outside of the Exchange will help increase consumers' ability to compare plans. However, states can improve these policies by framing regulations around the ways in which consumers will actually use the new information.

For example, new reporting requirements for insurers will enhance individuals' opportunities to evaluate the tradeoffs and pricing of different plans. The ACA requires insurers offering plans on the Exchange to disclose enrollment figures, rating practices, cost-sharing when enrollees go out of the network, and other relevant data (Jost 2010, 14). In addition, the Secretary of Health and Human Services is required to develop a standardized benefit and coverage summary for enrollees in all plans, to further ease comparison (Jost 2010, 14).[5]

Another positive development for increasing market transparency is the standardization of some components of health insurance plans. New regulations that require all insurance plans to lift limits on benefits will provide a minimum set of essential health benefits and eliminate cost-sharing for preventive screenings. This will create a minimum standard that could serve as a baseline for individuals comparing plans. However, the utility of this minimum standard may be lost if there is still a high level of variance in plans offered on the Exchange (Lueck 2009, 1). For

example, an individual reviewing ten plans on the Exchange website for the state of Florida may see that every single plan has the same minimum set of benefits, but the similarities end there. One plan may offer higher deductibles, but more covered benefits; another may offer more extensive coverage for treatments associated with hospital stays, while yet another may be geared toward the preventive care needs of families with small children. Studies in behavioral economics strongly suggest that more choices do not necessarily lead to better consumer assessments of value, especially in financially complex situations (Bertrand, Mullainathan, and Shafir 2006).

The baseline for coverage and increased disclosures helps prevent fraudulent plans from operating in the market, but may not go far enough to help create a simple set of comparisons for consumers looking for the best value for their healthcare needs. States should consider imposing limits on the variance in benefit package designs available in order to avoid overwhelming consumers with a myriad of options, as suggested in a 2009 report from the Center for Budget and Policy Priorities (Lueck 2009, 1).

Another component of the ACA that attempts to standardize plans and increase transparency is the use of an actuarial standard to assign relative values to plans. Each Exchange will offer plans separated into four categories by their actuarial values:

- Platinum-tier plans will have an actuarial value of 90 percent;
- Gold-tier plans will have an actuarial value of 80 percent;
- Silver-tier plans will have an actuarial value of 70 percent; and,
- Bronze-tier plans will have an actuarial value of 60 percent.[6]

The ACA's reliance on the actuarial value is intended to illustrate the relative "richness" of plans in covering treatment costs. However, the centrality of the actuarial value standard raises concerns for consumer transparency in two respects. First, arguably very few consumers or even policy makers can state exactly what an actuarial value describes about the value of a plan to an individual consumer. Second, the Center for Budget and Policy Priorities notes that without adequate regulation of plan components, insurers may be able to manipulate plans in a way that allows them to "fix" actuarial values to meet a particular standard, while still providing benefits that may not be of value to certain consumers (Lueck, 2009, 2). Both of these dynamics raise doubts about how effective an

actuarial value standard will be in helping consumers better understand what they are getting for their money.

States should closely monitor how useful the actuarial value tiers are in setting standards and clarifying relative value to consumers during the inaugural years of the Exchanges. In their regularly required reporting to Congress and state legislatures, Exchange administrators and insurance regulators should include details on how insurers use the actuarial values to establish plan components and whether enrollees appear to understand the measurement.

Finally, one of the frequently touted transparency benefits of the new Exchanges in its format is a website that has been described as similar to Travelocity and other travel planning sites. If this web interface is properly designed, it will enhance the capacity of consumers to seek out and compare insurance plans. There is likely to be a sharp learning curve as states unroll their individual Exchange sites and share information about best practices.

States should pay special attention to those who do not have ready or easy access to the Internet, especially consumers who are in the lower income ranges and may be buying insurance for the first time with premium credits and subsidies through the Exchange. In order to assist these lower income insurance purchasers and those consumers who may have trouble moving between public and private plans, states should turn to the team of healthcare navigators that will be required by law for Exchanges.[7] The healthcare navigator role could be expanded to include regular outreach to populations where there is limited access to public libraries with Internet capability, or assisting enrollees with changing insurance needs.

On balance, the transparency measures for Exchanges outlined in the ACA are an improvement over the current market. However, states should consider the benefits of increased plan design regulation, and closely study how consumers use the new information available to them.

Competitive Market Condition #2: Meaningful Set of Plan Choices

Health economist James Robinson and others have suggested that the small number of insurers in some markets has diminished pressure on insurers to compete, and enabled insurers to increase profit margins by passing costs on to consumers (Robinson 2004; Blumberg 2009; Emmons, Guardado, and Kane 2010).[8] Several previous attempts at establishing exchanges or purchasing pools in states or counties suffered from a number of issues, including a failure to attract enough insurers to the pool to create mean-

ingful choices for consumers.

The individual mandate is the most prominent aspect of the ACA that may influence insurers to participate in the Exchange. Exchanges are meant to serve those who do not have ready access to employer- or government-sponsored healthcare, and are likely to be used by many individuals who were not previously in the insurance market. The Exchanges could serve as a primary venue for new customers for the insurance industry, which is likely to initially attract insurers' participation. While the Exchanges represent new opportunities for insurers, whether they can create enough economic incentive for insurers to attempt to penetrate saturated or highly monopolized markets remains to be seen.

The relative appeal of markets outside of the Exchange is an additional key factor in determining how attractive the Exchanges are to new insurers. Several measures in the ACA are designed to create parity between the markets inside and outside of the Exchange to prevent "cherry-picking" by insurers trying to shift healthier individuals to more profitable plans outside of the Exchanges. These measures include a requirement that all plans carry a minimum set of essential health benefits with at least a 60 percent actuarial value, and that identical plans inside and outside of the Exchange have the same pricing. If insurers view measures to increase regulation of plans inside the Exchange as cumbersome, it could dampen their participation in the plan.

Enhancing the attractiveness of the Exchange markets will be a tricky balancing act for state administrators and designers. Increasing the Exchange's leverage over participating insurers may enhance its capacity to demand lower prices or create regulations that help make plans more valuable for consumers. However, these dynamics could make the external market much more attractive to insurers, thus reducing the real choices available to consumers within the Exchange.

The clear advantage for states caught in this balancing act is federal support for premium credits and subsidies. Since the availability of federal assistance is likely to create a large base of new customers for insurers, states should reassure insurers considering entering the market that they will be an aggressive partner in encouraging new enrollees to enter the market.

Competitive Condition #3: Effective Management of Risk Pooling

With notable exceptions in Massachusetts and several smaller programs, previous state-based attempts at risk pooling or insurance Exchanges have failed. Many of those failures have been blamed on uncontrolled adverse

selection, as insurers increasingly shifted healthier and cheaper enrollees out of the pools. States learned that if insurers shift low-risk enrollees out of the Exchanges on a continuing basis, the remaining enrollees in the Exchange would be high-risk and high-cost, which could cause insurers to leave the market and render it less competitive or even dysfunctional.

The ACA's defenses against adverse selection continue to be heavily debated by economists and health policy analysts. The general consensus appears to be that while there are significant measures to reduce adverse selection, there is nothing in the federal legislation that will definitely prevent it.

The ACA provides three promising features to minimize lopsided risk: 1) the premium credits and subsidies are exclusively available within the Exchanges; 2) many of the insurance industry reforms that affect benefit design and cost are the same for plans sold within and outside of the Exchanges; and 3) risk adjustment and reinsurance programs have already been written into the Exchange implementation to help deal with unstable and low-evidence enrollee pools (Cunningham 2010; Lueck 2010; Jost 2010).

These measures are largely a product of lessons learned from the failures of other pooling attempts. However, previous failures also point to a possible need for states to build on the federal requirements and possibly do more to prevent runaway adverse selection from collapsing Exchanges. For example, Alain Enthoven, the American Association of Actuaries, and a number of prominent analysts have raised concerns over the low penalty associated with failure to comply with the individual mandate (Stewart 2010; Cunningham 2010; Bingham 2010). In a recent report, Peter Cunningham of the Center for Health System Change emphasized that some individuals who are eligible for credits and subsidies may have to pay as little as $695 annually for individual mandate penalties—often less than the premiums for a subsidized insurance plan (Cunningham 2010, 2).

States with wealthier, younger, or healthier uninsured populations should study whether increasing the insurance penalty will incentivize those populations to participate in the Exchange. The increased penalty could be marketed to the public much like "sin taxes" on tobacco and alcohol use, but should be carefully evaluated to ensure that it does not indirectly punish uninsured lower-income individuals.

Another risk pitfall that states should avoid is choosing to separate the small employer and individual risk pools within the Exchanges for political reasons. This may result in pools that are too small to withstand healthcare utilization volatility and are subject to lopsided risk. Health law

analyst Timothy Jost has suggested that a policy likely to enlarge the pool and shore up the positive risk in Exchanges is to urge states to allow large employers to enter the Exchanges in 2017, the earliest date permitted by the ACA, and adopt close monitoring to ensure that employers with very high-risk populations are not the only ones transitioning into the Exchange (Jost 2010, 10).

Finally, the continued existence of markets outside of the Exchanges will mean that there is always a possibility of adverse selection. The ACA requirement that insurers in the Exchange sell at least one Gold and one Silver tier plan, and other additional requirements for Exchange certification, could make offering plans both inside and outside of the pool more expensive than simply offering plans outside. There is no rule to preclude an insurer from exclusively selling high cost-sharing plans outside of the Exchange. If low-risk individuals see those plans as more affordable than plans inside the Exchange, adverse selection cycles may occur and lead to negative consequences for the market and consumers (Lueck 2010, 2).

The potential state "fixes" for the dual markets problem with adverse selection pose serious tradeoffs for consumers and insurers. However, plans are likely to be more affordable for consumers if adverse selection is aggressively managed. This could be achieved by requiring all insurers to meet the same requirements inside and outside of the Exchange, such as the requirement to sell both Silver and Gold tier plans in order to sell lower actuarial value plans. Another possibility could be that states may require all sales of insurance to be conducted through the Exchange.[9] Both of these options are likely to meet stiff resistance from insurers, who may claim that such extensive regulation will drive insurers out of the state market entirely. However, in instances where there are captive buyers, sellers (the insurers) will be hard pressed to stay away.

In many ways, the creation of more competitive markets within the Exchanges is one of the most experimental components of the ACA. There will be fifty different test labs in the states, which will learn a great deal about the limits and benefits of managed competition. If states create competitive marketplaces that prioritize the empowerment of consumers with clear information, real choices, and effectively managed risk, insurers will be able to compete on price and quality. By forcing insurers to compete on price, consumers and Exchanges will compel plan issuers to maintain high value for consumers.

Making the insurance market more likely to produce affordable coverage products is one half of the primary role that Exchanges will have in increasing access to affordable care. The other half will be Exchanges' role

as a means for individuals and families to afford seamless coverage that meets their changing needs.

III. Exchanges as a Venue for Direct Assistance to Individuals and Families

The ACA outlines ranges of acceptable levels of expenditure and "affordability" for health insurance. According to Section 1411 of the law, an "unaffordable" employer plan is one that requires an employee contribution of 9.5 percent or greater of the employee's income.[10] In other sections, the law establishes the required caps on healthcare costs for individuals under 400 percent of the Federal Poverty Level (FPL), and stipulates that the fraction of an individual's income that may go towards healthcare premium costs ranges from 2 to 9.5 percent.[11] The ACA's prescribed limits on premium and out of pocket costs are more affordable than the status quo—no limits whatsoever—but still may fall short of true affordability for certain families.

An analysis of 2007 health spending survey data by the Center for Health System Change found that the percent of families with medical bill problems rapidly increased when out of pocket spending on healthcare costs exceeded 2.5 percent of family income (Cunningham 2008, 1). At lower income levels, the same study also concluded that even modest health expenditures below 2.5 percent of income could cause considerable financial strain, especially for those already experiencing trouble paying their medical bills (Cunningham 2008, 3).

Following a brief discussion of the mechanisms for delivering premium credits and cost-sharing subsidies to families and individuals, this section analyzes how a health crisis may render the currently prescribed levels of assistance with health insurance costs inadequate for certain families and individuals. It concludes by suggesting policy solutions to reduce this occurrence.

How the Premium Credits and Cost-sharing Subsidies will Work

Within the ACA's overall scheme for expanding access to health insurance, the Exchanges are meant to be an access point for those who are ineligible for Medicaid or Medicare, and who do not have access to an affordable employer-sponsored plan. In order to make individual plans more affordable, eligible enrollees in certain plans on the Exchange will be able to receive federally funded premium credits and subsidies meant

to effectively cap the cost of health insurance. This assistance is only available on Silver-level plans on the Exchange, and only to eligible enrollees who do not have access to an affordable employer-sponsored plan and do not qualify for Medicaid or Medicare.[12] Table 1 summarizes the premium credits and subsidies provided for in the ACA, according to 2010 FPLs.

Table 1: Subsidies, Limits on Out-of-Pocket (OOP) Expenditures, and Plan Actuarial Values[13]

Under 100% FPL	Medi-Cal	
100-133% FPL[1]	Medi-Cal	
	Max Premium	2% of income
	Max OOP (Indiv/Family)	$2,196/$4,395[2] (1/3 of HDHP in 2014)
	Actuarial Value	94%
134-150% FPL	Max Premium	3-4% of income
	Max OOP (Indiv/Family)	$2,196/$4,395 (1/3 of HDHP in 2014)
	Actuarial Value	94%
151-200% FPL	Max Premium	4-6.3% of income
	Max OOP (Indiv/Family)	$2,196/$4,395 (1/3 of HDHP in 2014)
	Actuarial Value	87%
201-250% FPL	Max Premium	6.3-8.05% of income
	Max OOP (Indiv/Family)	$3,297/$6,593 (1/2 of HDHP in 2014)
	Actuarial Value	73%
251-300% FPL	Max Premium	8.05-9.5% of income
	Max OOP (Indiv/Family)	$3,297/$6,593 (1/2 of HDHP in 2014)
	Actuarial Value	70%
301-400% FPL	Max Premium	9.5% of income
	Max OOP (Indiv/Family)	$4,396/$8,791 (2/3 of HDHP in 2014)
	Actuarial Value	70%

Over 400% FPL	No Premium Credits or Subsidies	
	Max OOP (Indiv/Family)	$6,593/$13,187 (HDHP in 2014)
	Actuarial Value	70%

Table Notes:

1. The first tier of premium credits (2 percent of income) was designed for legal immigrants subject to the five-year bar who have incomes below 133 percent of the FPL, but are ineligible for Medicaid.
2. Out-of-pocket estimates are based on costs for the second lowest Silver plan. These costs are pegged to High Deductible Health Plan's (HDHP) out-of-pocket limits. See Internal Revenue Code, 26 U.S.C. § 223(c). The limits ($5,950 for individual coverage and $11,900 for family coverage in 2010) increase annually and are indexed using cost-of-living adjustments. The amounts in this table are predicted using 2010 as a base year, assuming an annual increase of 2.6 percent. The ACA reductions in HDHP out-of-pocket limits are by 2/3, 1/2 and 1/3, for each respective income tier. ACA § 1402 (c)(1)(A).

Source: Phillips, 2010. "Gains, Gaps and New Choices," Figure 4.

On balance, the ACA's scheme of credits and subsidies for individual purchasers on the Exchange is an improvement on the status quo, which does not impose any limit on how much of a family's income can be spent on healthcare coverage. However, a weakness of the scheme is the relative difficulty in understanding exactly how much of their income a particular individual or family will be expected to pay. There are two types of health insurance costs for consumers: premiums, which are paid regardless of healthcare utilization, and out-of-pocket costs, like copayments, which vary directly with healthcare utilization. The ACA creates two separate caps for both types of costs, despite the fact that one family or individual's income will be paying for both. This separation of caps can lead to misleading conclusions about affordability, and potentially obscure the unsustainably high costs that a family with high healthcare utilization could still be expected to pay.

High Healthcare Utilization and Costs: Credits and Subsidies May Not Go Far Enough

In a year of average health and healthcare utilization, the credits and subsidies offered through the plans may help make healthcare costs more

affordable (Congressional Budget Office 2010, 194). Unfortunately, every year many families face health crises, and conditions that are chronic in nature or require many specialist interventions or regular treatments can quickly lead to escalating healthcare costs. The good news for families who are eligible for assistance in the Exchange is that the program will employ new caps on these costs in 2014. However, those families near the lower end of the income scale may need more assistance than these caps currently permit.

Table 2 presents the maximum amount of a family's income that may be spent on healthcare under the ACA to illustrate the "helpfulness" of the caps in a high healthcare utilization year. The left column of the table includes different family types and the income levels that correspond to that type of family at different percentages of the FPL.[14] The four columns to the right (H+H: Alameda County, for example) contain the percentage of each family's income that would typically go to housing and healthcare in a high utilization year.[15] This combined cost of housing and healthcare is included in order to help contextualize the impact of high healthcare costs on a family's budget. [16]

To understand how families with children eligible for public health insurance programs will factor into the affordability, this analysis assumes that the families live in California, which has one of the more generous levels of eligibility for coverage under the State Children's Insurance Plan at 250 percent of FPL. The four counties listed in the combined healthcare and housing costs are a mix of rural and urban areas (Alameda and Los Angeles County are urban, Fresno County is urban and rural, and Colusa County is rural).

Table 2 illustrates the wide-ranging financial impact that a year of high healthcare utilization may have on families eligible for premium credits and cost-sharing subsidies. A calculation for families at 75 percent of FPL has been included in order to illustrate a particular problem that persists for legal immigrants subject to the five-year bar on federal assistance. Permanent residents with less than five years of status will be required to have health insurance, but will not be eligible for Medicaid.[17] Additionally, the ACA states that those with incomes under 100 percent of FPL will, for the purposes of calculating subsidies and credits, have an assumed income of 100 percent of FPL.[18]

Table 2: Maximum Costs for Healthcare under the ACA

Family Structure	Income	% FPL	Max Health Cost Exposure Per ACA (Dollars)	% of income for Healthcare	H+H Alameda county	H+H: Fresno county	H+H: LA county	H+H: Colusa County
Single Person, Alone (Studio Apartment)	8123	75	2200	27%	169%	123%	166%	122%
	14404	134	2365	16%	97%	70%	95%	70%
	21660	200	4265	20%	73%	55%	72%	55%
	27183	251	5096	19%	61%	47%	60%	47%
	43320	400	9236	21%	48%	39%	47%	39%
Two Adults, No Children (1 Bedroom Apartment)	10928	75	4258	39%	167%	117%	164%	110%
	19379	134	4448	23%	95%	67%	93%	63%
	29140	200	7636	26%	74%	55%	73%	53%
	36570	251	8754	24%	62%	47%	61%	45%
	58280	400	17137	29%	53%	44%	53%	43%
Single Parent, One Child (1 Bedroom Apartment)	10928	75	4258	39%	167%	117%	164%	110%
	19379	134	2365	12%	84%	56%	83%	52%
	29140	200	7636	26%	74%	55%	73%	53%
	36570	251	8754	24%	62%	47%	61%	45%
	58280	400	17137	29%	53%	44%	53%	43%

Two Parents, One Child (2 Bedroom Apartment)	13733	75	4333	32%	133%	94%	131%	88%
	24353	134	4448	18%	86%	60%	88%	60%
	36620	200	8107	22%	67%	50%	69%	50%
	45958	251	9513	21%	57%	43%	58%	43%
	73240	400	18558	25%	48%	39%	49%	39%
One Parent, Two Children (2 Bedroom Apartment)	13733	75	4333	32%	133%	94%	131%	88%
	24353	134	2365	10%	78%	51%	80%	51%
	36620	200	8107	22%	67%	50%	69%	50%
	45958	251	9513	21%	57%	43%	58%	43%
	73240	400	18558	25%	48%	39%	49%	39%
Two parents, Two Children (2 Bedroom Apartment)	16538	75	4408	27%	111%	78%	109%	74%
	29327	134	4448	15%	72%	50%	73%	50%
	44100	200	8578	19%	57%	42%	58%	42%
	55345	251	10271	19%	48%	37%	49%	37%
	88200	400	19979	23%	41%	34%	42%	34%

Source: Calculations performed using UC Berkeley Center for Labor Research and Education Health Reform Calculator and U.S. Department of Housing and Urban Development Fair Market Rent Estimates, 2010.

The top row of each family section demonstrates that very low-income, recently legal permanent resident families may face an impossible financial situation if there is a health crisis. These families may currently opt not to buy insurance and instead utilize community and emergency clinics, but will be required to obtain coverage in 2014. The maximum percentage of income spent on healthcare for this income level and family type is consistently much higher than for the other categories, and in most cases is nearly a third of total income. When combined with housing costs, the situation appears even bleaker.

Aside from very low-income, recent legal immigrants, other families may also face financial hardships during a health crisis. Families just beyond eligibility for the newly expanded Medicaid (134 percent FPL) are at the lower end of the maximum percentage of income going toward healthcare costs, but in nearly all cases could still pay over 10 percent of their income for healthcare. Housing and healthcare costs combined for families in this situation are consistently over 50 percent of income.

Overall, Table 2 illustrates that there is high variance in how well the system of premium credits and subsidies would serve the financial needs of families and individuals facing healthcare crises. Families just above the threshold for Medicaid eligibility (133 percent FPL), for example, could have between 50 and 97 percent of their income absorbed by healthcare and housing expenses during a high utilization period depending on income, family composition, and location. The next section proposes policy recommendations that aim to mitigate the financial impact of a devastating healthcare event for families receiving credits and subsidies.

IV: Conclusions and Policy Recommendations to Improve Seamless, Affordable Coverage in Exchanges[19]

The economic and political climate in the years leading up to the opening date for the Exchanges promises to be full of uncertainty for many issues affecting their implementation. Can strapped state budgets handle the development of an entirely new agency and state program? Will the individual mandate prevail in court? How many more Americans will lose their access to employer-sponsored coverage by 2014? The potential chaos associated with these questions underscores the need for state Exchange developers to advance a set of guiding principles and goals against which implementation policies can be measured. Given the current economic climate, the central guiding principle must be whether Exchanges improve

the likelihood that everyone who needs to purchase healthcare in the individual market will be able to afford it. [20]

The Exchanges are designed to improve the affordability of health insurance by creating a better, more competitive market for consumers and establishing a mechanism for directly helping individuals pay for the plans that they find. The Exchanges, as conceived in the ACA, fulfill each of the three conditions of a competitive market to varying extents. Federal measures to improve transparency are the strongest of the trio. New federal regulations for the insurance industry will help Exchanges avoid adverse selection and attract insurer participation, but much more (and more politically difficult) state action will be necessary to avoid the pitfalls of previous health insurance pools. States should strike a careful balance between plan content regulation and the creation of an attractive environment for insurer participation. State measures that build on the ACA to help alleviate the confusion and inherent complexity of shopping for healthcare coverage will go a long way to ensuring that consumers are empowered in the new Exchange marketplace.

For the average family in an average healthcare utilization year with relatively stable income, direct assistance through the Exchanges will improve affordability. However, in the case of a healthcare crisis that drives utilization and costs near the upper thresholds of subsidies, there is wide variance in the total impact that healthcare costs may have on a family or individual budget.

State Exchange designers should consider further financial assistance measures when health crises push costs towards the upper limits of the caps established by the ACA. The development of electronic medical records and extensive web-based systems for the Exchanges could be an opportunity to develop assistance measures triggered by healthcare spending levels. States should closely study families and individuals who fall into categories where 50 percent or more of income could be spent on housing and healthcare during a high healthcare utilization period. If studies show that certain families in these vulnerable categories are reaching subsidy cap limits, states should consider implementing targeted, automatic supplement systems that are randomly audited. Another approach might be to develop an emergency appeal and grant program that could be automatically initiated when families below a certain income level reach 65 percent of their ACA-designated healthcare-spending cap. Health insurance, once obtained, works best to increase efficiency in healthcare spending and improve outcomes when coverage is consistent. By proactively supporting families before they reach the point of dropping coverage due to inability

to afford coinsurance, states can improve the likelihood of seamless access to care.

States should also take advantage of electronic data availability to monitor the impact of high healthcare spending for recipients of credits and subsidies on participation in other programs. For example, if everyone at 251 percent of FPL who reaches their upper threshold for health spending also enrolls in food or housing assistance programs during the same period, it may indicate a need to either supplement the credits and subsidies, further expand Medicaid eligibility, or both. Exploiting the availability of this electronic data during the course of implementation can help states learn the most efficient and effective ways to supplement the federal subsidies, if necessary.

The administration of individually based premium subsidies and credits for health coverage can be very costly, as demonstrated in a 2008 analysis of the Healthcare Tax Credit Program for displaced workers (Dorn 2008). Under that program, the administration costs for subsidies absorbed 34 percent of the total budget of the program (Dorn 2008). However, in order to make mandatory insurance coverage affordable, the required financial contributions must be flexible enough to realistically reflect real-time circumstances. To discover the best balance between administrative simplicity and individual accommodation, the federal government can give states that have taken the lead grant funds to implement reduced-scale pilot studies of the subsidy and credit system ahead of 2014.

While the Exchanges' system of premium credits and subsidies is a vast improvement over the current situation for individuals and families buying health insurance on the individual market, states should design Exchange policies to address possible shortfalls in assistance that carry potentially serious consequences. State Exchange administrators should prioritize policies that address the needs of many who will be entering the insurance market for the first time, and who may require additional assistance beyond federal provisions to have continuity of coverage and care.

Notes

[1] This paper will only focus on potential enrollees in the individual market pool of the Exchanges. While the small employer pools will have an impact on the individual market in the Exchange, the range of issues involved in the small employer pool are beyond the scope of this paper.

[2] For an overview of the American health insurance market, see Austin et al. 2009.

[3] There have been numerous attempts at creating Exchanges to pool risk for small

business employee and individuals. For example, see Faulkner et al. 2009.

[4] Managed Competition "is defined as a purchasing strategy to obtain maximum value for consumers and employers, using rules for competition derived from microeconomic principles. A sponsor … structures and adjusts the market to overcome attempts by insurers to avoid price competition. The sponsor establishes rules of equity, selects participating plans, manages the enrollment process, creates price-elastic demand, and manages risk selection." (Enthoven 1993, 25)

[5] Another measure to increase transparency—reporting what fraction of revenues are spent on administrative costs and profits—has been met with some skepticism by insurers and others. The usefulness of this figure, called the Medical Loss Ratio, will largely depend on how closely its inputs are regulated and whether insurers alter practices in a way that distorts the comparative usefulness of the ratio.

[6] ACA § 1411.

[7] ACA § 1311. The ACA provides for each Excahnge to establish a group of healthcare navigators, and outlines their duties as including dissemination of information, referrals to public programs, and facilitation of enrollment in Exchange programs.

[8] Since suppliers (healthcare providers and hospitals) often determine demand and costs, Dr. Uwe Reinhardt has raised questions about whether fewer insurers are actually desirable, since fewer insurers could have more leverage to negotiate lower prices for healthcare services and pass those savings on to plan enrollees (Reinhardt 2010).

[9] This option may also create problems for the undocumented immigrants who are excluded by federal law from purchasing health insurance on the Exchange. The elimination of markets outside of the Exchange will mean that millions of families who are denied access to public coverage programs and often do not have access to employer-sponsored plans will be completely shut out of insurance options.

[10] ACA § 1411

[11] ACA § 1402

[12] ACA § 1501. Eligible enrollees do not include undocumented immigrants or incarcerated individuals.

[13] ACA § 1402(c) describes both the actuarial value limits and the out of pocket cost limits.

[14] The UC Berkeley Center for Labor Research and Education Health Reform Calculator was used to estimate the maximum healthcare expenses that a family facing high healthcare utilization would be expected to pay under the ACA. The calculator can be found at http://laborcenter.berkeley.edu/healthpolicy/index.shtml under the heading: "Calculator: How Much Will a Family Spend Under the Health Reform Law?" It is designed to calculate maximum possible financial exposure

for healthcare costs under the ACA, based on family size and income.

[15] To determine housing costs, estimates of the most conservative housing options for each family's size and composition were used. This was supplemented by the U.S. Housing and Urban Development Departments' estimates of the 50th percentile rental costs for that type of unit in each county. See the Fair Market Rent figures for 2010 at http://www.huduser.org/portal/datasets/fmr.html.

[16] The cost of housing was used to help illustrate the potential financial burden of high health costs due to the tendency for housing, whether under a rental, lease or mortgage contract, to be a long-term cost that is not easy to adjust or eliminate. Other costs in a family's budget, while extremely necessary, may vary month to month.

[17] ACA § 1411

[18] ACA § 1401(c)(1)(B)

[19] The affordability issues also point to a broader need to reconsider the way the government calculates financial hardship. See Blank 2007; Hacker et al. 2010.

[20] The Exchanges, and in fact the entire ACA, exclude undocumented immigrants. The Congressional Budget Office estimates that in 2019, approximately one-third of the 23 million still uninsured will be undocumented immigrants (Congressional Budget Office 2010).

References

Aaron, Henry. 2010. Systemic Reform of Health Care Delivery and Payment. *The Economists' Voice* 7(5).

Adamy, Janet. 2010. Court Strikes at Health Law. *Wall Street Journal*, December 13.

Austin, D. Andrew, and Thomas Hungerford. 2009. *The Market Structure of the Health Insurance Industry*. Congressional Research Service, November 17.

Autor, David. 2010. *The Polarization of Job Opportunities in the U.S. Labor Market*. The Brookings Institution – The Hamilton Project.

Bertrand, Marianne, Sendhil Mullainathan, and Eldar Shafir. 2006. Behavioral Economics and Marketing in Aid of Decision Making among the Poor. *Journal of Public Policy & Marketing* 25(1): 8–23.

Bingham, Alfred. 2010. Letter to Congressional Leaders Re: Patient Protection and Affordable Care Act (H.R. 3590) and Affordable Health Care for America Act (H.R. 3962), January 10.

Blank, Rebecca. 2007. How to Improve Poverty Measurement in the United States. Presented at the Presidential Address for the Association of Public Policy and Management.

Blumberg, Linda. 2009. Improving Health Insurance Markets and Promoting Competition under Health Care Reform. Testimony before the House Committee on Ways and Means, April 22.

Blumberg, Linda and Karen Pollitz. 2009. *Health Insurance Exchanges: Organizing Health Insurance Marketplaces to Promote Health Reform Goals*. Urban Institute.

Broadus, Matt, and Rory Thompson. 2010. *Employer-Based Health Coverage Declined Sharply Over Past Decade*. Center on Budget and Policy Priorities, December 1.

Buettgens, Matthew, and John Holahan. 2010. *America Under the Affordable Care Act*. Urban Institute, December 7.

Congressional Budget Office. 2010. Selected CBO Publications Related to Health Care Legislation, 2009-2010.

———. 2008. *Key Issues in Analyzing Major Health Insurance Proposals*.

Cunningham, Peter. 2010. *Who are the Uninsured Eligible for Premium Subsidies in the Health Insurance Exchanges?* Center for Health System Change. Research Brief No. 18.

———. 2008. *Living on the Edge: Health Care Expenses Strain Family Budgets*. Center for Studying Health System Change.

Curtis, Rick. 2009. *What Health Insurance Exchanges or Choice Pools Can and Can't Do about Risks and Costs.* Institute for Health Policy Solutions.

Dorn, Stan. 2008. *Health Coverage Tax Credits: A Small Program Offering Large Policy Lessons*. Urban Institute.

Duggan, Mark and Robert Kocher. 2010. The Economics, Opportunities, and Challenges of Health Insurance Exchanges. *The Economists' Voice* 7(5).

Emmons, David, Jose Guardado, and Carol Kane. 2010. *Competition in Health Insurance: A Comprehensive Study of U.S. Markets, 2010 Update*. American Medical Association.

Enthoven, Alain. 1993. The History and Principles of Managed Competition. *Health Affairs* 12, Supplement: 24-48.

Faulkner, Deb, Amy Lischko, and Deborah Chollet. 2009. *Considering a Health Insurance Exchange: Lessons from the Rhode Island Experience*. Robert Wood Johnson Foundation – State Coverage Initiatives.

Gawande, Atul. 2009. McAllen, Texas and the High Cost of Health Care. *The New Yorker*, June 1.

Hacker, Jacob, Gregory Huber, Phillip Rehm, Mark Schlesinger, and Rob Valletta. 2010. *Economic Security Index: Putting a Face on American Economic Security*. The Rockefeller Foundation.

Herszenhorn, David. 2010. Obama Vows to Keep Fighting for Health Care Overhaul. *New York Times*, January 22.

Holahan, John. 2011. The 2007–09 Recession and Health Insurance Coverage. *Health Affairs* 30(1): 145-152.

Jost, Timothy. 2010. *Health Insurance Exchanges and the Affordable Care Act: Key Policy Issues*. The Commonwealth Fund.

———. 2009. Health Insurance Exchanges: Legal Issues. *Journal of Law, Medicine and Ethics* 37, Supplement s2: 51-70.

Kaiser Family Foundation. 2010. *Summary of New Health Reform Law*. Publication 8061.

———. 2010. Kaiser Health Tracking Poll -- December 2010 – Findings.

Kingsdale, Jon, and John Bertko. 2010. Insurance Exchanges Under Health Reform: Six Design Issues For The States. *Health Affairs* 29(6): 1158 -1163.

Long, Stephen H., and M. Susan Marquis. 2001. Have Small-Group Health Insurance Purchasing Alliances Increased Coverage? *Health Affairs* 20(1): 154 -163.

Lueck, Sarah. 2010. *States Should Structure Insurance Exchanges to Minimize Adverse Selection.* Center for Budget and Policy Priorities.

———. 2009. *Rules of the Road: How an Insurance Exchange Can Pool Risk and Protect Enrollees*. Center for Budget and Policy Priorities.

McGarr, Cappy. 2009. A Texas-Sized Health Care Failure. *New York Times*, October 6.

McKethan, Aaron, Adam Wilk, and Wes Joines. 2007. *Options to Expand Health Insurance Coverage for Workers in Small Businesses in North Carolina*. The Lewin Group.

National Health Law Program. 2010. National Health Law Program Analysis of the Health Care Reform Law.

Obama's Remarks at the Health Care Bill Signing. 2010. *New York Times*, March 23.

Pecquet, Julien. 2010. President Obama Unveils Patient's Bill of Rights. *Healthwatch: The Hill's Healthcare Blog*, June 22.

Phillips, Christian D. 2010. *Gains, Gaps and New Choices: The Impact of the Affordable Care Act in California*. UC Berkeley Center for Health, Economic and Family Security.

Ravner, Julie. 2010. Friends Of Health Overhaul Defend It In Federal Court Case. *Shots: NPR Health Blog*, November 16.

Reinhardt, Uwe. 2010. Does Competition Benefit Health-Insurance Subscribers. *New York Times Economix Blog*, October 29.

Riemer, David, and Alain Enthoven. 2009. The Only Public Health Plan We Need. *New York Times*, June 25.

Robinson, James C. 2004. Consolidation and the Transformation of Competition in Health Insurance. *Health Affairs* 23(6): 11 -24.

Short, Pamela, Deborah Graefe, and Cathy Schoen.2003. *Churn, Churn, Churn: How Instability of Health Insurance Shapes America's Uninsured Problem*. The Commonwealth Fund.

Stewart, Ian. 2010. Ghost of 'Adverse Selection' Looms over Health-Care Reform Effort. *Capitol Weekly*, December 27.

Tulley, Shawn. 2010. Health Care's New Hidden Danger. *Fortune Magazine*, February 2.

Wicks, Elliot. 2009. *Building a National Insurance Exchange: Lessons from California*. The California Healthcare Foundation.

York, Anthony. 2010. Schwarzenegger Names Susan Kennedy to Top Healthcare Post. *Los Angeles Times PolitiCal Blog*, December 31.

2

A Step Short of the Bomb: Explaining the Strategy of Nuclear Hedging

Ches Thurber

The global spread of technology will inevitably result in more states gaining access to the scientific and technical means to create a nuclear weapon. In order to confront this reality, policymakers and analysts must develop a better understanding of why some states feel compelled to conduct overt nuclear tests while others are content to pursue a strategy of hedging: developing the capability but not actually testing or deploying nuclear weapons. Using case studies from Japan and South Asia, this article seeks to explain nuclear policy through a combination of two factors. First, states attempt to maximize their relative security vis-à-vis their rivals by balancing the value of deterrence with the risk of proliferation. Secondly, domestic political sentiment and the balance of power amongst competing bureaucratic factions may either enable restraint or push a state toward conducting a nuclear test. Finally, by applying these two factors to the case of Iran, this article will evaluate the drivers of Iranian nuclear behavior and offer policy recommendations to increase the odds that Iran will pursue a latent, rather than an overt, nuclear capability.

Ches Thurber is a doctoral student at Tufts University's The Fletcher School of Law and Diplomacy, where he focuses on international security. He can be reached at richard.thurber@tufts.edu.

I. Introduction

In the wake of yet another failed negotiation round in January 2011, the Islamic Republic of Iran appears increasingly unlikely to cease nuclear technology development. For Iran, the perceived benefits of an active nuclear program—whether in terms of security, technological advancement, or national prestige—outweigh anything the West can offer in compensation. While damage inflicted by the "Stuxnet" computer virus has reportedly slowed Iran's progress, nuclear security analysts still believe that Iran may be able to develop both the technological expertise as well as the sufficient quantity of fissile material necessary to construct a nuclear device in as few as three years (Dilanian 2011). Even more worrisome is that Iran is not alone. As technological barriers are shattered, nuclear capability—once reserved for the world's most advanced nations—is coming within reach of dozens of nations. However, despite the alarming inevitability of nuclear capability proliferation, nuclear *weapon* proliferation is not inevitable. Instead, we may see states adopt a strategy of nuclear "hedging;" whereby they seek to acquire the technology necessary to construct a nuclear weapon but refrain from becoming overt nuclear powers. The key question will become how nuclear-capable states can be dissuaded from testing and deploying the world's most dangerous weapon.

Winston Churchill provided an early articulation of nuclear hedging in 1951 when he summarized Britain's nuclear strategy: "I have never wished since our decision during the war that England should start the manufacture of atomic bombs. Research, however, must be energetically pursued. We should have the art rather than the article" (Gowning 1976, 401; Levite 2002, 70). Some analysts have proposed that contemporary Iran may be following a similar strategy of seeking only to acquire a "latent" nuclear capacity (Cole 2009). This idea has also been gaining traction in policy circles. In a January 2010 memo to President Barack Obama, Secretary of Defense Robert Gates expressed concern that Iran might develop all of the materials necessary to build a bomb but stop short of actually assembling it (Sanger and Shanker 2010).

The Churchill model, however, is not cause for optimism; the United Kingdom tested a nuclear device in 1952, the year after Churchill's statement. Indeed, some of the most obvious benefits of nuclear weapons possession do not seem to apply when a state has only a latent nuclear capability. For instance, nuclear latency is not as effective a deterrent as actual possession and deployment of a weapon as it poses no threat for counterattack. Neither is nuclear latency likely to galvanize national pride

in the same way as a country's first nuclear test. Despite these limitations, many states with the requisite materials and capabilities have opted not to test or deploy a nuclear device. Argentina, Belgium, Brazil, Germany, Japan, the Netherlands, South Africa, and Sweden could all be considered latent nuclear powers today. Why have these states been able to maintain a strategy of nuclear hedging, while others have felt compelled to test and deploy weapons?

This article will argue that nuclear policy can be explained through a combination of two factors. First, states seek to maximize their relative security vis-à-vis their rivals by balancing the value of deterrence with the risk of proliferation. Secondly, domestic political sentiment and the balance of power amongst competing bureaucratic factions may either enable restraint or push a state toward conducting a nuclear test. To illuminate these factors, the paper will present definitions of nuclear latency and nuclear hedging and apply theories of nuclear proliferation to these concepts. It will then examine two case studies, using the framework described above to explain why Japan has remained a latent nuclear state in the face of rising regional tensions, while India and Pakistan felt compelled to become full-fledged nuclear states after a period of latency. The argument will then be applied to the current challenge of Iran's nuclear development in attempt to provide a better understanding of the drivers of Iran's nuclear policy. The article concludes with policy recommendations to increase the chances that Iran will pursue a latent rather than an overt nuclear capability.

II. Defining Latency and Hedging

The traditional demarcation of a nuclear power has been the successful detonation of a nuclear device. However, the proliferation of advanced nuclear technologies for ostensibly civilian use has made the once clear "red line" murky. It is now possible for a nonnuclear state to develop a "civilian" nuclear program that puts it within months, or even weeks, of being able to construct a weapon, all while maintaining compliance with Nuclear Non-Proliferation Treaty (NPT) and International Atomic Energy Agency (IAEA) regulations. Furthermore, a number of states have sought to conceal their nuclear weapons capability. Israel is widely acknowledged to have upwards of one hundred nuclear weapons, though it has never overtly tested or publicly acknowledged its nuclear status. South Africa assembled six nuclear weapons in the 1980s, though their existence was not publicly confirmed until 1993, three years after they were dismantled (Nuclear Threat Initiative 2010b).

Such ambiguity requires the international community to pay closer

attention to states' nuclear programs early in the development process. Wohlstetter et al. have argued that the nuclear threshold must now be pushed back. They claim that states with the capabilities to produce a weapon should be considered nuclear states just as states that have actually tested devices:

> If, in fact, technological transfers can bring a "nonnuclear weapon state" within weeks, days or even hours of the ability to use a nuclear explosive, in the operational sense that "non-nuclear weapon state" will have nuclear weapons. The point is even more fundamental than the fact that effective safeguards mean timely warning. A necessary condition for timely warning is that there be a substantial elapsed time. But if there is no substantial elapsed time before a government may use nuclear weapons, in effect it *has* them (Albert Wohlstetter, Jones, and Roberta Wohlstetter 1979, 39).

Wohlstetter et al. thus provide an early conception of "latent" nuclear capacity. In doing so, they highlight one of the key challenges presented by nuclear latency: while NPT requirements and IAEA inspections are supposed to prevent a state from diverting civilian materials toward military use, nuclear latency might allow a state to renounce the NPT, throw out inspectors, and assemble a nuclear weapon in an extremely short time frame. Such a scenario would leave the international community few, if any, opportunities to intervene.

"Hedging" describes a state's strategy of seeking to develop a latent nuclear capability without actually testing a device. Levite defines nuclear hedging as "a national strategy of maintaining, or at least appearing to maintain, a viable option for the relatively rapid acquisition of nuclear weapons, based on an indigenous technical capacity to produce them within a relatively short timeframe ranging from several weeks to a few years" (Levite 2002, 69).

While latency refers to a state's capability, hedging refers to a state's policy of seeking or maintaining nuclear latency. Rather than enter into the debate over what level of nuclear capability constitutes nuclear "latency," this article will focus more specifically on the strategy of hedging. The article also addresses the question of what external and internal political pressures convince some states to pursue a policy of hedging while others feel compelled to overtly test and declare their nuclear weapons.

III. Theories of Nuclear Pursuit and Nuclear Abstinence

The field of international relations includes an expansive literature on the causes of nuclear proliferation, why some states seek these weapons and others do not, and why some states relinquish their nuclear weapons. Policy decisions in this realm were traditionally considered binary: a state could either go nuclear or not. Nuclear latency and the related strategy of nuclear hedging open a spectrum of gray between these stark black and white choices. The question becomes not just whether or not a state will go nuclear, but up to what point will it develop its nuclear capabilities. New analysis is needed to address this question; yet the traditional theories of nuclear proliferation can still be applied. The principal variables that govern whether a state may or may not seek a weapon—the external security environment and domestic political pressures—may also help inform whether a state might pursue a strategy of nuclear hedging and to what degree it might seek to develop its nuclear program.

The Security Environment

Realist scholars consider external security conditions to be the primary driver of global nuclear proliferation. In this model, states seek nuclear weapons when they face a significant security threat that they cannot confront solely through conventional military means. This threat may take the form of a rival with superior conventional forces or one with a nuclear capability. Nuclear weapons scholar Scott Sagan has noted that proliferation often occurs in pairs or bunches: as one state acquires a nuclear weapon, its rivals feel compelled to acquire it as well (Sagan 1996, 57). History provides ample evidence for this theory. The knowledge that Germany was seeking a nuclear device compelled the United States, United Kingdom, and Soviet Union to scramble to develop a weapon first. India sought to acquire a weapon after neighboring China tested a device, and Pakistan conducted its first test just two weeks after India tested weapons in 1998.

This logic also helps explain why states may *not* want to acquire nuclear weapons. While a state might seek to develop a weapon if its rival has one, it may also choose to avoid provoking proliferation by its neighbors. If a state has a conventional military advantage, it will likely see a situation in which both it and its rival acquire a nuclear weapon—the great equalizer—as a decrease in its relative power. Thus, states will pursue a policy that Professor T.V. Paul terms "prudential realism," in which they seek to balance their desire to maximize their own capabilities with a desire

to minimize the degree to which they are perceived as a threat to their neighbors (Paul 2000, 5). Through the lens of prudential realism, nuclear latency becomes an attractive policy option. It allows a state to develop the capabilities to quickly build a weapon in the event of escalating regional tensions, but in the meantime is far less provocative than actually testing and deploying a weapon.

An alternative policy may be to seek security guarantees from a nuclear power. A nuclear umbrella may have a deterring effect on a state's adversaries without pushing them to develop a weapon of their own. In tandem, nuclear latency and inclusion in a nuclear umbrella may be a particularly effective policy package. A latent nuclear capability alone may not be a sufficient deterrent, as a rival can always attack before the latent state is able to assemble and deploy a weapon. Moreover, the nuclear umbrella strategy carries the risk that the ally's commitment to the security guarantee may waiver in the future. But, when combined, the nuclear umbrella deters potential adversaries while the latent nuclear capability provides an insurance policy in case that security guarantee weakens.

Domestic Political Dynamics

While external security factors provide a clear and relatively simple explanation for nuclear behavior, states' nuclear decision-making may be strongly influenced by internal factors as well. These domestic pressures may take multiple forms. Sagan's "domestic politics model" focuses on the effect of bureaucratic interests on nuclear policy (Sagan 1996, 55). He posits that nuclear programs are promoted by a coalition of industrial, scientific, and military actors that seek to benefit from a nuclear weapons development program. When these interest groups are able to unite and win the battle of bureaucratic politics, they can push a government to pursue a weapon even when external security conditions have not changed. Sagan points to India and South Africa as cases where internal bureaucratic politics can better explain the timing of key nuclear decisions than security factors. However, if domestic interest groups opposed to nuclear weapons are better organized and more readily mobilized, bureaucratic interests can push in the other direction. For a government caught between two competing bureaucratic coalitions, Levite argues that the strategy of nuclear hedging may carry the lowest political costs (Levite 2002, 74).

Popular beliefs and norms regarding nuclear weapons can also influence a government's nuclear strategy. As with bureaucratic coalitions, these norms can push for or against weaponization. The domestic public may see nuclear weapons as a symbol of prestige and a necessary step in order

for a nation to assume its rightful place among elite global powers. On the other hand, nuclear weapons can be seen as a taboo and contrary to a nation's core values (Sagan 1996, 76). Some states exhibit both of these normative trends; in these cases, nuclear latency could provide a compromise solution. Developing nuclear capabilities can allow political leaders to rally national pride over their country's scientific and technological accomplishments. For example they may choose to publicly announce and celebrate key milestones of nuclear development, such as the successful uranium enrichment or the construction of a new research reactor. However, by pursuing only a latent nuclear capability, these leaders are still able to avoid the risks that would come from actually testing a nuclear weapon.

The way in which a government responds to bureaucratic and popular pressures also depends on the nature of the regime in power. Changes in government have often been pivotal moments in states' nuclear trajectories. It is no coincidence that South Africa gave up its nuclear weapons program during its transition to a post-apartheid regime. Argentina and Brazil similarly abandoned their weapons programs as they moved toward liberalization. India's on-again, off-again posture towards further testing and weaponization from 1974 to 1998 corresponded with governmental changes (Sagan 1996, 66). Changes in government could increase the likelihood that a state will eventually test a weapon, based on the government's perception of its security environment or the ideology of its political constituency.

IV. Japan: The Classic Example of Latency

Japanese nuclear policy is inevitably shaped by the country's tragic distinction as the only nation to be the victim of a nuclear attack. Japan's constitution specifically renounces war and the use of offensive weapons. And while historians continue to debate the degree to which this language was drafted at the demands of the American occupiers at the end of World War II (Shoichi 1998, 82-83), it has become a firmly embedded principle of Japanese policy. Japan reiterated these principles and applied them specifically to the issue of nuclear weapons in the 1955 Atomic Energy Basic Law, which restricts the use of nuclear energy to purely peaceful purposes. In 1968, the Japanese Diet, the country's legislative body, adopted the "Three Non-Nuclear Principles," which stipulated that Japan would not manufacture, possess, or permit the introduction of nuclear weapons into its territory.

Despite its unique history and the resulting policies banning nuclear weapons, Japan has developed an extensive nuclear energy program. Nuclear

power plants account for 30 percent of Japan's energy production, and the country has the third highest number of nuclear reactors in the world with thirty-five (Nuclear Threat Initiative 2009). Furthermore, Japan has an active program for reprocessing spent nuclear fuel. As a result, it has the largest stockpiles of separated plutonium of any non-nuclear state (International Panel on Fissile Materials 2009, 23). The Japanese government maintains that this material will eventually be used as fuel in new reactors. In the meantime, however, the plutonium quantities are reportedly sufficient to produce up to 10,000 warheads (Nuclear Threat Initiative 2009). Analysts speculate that Japan is mere months away from being able to assemble a nuclear warhead if it so chooses (Levite 2002, 71). Moreover, in recent years, Japanese leadership has shown a willingness to broach the topic of reconsidering its nuclear policy. In 1999, Vice Defense Minister Shingo Nishimura told a magazine that the Japanese parliament should "consider that Japan might be better off if it armed itself with nuclear weapons," and in 2002 then Deputy Chief Cabinet Secretary (and later Prime Minister) Shinzo Abe argued that Japan's laws do "not necessarily ban the possession of nuclear weapons as long as they are kept at a minimum and are tactical" (Campbell and Sunohara 2004, 229).

The Security Environment: Japan under the U.S. Nuclear Umbrella

The protection of the U.S. nuclear umbrella has provided Japan with a strong deterrent while allowing it to maintain its nuclear abstinence. However, recent changes in the Asia-Pacific security environment may strain this policy. North Korea's nuclear tests in 2006 and 2009, as well as its missile tests over Japanese airspace, have led to a widespread perception that North Korea poses an escalating threat to Japan. China's rapid economic expansion and the corresponding surge in its military and defense spending represent another potential threat to Japan and the region at large. Furthermore, China is aggressively pursuing a nuclear energy program that could involve reprocessing on a large scale. This could result in China also having large stockpiles of separated plutonium (Jane's Information Group 2011). The end of the Cold War has compounded these potential threats by calling into question the United States' security commitment to Japan. Without the threat of the Soviet Union, and with its forces extended in Iraq and Afghanistan, would the United States still be able and willing to come to Japan's defense (Hughes 2007, 72)?

Japan has responded to these challenges by redoubling its efforts to maintain its close security relationship with the United States. For example,

it has sought explicit clarification that the United States would respond with force to North Korean aggression (Hughes 2007, 76). Such clarification was intended not only to provide greater confidence to Japan's leaders and population, but also to deter North Korea by emphasizing the potential consequences of aggression. It has also increased its military participation in alliance activities, including sending Japanese forces to Iraq in 2003 (Hughes 2007, 79).

The political costs of developing nuclear weapons are also likely to have influenced Japan's nuclear strategy. Going fully nuclear would anger the United States and thus damage its most important security relationship. Furthermore, it would likely escalate tensions with China and could spark a regional build-up of nuclear arms. Under these constraints, nuclear latency represents the optimal strategy for ensuring Japanese security. Since the U.S. nuclear umbrella currently provides an effective deterrent, Japan can maintain its long-held position against an internal nuclear weapons program without unduly compromising its security. But in the event that the U.S. security guarantee wanes, or is perceived by Japan's rivals to wane, Japan is not stuck at square one—it can quickly become a nuclear power if changes in the security environment convince it that it must.

Domestic Political Dynamics: The Shadow of Hiroshima and Nagasaki

Changes in Japan's regional security environment have sparked a revival of the nuclear debate within its domestic politics. In 2007, Prime Minister Shinzo Abe declared that acquiring a nuclear weapon might be legal under Japan's constitution, as it could be considered a defensive weapon (Hughes 2007, 84). However, such arguments are unpopular among the Japanese polity. Both the Basic Law of 1955 and the Three Principles of 1968 clearly prohibit Japanese acquisition of a bomb. Moreover, Japanese public opinion remains firmly against nuclear weapons. Even after the first North Korean nuclear test in 2006, polls indicated that only 17 percent of the Japanese population supported the idea of nuclear acquisition (Hughes 2007, 89). With such popular opposition and an increasingly competitive electoral climate, it would be politically reckless for a Japanese government to go forward with testing a nuclear weapon. Japan's nuclear latency can thus be explained both by its security interests and by the strong domestic pressures against a weapons program.

V. The Failure to Maintain Latency in South Asia

While the security environment and domestic politics have led Japan to maintain a latent nuclear capability, these same two factors led to the opposite outcome in India and Pakistan. South Asia presents a case where two rival states went beyond latency to fully test and deploy nuclear weapons. India first tested a nuclear device in 1974, largely in response to the Sino-Indian War and China's successful nuclear test in 1964. But India did not conduct any further tests or deploy deliverable nuclear weapons for another thirty-four years. Thus, while most would consider India to have been a full nuclear power as a result of its 1974 test, it more closely resembled a latent nuclear state. That is, it sought to demonstrate its nuclear potential but without incurring the costs of escalation that could have resulted from testing additional devices or mounting warheads on missiles. From the perspectives of threat and deterrence, the quasi-latent status of India's nuclear program also meant that in the event of a crisis, adversaries might question whether India truly had the capability to use nuclear force. Analysts have questioned the efficacy of India's 1974 nuclear test (Perkovich 2002, 180), and Prime Minister Indira Gandhi cancelled planned additional tests in 1982 (Nuclear Threat Initiative 2010c). It is far from clear when and if India actually had the capability to use a nuclear weapon in a military context during this time period.

In the mid-1970s after a series of wars with India, Pakistan began a nuclear program focused on uranium enrichment. By 1990, it was believed to have become a virtual nuclear power, possessing stockpiles of highly enriched uranium (HEU) estimated at between 580 and 800 kg—enough to make thirty to fifty bombs (Nuclear Threat Initiative 2010d). Pakistan's Foreign Secretary Shahryar Khan stated publicly in 1992 that Pakistan had "all the elements which, if hooked together, would become a [nuclear] device" (Ahmed 1999, 190).

Thus for a period of eight years beginning roughly in 1990, both Pakistan and India could be considered latent nuclear states. But on May 11 and May 13, 1998, India conducted its first nuclear tests in twenty-four years and subsequently declared itself a nuclear power. Pakistan responded almost immediately with tests of its own on May 28 and May 30 of the same year.

The Security Environment: A Security Dilemma in South Asia

While both India and Pakistan were pursuing nuclear hedging strategies

in the early and mid- 1990s, they also continually worked to improve their weaponization and delivery systems. In 1994, knowing that Pakistan had stockpiles of nuclear material, India developed the capacity to deliver a nuclear weapon by aircraft and subsequently designed a warhead that could be placed on its Prithvi-1 missile (Nuclear Threat Initiative 2010c). Pakistan responded by similarly seeking to improve its delivery missiles, acquiring key technologies from China and North Korea (Nuclear Threat Initiative 2010d).

Unlike Japan, India lacked a security guarantee from a nuclear power that might have strengthened its faith in deterrence in the face of Pakistan's increasing military capabilities. Indian leaders place its country's nuclear policies in the context of its history of non-alignment with great powers. As such, it is forced to rely on its own military capabilities for its defense. Jaswant Singh, a defense advisor to India's Prime Minister Atal Bihari Vajpayee, explains the rationale behind India's 1974 nuclear test in *Foreign Affairs* by saying that "with no international security guarantees forthcoming, nuclear abstinence by India alone seemed increasingly worrisome" (Singh 1998, 42). According to Singh, it was the same logic pushed India to renewed testing in 1998. While India raised concerns about the nuclear assistance that China was providing to Pakistan, "the United States was either unwilling or unable to restrain China" (Singh 1998, 46). India thus concluded that it could not rely on foreign help and that the only way to be sure of deterring Pakistan was to conduct another test to demonstrate its capabilities and signal its willingness to respond in kind to a potential nuclear attack.

Once India had tested its weapons, Pakistan, unsurprisingly, conducted a test as well. The potential negative consequence that had prevented Pakistan from testing—that it would spark renewed nuclear proliferation in India—was no longer applicable after the Indian demonstration. Through the logic of Paul's prudential realism, Pakistan no longer had anything to lose. While Pakistan has historically maintained a close relationship with China, it lacks the type of security guarantee that the United States provides to Japan and NATO member states. A Pakistani official declared, "We will never be able to remove the nuclear imbalance if we do not follow suit with our own explosion" (Ahmed 1999, 194). This supports the inference that nuclear latency is more difficult to maintain in arms races among rivals without security guarantees from nuclear powers.

Domestic Political Dynamics: Nuclear Weapons and National Pride

Rather than act as a restraint, domestic political dynamics in both India and Pakistan propelled the countries toward testing nuclear devices. Both felt compelled to demonstrate their capabilities not only for the security rationale of deterrence, but also as a symbol of national prestige. Jaswant Singh wrote, "Nuclear weapons remain a key indicator of state power. Since this currency is operational in large parts of the globe, India was left with no choice but to update and validate the capability that had been demonstrated twenty-four years ago in the nuclear test of 1974" (Singh 1998, 44). In fact, public opinion polls taken after the original 1974 test demonstrated the political popularity of nuclear tests in India; 90 percent of respondents said they were personally proud of India's achievement (Sagan 1996, 68).

Changes in India's nuclear policies can also be linked to transitions in political leadership. Following India's first test under Prime Minister Indira Gandhi, the nuclear program was put on hold in 1977 when the Janata Party came to power. But Gandhi revived the program when she returned to power in 1980. This research laid the groundwork for her son, Rajiv Gandhi, to authorize development of weapons-specific technologies in the late 1980s (Nuclear Threat Initiative 2010c).

In the 1990s, the Bharativa Janata Party (BJP) (a successor to the previously mentioned Janata Party) became the national advocate for testing a weapon and formally declaring India's nuclear status. Prime Minister Vajpayee planned to conduct a test in 1996 but his party lost its seats before executing a test. In the 1998 election, Vajpayee and the BJP successfully ran on a platform that pledged to "reevaluate" India's nuclear policy and to "exercise the option to induct nuclear weapons" (Ahmed 1999, 193). While the extent to which this issue played a role in the election is debatable, a nuclear test was clearly part of BJP's domestic political strategy and, once in power, it carried out the campaign promise.

Pakistan's decision to test a nuclear weapon can also be traced to bureaucratic politics, in which its military was the primary driver of a more assertive nuclear policy. India's test sparked a debate within Prime Minister Nawaz Sharif's cabinet, with Sharif himself initially taking a more restrained approach (Ahmed 1999, 194). The military, however, argued forcefully for Pakistan to respond with its own nuclear tests; while the cabinet may have been divided, the internal balance of power in Pakistan strongly favored the military. The military and civil bureaucracy, "with the acquiescence of the political leadership," thus decided to proceed with the tests (Ahmed

1999, 179). While the military's central role in this decision suggests that security concerns played an important role, national prestige was clearly at play as well. Ahmed writes, "for Pakistani policymakers, particularly the military, a nuclear stature less than India was unacceptable" (Ahmed 1999, 195).

Public opinion also plays an important role in the decision to conduct tests. Even before India's test, a 1996 Gallup poll found that 80 percent of Pakistanis would support a nuclear test if India were to conduct a test first (Koch 1996, 4). In short, the inability of India and Pakistan to maintain a status quo of mutual nuclear latency was caused by both a security environment that could not prevent escalation and internal political dynamics that pushed the countries toward conducting nuclear tests.

VI. Assessing Iranian Ambitions

An analysis of the external security environment and domestic political factors helps to explain why Japan has maintained nuclear latency despite growing regional threats and why India and Pakistan felt compelled to end their respective periods of latency and become full-fledged nuclear powers. Applying this framework to Iran may therefore be useful in evaluating whether Iran is likely to test a weapon or content itself with a latent nuclear capability. While drawing analogies between Iran and the countries detailed above will not produce any definitive conclusions about its nuclear ambitions, exploring Iran's security environment and internal pressures provides a framework that will more clearly elucidate the factors at play.

The Security Environment: Comparing the Fates of Iraq and North Korea

Iran faces possible security threats from its Arab neighbors, Israel, and the United States, each of which could push Iran toward acquiring a nuclear weapon. Locally, Iran seeks to extend its influence over the Arab Gulf states. Iran scholar Ray Takeyh writes that, "From the Islamic Republic's perspective, the Gulf is its most important strategic arena, constituting its most reliable access to the international petroleum market" (Takeyh 2004, 53). Moreover, Iran's eight-year war with Iraq cost tens of thousands of lives and remains firmly imprinted in the memories of Iran's leaders today. President Ahmedinejad himself served as a soldier in the Revolutionary Guard Corps (IRGC). Further, with Saddam Hussein removed from power, Iran possesses the most powerful conventional military in its immediate region. Under the logic of prudential realism, deploying nuclear weapons would actually damage Iran's regional security by sparking proliferation

amongst its rivals and thus negating its conventional military advantage.

Israel, however, is widely understood to have a nuclear weapon (Cohen 2010; Nuclear Threat Initiative 2010a). Thus, an Iranian bomb could be a deterrent against a potential Israeli attack. But despite a history of hostile rhetoric and proxy battles via the Lebanese Islamist group Hezbollah, there has been no direct conflict between Israel and Iran. Iran did not participate in any of the Arab wars against Israel, and Israel has not shown any inclination to retaliate against Iran for Hezbollah's violent actions. Indeed, it is Iran's nuclear program itself that has raised the possibility of an Israeli attack. While Israel should rightly be concerned about the possibility of an Iranian nuclear weapon, it is hard to accept an Israeli threat as rationale for Iran acquiring the bomb. As former weapons inspector David Kay puts it, "Iran does not worry that Israel can organize and build a regional coalition that will limit the power of Iran or might even topple the regime. Only one state has that power in the eyes of Tehran—the United States" (Kay 2008, 13-14).

Indeed, the United States and Iran have a history of tensions—from the coup in 1953 to the hostage crisis of 1979—that might cause the Iranian regime to see the United States as an existential threat. Meanwhile, the overthrow of Saddam Hussein demonstrated American willingness to use force to pursue its interests in the region. As one Iranian official stated, "The fact that Saddam was toppled in twenty-one days is something that should concern all the countries in the region" (Reuters 2003). The appeal of a nuclear deterrent may be even stronger for Iran as it compares the respective fates of Iraq and North Korea. The United States targeted Iraq for forcible regime change, believing Iraq to be developing nuclear weapons but not yet capable of detonating a weapon. Conversely, the United States has not pursued coercive action against North Korea, which demonstrated its nuclear capability in 2006, and has repeatedly offered North Korea economic incentives in exchange for more cooperative policies. Takeyh argues that the Iranian regime may interpret U.S. behavior as suggesting that a developing or latent nuclear capability could put Iran in jeopardy, while a proven ability to detonate a weapon would be Iran's only sure protection against an American attack (Takeyh 2004).

Domestic Political Dynamics: Islam and the Bomb

Iran's domestic politics are characterized by factional disputes between the clerics, revolutionary guard, so-called "pragmatic" conservatives, and reformists. However, all groups seem to agree that Iran must develop a high level of nuclear capability. For example, even Mir-Hossein Moussavi,

the primary challenger to President Mahmud Ahmedinejad in the 2009 elections, declared that "No one in Iran would accept suspension" of the country's uranium enrichment program (Financial Times 2009). In 2005, Supreme Leader Ali Khamenei issued a *fatwa* stipulating that no Islamic state may possess or use atomic weapons. In February 2010, Khamenei reiterated that Islam is "opposed to nuclear weapons" (Yeranian 2010). But he has simultaneously been a vocal proponent of developing Iranian nuclear capacity, even for security purposes. The conservative newspaper *Jumhuriye-Islamii*, largely considered a mouthpiece for the Supreme Leader by scholars of Iranian politics, wrote "in the contemporary world, it is obvious that having access to advanced weapons shall cause deterrence and, therefore, security" (Takeyh 2004, 56).

Pragmatic conservatives and reformists, interested in opening Iran for greater economic growth, may also be wary of the diplomatic and economic isolation that a nuclear test would produce. Many have hoped that if these leaders were to come to power, Iran might be more willing to cut a deal with the West and, in essence, to exchange its nuclear program for economic benefits. However, Iran's 2009 political campaign should cast some doubt on such optimistic assumptions. Even the reformist candidate Mir Hossein Moussavi struck a defiant stance on the Iranian nuclear program, stating that Iran would never halt its enrichment of uranium (PressTV 2009).

Finally, the religious edicts of the Supreme Leader and the liberalizing instincts of the pragmatists and reformists may be irrelevant if, as Iran analyst Gary Sick suggests, the IRGC has essentially carried out a "soft coup" and assumed exclusive control of Iran's national security decision making (Sick 2009). The IRGC is less likely to be influenced by religious dogma or economic incentives and may view a successful nuclear test as the only way for Iran to establish a credible deterrent against the United States and assume its rightful place in the elite group of world nuclear powers. Depending on which faction emerges victorious out of the past year's turmoil, Iran's internal political dynamics could either restrain it from moving beyond latency or propel it toward a nuclear test.

VII. Policy Recommendations

As Iran progresses toward developing a nuclear weapon capability, the international community has a compelling interest in preventing it from testing a nuclear device. Although most states would prefer Iran refrain from developing even a latent capability, Iran's interests and behavior indicate that this is highly unlikely. Iran's consistent defiance of UN Security

Council and IAEA resolutions, and its slow but steady uranium enrichment (even with the most recent "Stuxnet" setbacks), both indicate that it is only a matter of time before Iran acquires the technological capability to build a nuclear weapon.

However, it is not inevitable that Iran will test a nuclear device upon acquiring the capability to do so. Given this situation, the international community should dedicate its energies toward encouraging Iranian nuclear latency and preventing nuclear testing. This scenario greatly reduces the risks that other states in the region would feel compelled to test weapons or that dangerous non-state groups would try to acquire a nuclear weapon. The theories and cases examined in this article suggest the following steps for the international community to increase the likelihood of Iran pursuing only a latent nuclear capability:

End Talk of Preventive Military Strikes on Iran: The crucial disadvantage of the hedging strategy for a potential nuclear power is the lack of a credible nuclear deterrent. As long as Iran does not have a deployed nuclear weapon that it can use to threaten counterattack, Iran will feel vulnerable to military strikes by Israel and/or the United States. This threat would likely be the primary impetus for Iran to develop an overt nuclear capability. In fact, Iran's current nuclear development efforts could be seen as a race to develop a nuclear capability before Israel or the United States decides to attack. Israel and the United States should seek to reverse this logic by downplaying the possibility of a preventive strike as long as Iran refrains from testing or deploying a weapon. If Iran were to conduct a nuclear test, all bets are off, and such an action may invite a forceful response. By clarifying this "red line," the United States and Israel will alter the incentives facing Iran in a way that will encourage Iran to pursue a latent rather than overt nuclear capability.

Support from Russia and China: The U.S. nuclear umbrella has been a key factor in dissuading Japan from advancing beyond nuclear latency. It is unlikely and undesirable for any state to similarly accept a nuclear-backed security guarantee from Russia and China. However, it may be in the international community's interest for China and Russia to continue playing the role of Iran's supporters. If Iranian leaders believe that such great powers as China and Russia will help protect Iran's security, they may forego deploying their own nuclear deterrent. The message from Russia and China to Iran must be crystal clear in order for such a strategy to work. They must convey that while they are willing to support Iran's nuclear development and its conventional military forces, all assistance would end if Iran were to test a nuclear device.

Nuclear Technology in the Arabian Peninsula: The international community must increase the perceived costs to Iran of testing a nuclear weapon. One way of doing this is to make a nuclear test a "red line" that might invite a forceful response as described above. Another strategy is to continue to provide limited materials and assistance to Iran's rivals in the Arabian Peninsula. Avoiding proliferation among one's rivals is a key reason that states pursue a policy of nuclear hedging. Iran may be more likely to adopt such a policy if it fears proliferation amongst its Arab rivals. Of course, one of Iran's regional rivals, Israel, is already presumed to have a nuclear capability. However, it has to date been unwilling to officially acknowledge its arsenal and, more importantly, unwilling to threaten its use in regional conflicts. Iran should fear that conducting an overt nuclear test might spark a change in Israel's nuclear policy. An Israel that is willing to threaten nuclear retaliation against Iran for attacks launched by Hezbollah would put Iran in a very dangerous situation.

Avoid Intervention in Iranian Domestic Politics: Domestic politics also play an integral role in states' nuclear policies. The international community can do little in this regard; efforts by the international community to intervene in Iranian domestic politics are likely to only exacerbate the situation. Intervention may increase the chances that Iran will test a nuclear weapon in an effort to incite nationalist sentiment in support of the ruling government and to bolster Iranian prestige in the international community. The United States may wish to continue pursuing public diplomacy efforts to help mitigate its image as a threat to the country. However, it should strive not to be seen as taking sides between political factions or supporting efforts to overthrow the regime. In fact, even many dissident groups themselves have argued that support from the United States is counterproductive and undermines their mission (O'Rourke 2007).

VIII. Conclusions

As technologies will inevitably continue to spread, an increasing number of states will possess at least the scientific capability to produce a nuclear weapon. Thus, the main barrier to acquiring the bomb will no longer be technology, but political choice.

It is impossible to predict with any degree of certainty whether and when Iran might detonate a device and declare itself a nuclear state. However, the cases of Japan and South Asia shed some light onto the factors that influence states' nuclear policy choices. The possibility of future U.S. intervention provides the most compelling motive for Iran to develop a demonstrable nuclear deterrent. The contrasting examples set by U.S.

policy toward North Korea and Iraq make it likely that Iran will continue to pursue such a deterrent presence unless it can be convinced that weapons tests will lead to more dire security consequences than abstinence.

The world is likely to face many more cases similar to Iran in the future. Minimizing the number of states that test and deploy nuclear weapons will require the international community to address both the security and domestic political factors that influence whether latent states choose to become nuclear states. Global powers must provide incentives on both fronts, from expanding nuclear umbrellas to strengthening norms against proliferation, if they hope to convince the expanding ranks of nuclear capable states not to build the bomb.

References

Ahmed, Samina. 1999. Pakistan's Nuclear Weapons Program: Turning Points and Nuclear Choices. *International Security* 23, no. 4 (Spring).

Campbell, Kurt M., and Tsuyoshi Sunohara. 2004. Japan: Thinking the Unthinkable. In *The Nuclear Tipping Point*. Washington, DC: Brookings Institution Press.

Cohen, Avner. 2010. *The Worst-Kept Secret: Israel's Bargain with the Bomb*. Columbia University Press.

Cole, Juan. 2009. Does Iran really want the bomb? *Salon.com*. August 7. http://www.salon.com/news/opinion/feature/2009/10/07/iran_nuclear/index.html (accessed July 6, 2011).

Dilanian, Ken. 2011. Iran's Nuclear Program and a New Era of Cyber War. *The Los Angeles Times*, January 17.

Financial Times Interview: Mir-Hossein Moussavi. *Financial Times*, April 13. http://www.ft.com/cms/s/0/a2466224-2824-11de-8dbf-00144feabdc0.html#axzz1F6CKu3oK (accessed July 6, 2011).

Hughes, Llewelyn. 2007. Why Japan Will Not Go Nuclear (Yet). *International Security* 31, no. 4.

Hymans, Jacques E. 2010. When Does a State Become a "Nuclear Weapons State"? *The Nonproliferation Review* 17, no. 1: 161-180.

International Panel on Fissile Materials. 2009. *2009 Global Fissile Material Report: Nuclear Weapon and Fissile Material Stockpiles and Production*. http://www.fissilematerials.org/ipfm/pages_us_en/inventories/inventories/inventories.php (accessed July 6, 2011).

Jane's Information Group. 2011. *Strategic Weapon System, China*. Jane's Sentinel Security Assessment. China and Northeast Asia. January 17.

Kay, David. 2008. Nuclear Fallout. *The National Interest*.

Koch, Andrew. 1996. Nuclear Testing in South Asia and the CTBT. *The Nonproliferation Review*: 98 - 104.

Nuclear Threat Initiative. 2009. Japan Profile. *NTI: Research Library: Country Profiles: Japan*. October. http://www.nti.org/e_research/profiles/Japan/Nuclear/index.html (accessed August 31, 2010)..

———. 2010a. Israel Profile. *NTI Research Library: Country Profiles: Israel*. January. http://www.nti.org/e_research/profiles/Israel/index.html (accessed April 15, 2011).

———. 2010b. South Africa Profile. *NTI: Research Library: Country Profiles: South Africa*. January. http://www.nti.org/e_research/profiles/SAfrica/index.html (accessed August 31, 2010).

———. 2010c. India Profile. *NTI: Country Overviews: India: Profile*. February. http://www.nti.org/e_research/profiles/India/index.html (accessed August 31, 2010).

———. 2010d. Pakistan Profile. *NTI: Country Overviews: Pakistan: Introduction*. February. http://www.nti.org/e_research/profiles/Pakistan/index.html (accessed August 31, 2010).

O'Rourke, Breffni. 2007. Iran: Dissidents Debate Merits of U.S. Democracy Aid. *Radio Free Europe*, November 2. http://www.rferl.org/content/article/1079070.html (accessed July 6, 2011).

Paul, T.V. 2000. *Power Versus Prudence: Why Nations Forgo Nuclear Weapons*. Quebec City: McGill-Queen's University Press.

Perkovich, George. 2002. *India's Nuclear Bomb: the Impact on Global Proliferation*. University of California Press.

PressTV. 2009. Mousavi: Iran will never halt enrichment. *Pavyand News*. April 14. http://payvand.com/news/09/apr/1156.html (accessed July 6, 2011).

Reuters. 2003. Iran's New Anxieties. *Reuters*, April 19.

Sagan, Scott. 1996. Why Do States Build Nuclear Weapons? Three Models in Search of a Bomb. *International Security* 21, no. 3: 54-86.

Sanger, David E., and Thom Shanker. 2010. Gates Says U.S. Lacks a Policy to Thwart Iran. *The New York Times*, April 17.

Shoichi, Koseki. 1998. *The Birth of Japan's Postwar Constitution*. Boulder, CO: Westview Press.

Sick, Gary. 2009. Iran's Political Coup. *Gary's Choices*. June 13. http://garysick.tumblr.com/post/123070238/irans-political-coup (accessed July 6, 2011).

Singh, Jaswant. 1998. Against Nuclear Apartheid. *Foreign Affairs* 77, no. 5 (October).

Takeyh, Ray. 2004. Iran Builds the Bomb. *Survival* 46, no. 4 (Winter -2005).

Wohlstetter, Albert, Gregory Jones, and Roberta Wohlstetter. 1979. *Towards a New Consensus on Nuclear Technology*. Los Angeles: Pan Heuristics.

Yeranian, Edward. 2010. Iran's Supreme Leader Khamenei Says Islam Opposes Nuclear Weapons. *Voice of America News*, February 19. http://www1.voanews.com/english/news/middle-east/Supreme-Leader-Khamenei-Says-Islam-Opposes-Nuclear-Weapons-84771247.html (accessed July 6, 2011).

3

Balancing Financial Terror: The Game Theoretic Dynamics of Massive Debt

Gregory Hudson

The "balance of financial terror," first coined by Lawrence Summers in 2004 in the context of U.S.-China relations, characterizes the situation when one country owns such a significant portion of another country's national debt that it is unclear which country more credibly threatens the other. Debtors threaten to devalue their currency while creditors threaten to dump the debt, putting both countries in a precarious balance. Using game theory, this paper finds that the balance of financial terror perpetuates because neither the creditor nor the debtor country can unilaterally change the status quo without massive retaliation. Under a sequential move scenario, the balance is likely to perpetuate until some external factor raises the payoffs to unilateral change for one country. Alternatively, if the relationship were to transform from a sequential into a simultaneous game, the iterated elimination of dominated strategies suggests that the balance could collapse as creditors and debtors rush to preempt the worst-case scenario moves of the other side. For the case of the United States and China, factors that could upset the balance include quantitative easing by the Federal Reserve, which could reduce the payoff for holding U.S. debt.

Gregory Hudson is a recent graduate of the Master's program in International History and Politics at the Graduate Institute of International and Development Studies in Geneva, Switzerland. He can be reached at gregory.hudson@graduateinstitute.ch.

I. Introduction

John Maynard Keynes once said, "If you owe your bank manager a thousand pounds, you are at his mercy; if you owe him a million pounds, he is at your mercy" (Branegan, Taylor and Ungeheuer 1982). The point illustrated here is that a debtor-creditor relationship is far from a one-way street; rather, at some point, the balance of power changes from the creditor to the debtor. Unfortunately, the academic literature on this issue is lacking, and it is not always clear precisely when this balance changes. After all, if you owe the bank just half a million pounds ~~or~~ the equivalent in units of any valuable currency—which of you is at the mercy of the other?

Today, the international economy is grappling with exactly this problem. First described as the "balance of financial terror" by Lawrence Summers in 2004 (Davis 2010), the United States and China are locked in a debtor-creditor relationship where it is unclear just which government holds more power over the other. Since the Asian financial crisis in 1997-1998, China has sought to expand its foreign exchange reserves with investments in "safe" assets, such as U.S. Treasury Bonds. Meanwhile, since 2001, deficit spending has induced the United States to seek creditors from abroad. The result has been a marriage of convenience in which China receives safe assets and in return the United States receives low interest rates. Indeed, in the decade since the turn of the millennium the United States has deeply indebted itself to China. In October 2010, China held almost $1.2 trillion out of $13.6 trillion in U.S. public debt, or 9 percent of the debt (U.S. Treasury 2010). As a result, China is the single biggest holder of U.S. debt, and the most important creditor relationship for the United States. Similarly, these debt holdings account for roughly 33 percent of China's $2.65 trillion in foreign exchange reserves (Dean, Browne and Oster 2010). As a consequence, both sides exert substantial leverage over the other: the Americans do not want the Chinese dumping their debt holdings, which would send interest rates in the United States soaring, nor do the Chinese want the Americans to devalue their currency and thereby escape paying back the debt in full real terms. Both sides have reason to worry about provoking the other into massive financial retaliation, to say nothing of other political forms of retaliation, which could further compound the situation but are not discussed in this article. The financial retaliatory actions, labeled above as "dump" and "devalue," will receive more detailed explanation in Section III.

For policy makers, game theory can shed light on the strategic essence of this financial situation. This article posits a model of the balance of

financial terror based on the current U.S.-China dynamic, but one that could be theoretically applicable to any case in which countries seek to escape each other's terror and in which one country is highly indebted to another with debt denominated in the debtor's home currency. This article will demonstrate that self-denominated debt is a necessary condition for a true balance of terror. While today the United States is the debtor and China is the creditor, in the future, other countries could eventually substitute for these roles. Hence, the model presented here will prove useful in two dimensions: first, as a specific tool for the actual U.S. and Chinese policy makers who find themselves locked in the current debt situation, and second, as a theoretical tool for future policy makers who may find themselves in an analogous situation.

Specifically, this article's model suggests that debtors and creditors in this situation are likely to perpetuate the game until either the player sequence or payoffs change. In this model, countries have three options in a continuum of escalation. Nine distinct possible outcomes exist as a result, with countries (or players) ranking the outcomes differently. However, the logic of the balance of financial terror makes only a few of these outcomes plausible. Specifically, the outcomes highly depend on the sequence of moves; modeling the game simultaneously results in a disastrous outcome for both debtor and creditor countries, while modeling it sequentially with the debtor moving first retains the status quo.

Further, the balance of financial terror is highly sensitive to exogenous changes to the payoffs. By tweaking payoffs of the medium option to incorporate incidental effects, this model finds that both countries will undertake a medium outcome of escalation. Therefore, the model's power arises from its ability to predict big changes in players' behavior—and the game's outcome—from small exogenous changes to the players' circumstances. The model illuminates why the balance of financial terror has persisted as long as it has, and suggests how the balance could eventually be upset.

In the final analysis, this article argues that the most appropriate way to model the debtor-creditor relationship is to have the debtor move first. With the debtor moving first, and no exogenous considerations incorporated to the debtor's payoffs, the debtor is most likely to maintain the value of its currency so as to forestall the creditor from dumping the debtor's debt. Such a dynamic characterizes the U.S.-China debt relationship thus far, and this article will turn repeatedly to that example for illustration. Further, this article compares three different extensive form versions of the balance of financial terror game, each corresponding to a different player

sequence. By comparing them, this model illustrates why a creditor (or China, for example) would and should prefer to move after the debtor.

Ultimately, this article aims to use game theory to deepen our insight into why balances of financial terror sustain themselves, and how such balances could be upset. From a policy perspective, the model presented here should be useful to policy makers on both sides of the U.S.-China divide, who currently find themselves locked in this game. As will be shown below, the balance of financial terror only perpetuates under very specific assumptions. External changes to these assumptions, such as reversing the game sequence or heightening the payoffs of escalatory moves, can result in disaster.

Before turning to these elaborations, this article will first explain the underlying assumptions behind a basic game of the balance of financial terror—namely player assumptions, move assumptions, and outcome assumptions. The article justifies each below.

II. Players

In a basic version of the balance of financial terror, there are two players: the Debtor and the Creditor (hereafter always capitalized). For both the matrix and extensive form games that will appear later in this article, the Debtor will always appear in light gray and the Creditor in dark gray. To keep the players straight, the reader may find it helpful to think explicitly of the principal examples of the Debtor and Creditor from which this game drew its model: the United States for the Debtor in light gray and China for the Creditor in dark gray.

Also important to keep in mind is that like most applications of game theory, the player assumptions in this model simplify a more complicated reality. For example, naming the players as the U.S. Federal Reserve and the People's Bank of China can concretize the analysis, since monetary authorities primarily control the levels of devaluation and debt purchases in a balance of financial terror. The Federal Reserve indirectly controls the U.S. exchange rate through monetary policy, while the People's Bank of China buys and sells U.S. Treasury Bonds through their State Administration of Foreign Exchange. However, other actors from each government, such as the U.S. Treasury Department and China's sovereign wealth fund, the China Investment Corporation, also influence exchange and debt holding policy, even while the central banks remain the primary players. Hence, with this caveat in mind, for the purposes of this model and its focus on the bigger picture, we can simply streamline these government entities as the United States and China, in keeping with the common international

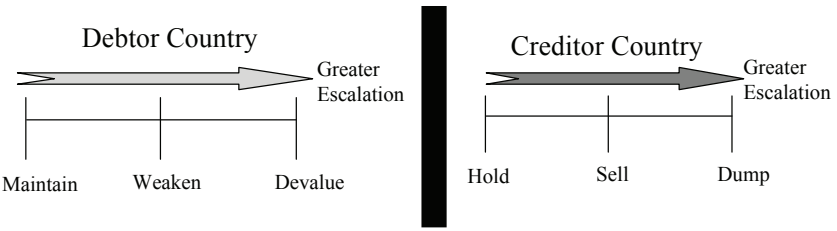
relations practice of analyzing states as whole units.

This model makes assumptions about who is playing the game in order to keep the focus on a particular dynamic in the debt relationship. For example, this model focuses on only one holder of U.S. debt, China, when in reality there are many holders of U.S. debt. Japan, for instance, is the second greatest single foreign holder of U.S. debt, at 6.4 percent of the total holdings (U.S. Treasury 2010). After Japan, the size of foreign holdings rapidly diminishes, with the United Kingdom as the third single greatest holder at 1.5 percent, Brazil at 1.3 percent, and so forth (U.S. Treasury 2010). One could add player assumptions to this model to simulate the Japanese version of the game as well, and continue to do so for an increasing number of players until a model covers all the holders of U.S. debt. However, crowding the model with players would so complicate the model as to render it useless for policy makers, and would distract from the core insights that simpler models can generate. This model attempts to strike a balance between completeness and clarity.

III. Moves

To reflect the mutual fears of the balance of financial terror, the model in Figure 1 assigns the Debtor and the Creditor countries a continuum of escalatory moves to scare the other player into a preferred action. Basically, both players can retain their status quo position ("Maintain" currency strength for the Debtor, "Hold" debt for the Creditor), escalate some ("Weaken" currency for the Debtor, "Sell" debt for the Creditor), or escalate maximally along the continuum ("Devalue" currency for the Debtor, "Dump" debt for the Creditor). As the moves escalate, the outcomes not only become more costly for the intended target, but also for the move-taker.

Figure 1: Move Continuum



Debtor Moves Described

Maintain: The Debtor country keeps the value of its currency strong, or at least maintains the currency at a level where the Creditor will not lose real value in its debt holdings. The game presumes at the outset that the Creditor loaned to the Debtor when the Debtor's currency was relatively strong, and therefore the Creditor potentially has something substantial to lose if the Debtor's currency lost its value.

Weaken: The Debtor country gradually weakens the value of its currency, but does not weaken it so sharply as to collapse the currency. Hence, "Weaken" is a move of intermediate escalation, but not an attempt by the Debtor to escape paying its debt outright.

Devalue: The Debtor sharply decreases the value of its currency, to the point where it owes very little to the Creditor in real terms. The Debtor can devalue through either formal or informal means: Debtors with fixed exchange rates can formally announce new exchange rates, while Debtors with floating exchange rates, such as the United States, can use highly inflationary monetary policy to sharply reduce their home currency's purchasing power. Regardless of the form it takes, devaluation is the greatest escalation the Debtor can take; it essentially means the vast majority of the debt will not be repaid to the Creditor in real terms.

Creditor Moves Described

Hold: The Creditor keeps its investments in the Debtor's debt to retain them as assets. So long as the Creditor holds the Debtor's debt, the debt may be subject to the Debtor's attempts to weaken or devalue the currency.

Sell: The Creditor gradually or partially sells some of its debt holdings in order to cash them out for reinvestment in another asset. The Creditor does not cash the debt out entirely, making this move an intermediate escalation or a hedge between holding and dumping the debt.

Dump: The Creditor precipitously liquidates all of its debt holdings with the Debtor. Essentially, "Dump" is a vote of no-confidence in the Debtor's ability to repay, and so the Creditor seeks to withdraw its money as quickly as possible before it becomes worthless. Thus, this move is an extreme escalation and a decision by the Creditor to accept large losses now to avoid even more disastrous ones in the future.

Justification of Moves Selected for Modeling

This model attempts to capture the essence of the balance of terror dynamic. Yet, like all games, these moves are simplifications of the real world. In reality, the options available to both players are much more varied. For

example, one could immensely complicate the game by adding a large number of options between the theoretical endpoints of escalation, but the insights gleaned from such a complication would not be terribly different (and would come at the expense of cluttering the game's matrix and extensive forms).

Further, this model forgoes options to the left of the continuum above—i.e., those moves that would actually deepen the Debtor-Creditor relationship, such as "Buy[ing]" more debt for the Creditor, or "Strengthen[ing]" the currency for the Debtor. These moves seem bizarre or irrelevant in a true "balance of financial terror," as they run counter to an underlying assumption of the situation: both parties want to be less terrified of the other. For example, if the Creditor country bought even more of the Debtor's debt, the Creditor would just worsen its trapped predicament, and it would be further at the Debtor's mercy. As the opening quote to this article illustrated, the whole point of the balance of financial terror is that buying more debt no longer gives the Creditor leverage, but ironically worsens the Creditor's position. Hence, a Creditor should not wish to fulfill Keynes' maxim and tip the scales to the Debtor. Similarly, if the Debtor intentionally strengthened the value of its currency, instead of maintained, weakened, or devalued it, the Debtor would merely owe more to the Creditor in real terms. As a result, strengthening would defeat the Debtor's presumed desire to lighten its debt load.

Nevertheless, it is important to keep in mind that this model's selection of moves only applies to countries in very specific circumstances, like those of the United States and China today. In the U.S.-China relationship, the Debtor denominates its debt in its own currency, an option not available to many countries of the world. Some Debtor countries with foreign-denominated debts may feel forced to play "Strengthen" just to keep their exchange rate strong, and thus their real debt burdens low. Thailand, for example, strengthened its currency at the outset of the Asian financial crisis for this reason. However, this model does not attempt to explain that situation, which has fundamentally different assumptions and is not considered "balanced" financial terror. In this article's model, the Debtor currency's reserve status is exactly one source of the Debtor's leverage in a balance of terror. While one might hypothesize that even a Debtor with a reserve currency could attempt to strengthen its currency to bind itself up even closer with the Creditor, this move seems extremely risky to the point of stretching plausibility. Any strengthening of the currency by the Debtor would merely add value to the assets of the Creditor, and would not put the Creditor's previous gains at risk.

As a result, one can safely subsume the consequences of these strengthening moves into the left hand side of the continuum above. Effectively, in a balance of financial terror, where the point is gaining leverage over the other party, buying and holding the Debtor's debt has the same effect of adding no new pressure on the other party; similarly, strengthening and maintaining the Debtor's currency also adds no new pressure. Rather, they can both be subsumed as moves to preserve the status quo.

Finally, there is one other move that does not arise in this model: "Default" by the Debtor. In a default, the Debtor simply cannot pay back the Creditor and the Creditor must accept the loss. However, as stated earlier, since this model presumes the Debtor denominates its debt in its own currency—and therefore can print any sum to meet its nominal debt burden—default would not plausibly arise. Rather, devaluing the currency would give the Debtor the same escape from its burden, but avoid the technicality of a Default.

With the two players' three moves so defined, this article now forms a game and assigns payoffs to the nine possible outcomes, and justifies those assignments below.

IV. Preferred Outcomes and Matrix Form of the Game

Given the moves described in the game above, Figure 2 outlines a three-by-three matrix form of the game with the following payoffs. All payoffs are ordinal, or ranked by preference, which is sufficient to solve a game with pure strategies. Here, "0" is the least preferred payoff and "8" is the most preferred for any given player. Player 1 is the Debtor, with moves labeled on the left hand column, and Player 2 is the Creditor, with moves labeled on the top row.

Figure 2: Matrix form of the Balance of Financial Terror

	Hold	Sell	Dump
Maintain	(6,8)	(3,7)	(0,4)
Weaken	(7,5)	(4,6)	(1,3)
Devalue	(8,0)	(5,1)	(2,2)

This model uses ordinal payoffs in place of cardinal payoffs because of the difficulty of assigning precise intervals between outcomes. In other words, because it is speculative to specify to what *degree* a Creditor or

Debtor would prefer one outcome to another, ordinal preferences are a safer assumption. The style (Player 1's Move, Player 2's Move) notes each outcome that corresponds to a rectangle in the above matrix. Within a given rectangle in the matrix, the style (Player 1's Payoff, Player 2's Payoff) lists each payoff. So, for example, the outcome "(Maintain, Hold)" corresponds to the rectangle in the second row, second column of the matrix, where the Debtor maintains the value of its currency and the Creditor continues to hold the Debtor's debt. This outcome yields a payoff of six for the Debtor and eight for the Creditor.

Creditors and Debtors rank their outcomes differently. As a result, in a balance of financial terror Creditors and Debtors have divergent but at times overlapping interests. Explanations for the outcome rankings follow below, beginning with the Debtor.

Ranking the Debtor's Outcomes

This game ranks the Debtor's outcomes as follows:
Debtor's Preferred Outcomes: (Devalue, Hold) > (Weaken, Hold) > (Maintain, Hold) > (Devalue, Sell) > (Weaken, Sell) > (Maintain, Sell) > (Devalue, Dump) > (Weaken, Dump) > (Maintain, Dump)

In general, this game asserts that the Debtor would ideally like to escape from its debt burden through currency devaluation. So long as the Creditor would keep holding the debt nominally, and the Debtor could avoid paying back most of it in real terms, such an outcome would be a bargain for the Debtor. Indeed, in general the Debtor always prefers the Creditor to hold its debt rather than sell or dump it; so long as the Creditor holds the debt, interest rates will not go up. So, if the Debtor cannot devalue its currency while the Creditor holds its debt, then the Debtor would rather weaken its currency while the Creditor holds its debt, or maintain its currency while the Creditor holds its debt. Hence the first three preferred outcomes above.

The second best set of outcomes for the Debtor is that the Creditor merely gradually sells its debt holdings rather than dumps them outright. Gradual selling would put upward pressure on interest rates, but not intense pressure. In response, again, the Debtor would most prefer devaluing altogether to escape its debt payment. If not able to devalue, it would rather weaken to escape some repayment. If not able to weaken, it would rather maintain the strength of its currency.

The worst outcomes for the Debtor involve those where the Creditor dumps the debt. A dump would send interest rates soaring. But if already given a debt dump, the Debtor would most prefer to devalue whatever

is left, and then prefer to weaken whatever is left. Least preferred of all would be to try to maintain its currency at a strong level during a debt dump; this outcome would lead to the highest real repayment costs for the Debtor.

Ranking the Creditor's Outcomes

This game ranks the Creditor's outcomes as follows:
Creditor's Preferred Outcomes: (Maintain, Hold) > (Maintain, Sell) > (Weaken, Sell) > (Weaken, Hold) > (Maintain, Dump) > (Weaken, Dump) > (Devalue, Dump) > (Devalue, Sell) > (Devalue, Hold)

Generally in a balance of financial terror, the Creditor wants to maintain the value of its investments in the Debtor's debt. First, and ideally, the Creditor would like to see the Debtor maintain the value of its currency, so the Creditor can continue holding the debt with peace of mind that it will be repaid in full. Secondly, the Creditor would prefer the Debtor to maintain its currency and the Creditor sell part of the debt, to collect the debt at a reasonably high value.

After these options, the rationale for the Creditor's rankings becomes more complicated. The Creditor holds a preponderant proportion of the Debtor's debt, and any move by it to massively sell or dump the Debtor's debt will inherently decrease the value of its remaining holdings. In other words, as long as the Debtor cooperates by keeping its currency strong, the Creditor cannot credibly threaten to dump the holdings; to do so would mean lowering the value of its own remaining holdings. This is the logic of "China's Dollar Trap" as outlined by Paul Krugman, and it is the essence of the U.S.-China debt dynamic that remains in place almost two years later (Krugman 2009). The extent to which China is really "trapped" also determines the extent to which the financial terror is really "balanced," and so accurately modeling this point is crucial for policy.

If this trap is real, the self-inflicted pain from a "Dump" usually outweighs the gain the Creditor would make by cashing out early. Hence, in the matrix above, it is this disproportionate, self-inflicted pain that makes the Creditor prefer gradually selling or even holding a weakened currency rather than dumping the debt. In other words, the Creditor prefers outcomes "(Weaken, Sell)" and "(Weaken, Hold)" over any outcome in which the Creditor must dump its holdings and sustain a loss.

As a result, the only time "Dump" becomes attractive to the Creditor is when the Creditor knows the Debtor will devalue. In that case, the Creditor is better off cutting these mammoth losses short, because such will be the only time when the losses from the Debtor's move would out-

weigh the self-inflicted pain from a dump. Similarly, the Creditor would least prefer to get caught holding devalued currency, or almost as bad, to only partially sell the devalued currency when it could sell more. Finally, rounding out the middle of the Creditor's preference list, a Creditor would rather dump a highly maintained currency or a merely partially weakened currency to extract marginally more of its value, even if in practice these outcomes would never happen.

To understand why some of these outcomes would never happen—and indeed, to solve the game—this article analyzes the game in its extensive form below.

V. Extensive Forms of the Balance of financial terror: Player Sequences

With the players' outcomes ranked, the next question in modeling the balance of financial terror is the sequence of the players' moves. As we will see, the selection of which player moves first enormously impacts the game's outcome. While hypothetically either player could move first, this article asserts that it is more plausible that the Debtor rather than the Creditor will move first. Moreover, from a policy perspective, this article argues it is in both players' *interest* that the Debtor move first.

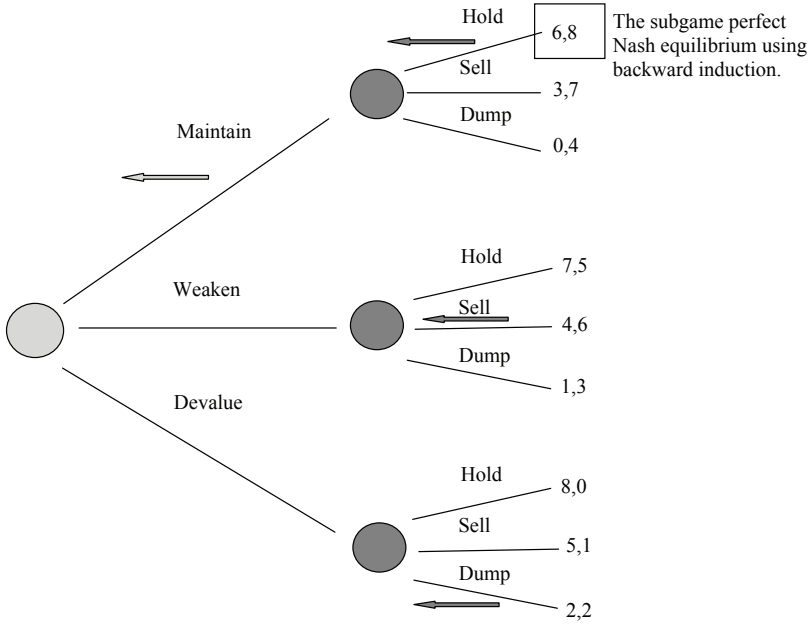
The Debtor moves first for one reason above all: if the Creditor moved first, it would always face a certain devaluation of the debt in response from the Debtor. Further, since the outcomes associated with devaluation compose the Creditor's bottom three preferences, it will always be in the Creditor's interest to wait and so avoid this response.

However, depending on the sequencing of the players, escalation is not necessarily symmetrical. While a Creditor would face devaluation by the Debtor in response to a Creditor's first move, a Debtor would *not face* dumping by the Creditor if the sequence of moves were reversed. In other words, the Debtor, if it moves first, can avoid an escalatory response, while a Creditor cannot. Consequently, the Creditor has every incentive to wait. Further, by moving first, a Debtor can set the game on an equilibrium path that is more preferred by both parties compared to the extremely escalatory outcome of "(Devaluation, Dump)." *Specifically, by moving first the Debtor can maintain its currency, resulting in the outcome "(Maintain, Hold)," which is worth more than the outcome "(Dump, Devalue)."* This article will contrast different extensive form versions in a later section. The next session outlines the extensive form of when the Debtor moves first.

Extensive Form: Debtor Moves First

Figure 3 below depicts the extensive form of the balance of financial terror, with the Debtor moving first. The illustration solves the game with pure strategies via backward induction, with arrows portraying a player's best response in any given subgame.

Figure 3: Extensive Form of the Balance of Financial Terror, Debtor



In backward induction, players anticipate the moves available at the end of the game, and work backwards to ensure their most preferred outcome occurs. Here, the Debtor would look forward and anticipate the Creditor's moves in each of the three possible branches; the anticipated moves are marked above with dark gray arrows. This anticipation is possible because both players' payoffs are common knowledge. After anticipating these moves by the Creditor, the Debtor can then select that move which results in the Debtor's best possible outcome, or the one with the light gray arrow above.

In this case, backward induction finds that when the Creditor moves second, it will essentially match the Debtor's level of escalation. Hence, if the Debtor devalues, the Creditor will dump; if the Debtor weakens, the

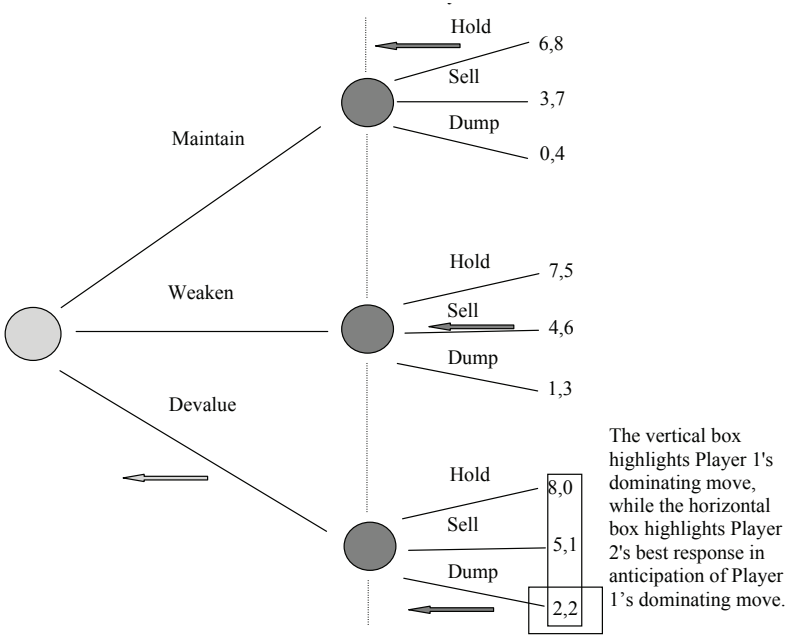
Creditor will sell; and if the Debtor maintains, the Creditor will hold (all marked with dark gray arrows above). Thus, the Debtor, anticipating these three best responses by the Creditor, will pick the game path that leaves the Debtor with the highest remaining payoff among the responses. In this case, the Debtor would rather maintain its currency at a strong level, so it does not face a partial debt sell-off, or worse, a debt dump. The final outcome is almost no escalation, or the status quo.

To date, this "(Maintain, Hold)" series of moves has been the essential dynamic of the U.S.-China relationship. The United States prefers that China continues to hold its debt, and so the United States avoids any moves that suggest too strongly it will cheat on its debt repayment through devaluation. China, in response, rewards the United States for this "good" behavior by continuing to hold U.S. debt. While both players may feel that they are in precarious position—the United States fears China could dump the debt at any moment, and China fears the United States will similarly inflate away its currency—China still moves second. Consequently, the United States avoids provoking a Chinese dump, and Chinese debt holdings remain safe for the period in question. This dynamic explains why both players find it so hard to escape the balance of financial terror. All things being equal, the balance sustains itself.

Extensive Form: Simultaneous Moves

A shift to a game with to simultaneous moves results in a different outcome. In Figure 4, the vertical line between the three light gray nodes indicates that the Creditor does not know the Debtor's move that came before it, and so the Creditor must make its own move in ignorance. In other words, any of the three light gray Creditor nodes could be the state of the game, and the Creditor cannot distinguish between them.

Figure 4: Extensive form of the Balance of Financial Terror, Players Moving Simultaneously



Simultaneous games cannot be solved by backward induction, since backward induction requires perfect information. However, other methods to solve the game are available. In this case, the game can be solved through the iterated elimination of dominated moves. A move is dominated if another move would yield a higher payoff for all the possible outcomes. Here, the Debtor's dominating move has its outcomes boxed in dark gray and the Creditor's best response anticipating this dominating move is boxed in light gray. Basically, for the Debtor the moves "Maintain" and "Weaken" are both dominated by the move "Devalue." Hence, for any given response by the Creditor, the Debtor's move "Devalue" will always yield a higher payoff. To see why devalue dominates, compare the Debtor's payoffs in the bottom branch of the game tree to the Debtor's payoffs in the upper two branches of the game tree. The bottom branch's payoffs are always higher for the Debtor for any given level of response by the Creditor.

Nevertheless, since the Debtor clearly has a dominating move, the Creditor can anticipate this move and respond accordingly. Hence the

Creditor, anticipating a dominating "Devalue" from the Debtor, will play "Dump" to be safe. In other words, dumping the debt is a best response a Creditor can do if it believes the Debtor will certainly devalue on its debt.

Thus, when we make the game simultaneous—or at least make the Creditor ignorant of the Debtor's choice and so effectively simultaneous—we find that the outcome for both parties is worse than if the Debtor had moved first. That is, instead of the balance of financial terror perpetuating under an equilibrium of "(Maintain, Hold)," the balance of financial terror collapses under "(Devalue, Dump)" where both players hedge against the most escalatory moves of the other player. By just slightly changing the information set available to the Creditor, we find that the outcome changes drastically. This drastic change for the worse for both players suggests that both players have an incentive to keep the other informed of its actions.

Still, for the United States and China, it is difficult to conceive of when a simultaneous game could most plausibly model the balance of terror. When one compares the simultaneous model here to the sequential model already covered above, China seems to have every incentive to wait and force the game to be sequential. Only by waiting to move in response to the United States can China credibly deter the United States from devaluation—otherwise, the United States would devalue anyway to yield its most dominant payoffs. Similarly, the United States has an incentive to move first instead of simultaneously, because it can anticipate the disaster that would result from simultaneous moves. In short, since both players know that perpetuating the balance of terror is preferred to mutually assured destruction, both players will structure the sequence of their moves to avoid such destruction.

If we set up the game so that the Creditor moves first and the Debtor moves second, we find a very similar outcome to the simultaneous game.

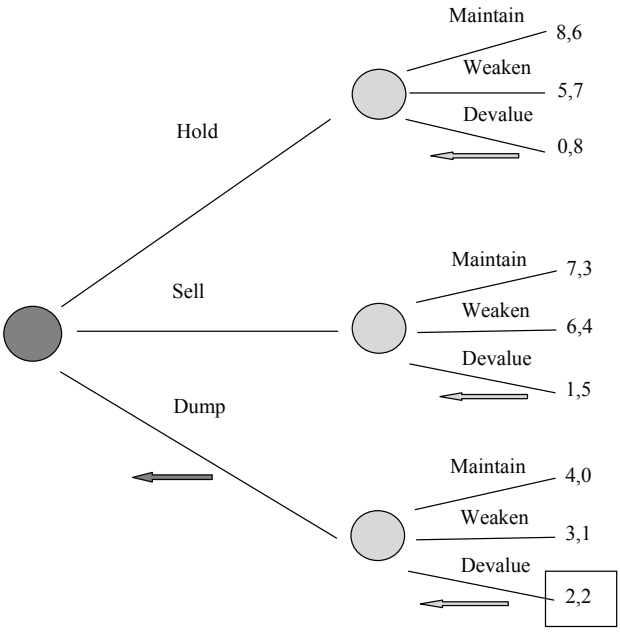
Extensive Form: Creditor Moves First

Figure 5 below depicts the game with the Creditor moving first, or the reverse of the first example. Since the moves are not simultaneous, backward induction can be used again. Nevertheless, this version of the sequential game results in the same outcome as the simultaneous game before; namely "(Dump, Devalue)." Note that this illustration reverses the payoff notation of the previous illustrations, to reflect the new sequence of the Creditor as Player 1 and the Debtor as Player 2.

The logic of dominated moves still holds when the Debtor moves second, just as it did when the players moved simultaneously. As a result, the Debtor's move "Devalue" dominates both in the simultaneous form

of the game and in the sequential version when the Creditor moves first. And like in the simultaneous game, the best response by the Creditor to "Devalue" is "Dump"; in short, the Creditor dumps in anticipation of devaluation.

Figure 5: Extensive form of the Balance of Financial Terror, Creditor Moving First



Nevertheless, all other things being equal, it is still unclear why the two players would ever allow the Creditor to move before the Debtor in the first place (just as it is unclear why the two would ever allow simultaneous moves). Both can anticipate the lower-level equilibrium that would result from the Debtor moving first, and China would presumably like to keep its deterrent against U.S. devaluation. When China moves first, it loses this ability to deter.

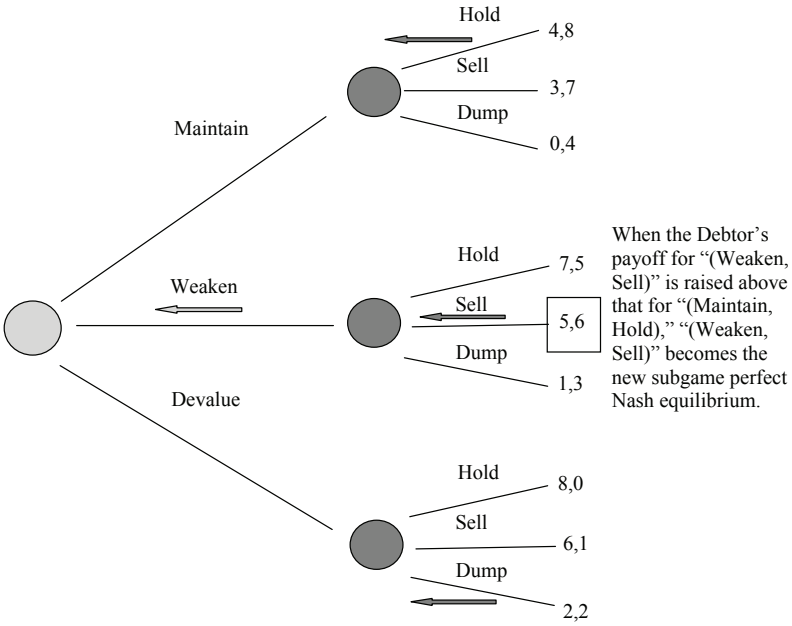
VI. Extensive Form of the Balance of Financial Terror: Modifying Payoffs

In the above versions of the balance of financial terror game, this article has changed only one aspect: the sequence of the players' moves. In contrast it has left the payoffs the same throughout. The article justified these payoffs in the "Preferred Outcomes" section, and for the most part, justified them according to reasons internal to the logic of the debtor-creditor relationship. Nevertheless, there is always a possibility that exogenous factors could affect the relative rankings of these payoffs and accordingly change the players' moves.

For the United States and China, one such exogenous shock particularly stands out: that the Debtor's payoffs from weakening its currency will outrank even the highest of those from maintaining the strength of its currency. Why would this happen? Consider that weakening a currency might not only be an intentional move by the United States to escape from its debt burden, but could also be an incidental effect from other actions. For example, if the U.S. Federal Reserve engages in quantitative easing to stave off deflation, this could have the incidental effect of weakening the U.S. dollar on foreign exchange markets and of reducing the value of China's U.S. debt holdings. For example, in fall 2010 the Federal Reserve considered such a quantitative easing policy. In response to this and other central bank interventions, some foreign governments warned of "currency wars" (Wei 2010) or the intentional weakening of national currencies to boost national exports. Regardless of the intent of quantitative easing, presume that it at least has the incidental effect of weakening the U.S. dollar. How should we expect the Debtor's payoff rankings and the Creditor's equilibrium response to change from our original game?

Arguably, quantitative easing makes the incidental effect of currency weakening attractive enough that the Debtor prefers the outcome "(Weaken, Sell)" even to "(Hold, Maintain)." As a result, when one solves the modified game through backward induction, the Debtor plays "Weaken" and the Creditor plays "Sell" in response.

Figure 6: Extensive Form of the Balance of Financial Terror, Payoffs Modified for QE



In this example, the Debtor country has stopped playing purely according to the logic of the balance of financial terror, even though the Creditor has continued to play according to such logic. By changing the Debtor's payoffs, it is as if we took the original game and knocked the Debtor off its old subgame-perfect move of "Maintain" and onto the subgame-imperfect move of "Hold." The Creditor still plays its best response, which is now the move "Sell."

Thus, for the United States and China, we might expect China to gradually sell its U.S. debt holdings in response to quantitative easing by the Federal Reserve. This result presumes that such quantitative easing truly weakens the U.S. dollar on foreign exchange markets; if the easing has no such incidental effect, the old payoffs would be upheld, and so would the old equilibrium of "(Maintain, Hold)." The point here is merely to make policy makers aware that exogenous factors must be considered when analyzing the balance of financial terror. Such factors could change the relative payoffs of one player, and elicit new responses as a result.

VII. Conclusion

By using game theory to model the balance of financial terror in three different player sequences, this article suggests that it is in both the creditor's and debtor's interest to have debtors move first on the values of their currencies. While such a debtor-first sequence perpetuates the balance of terror, both parties should prefer this perpetuation to the alternative—a retaliatory currency devaluation and dump of the debtor's debt. In short, the creditor has an incentive to wait and the debtor an incentive to move first, because doing so gives each a better outcome than the cataclysmic alternative. Additionally, policy makers from both sides of the balance should be attuned to the incidental effects of their policies. In the U.S.-China case, the Federal Reserve—an institution largely shielded from political influence and tertiary to American foreign policy—could affect the balance of financial terror by setting the United States on a currency-weakening equilibrium path. Such currency weakening could then provoke, in response, gradual selling of U.S. debt by China. Ironically, the balance of financial terror could then be resolved unintentionally via a path of medium escalation—not quite a cataclysmic end, but not quite a perpetual terror, either.

So far, the effect of the Federal Reserve's quantitative easing on China's payoffs remains uncertain. While Chinese treasury holdings gradually declined from a peak in October 2010 at $1175.3 billion to $1160.1 billion in December 2010 (U.S. Treasury 2010), such small fluctuations have occurred before between December 2009 and February 2010, as well as between April 2010 and June 2010. As a result, the recent declines do not necessarily signal that China has conclusively changed its debt purchasing policy as a response to the Federal Reserve's action. Hence, we will probably only have a better understanding after more time passes, and the data for the January to March period becomes available. If the Federal Reserve continues with its quantitative easing policy, and China continues to sell its debt, we might conclude that China's relative payoffs really have changed.

Until that or some other exogenous shock occurs, however, the balance of financial terror will likely remain surprisingly well-balanced for the foreseeable future. Good policy makers on both sides of the balance should be scenario-planning for other possible shocks, and using game theory to analyze their effects. Indeed, such shocks may not only be financial: geopolitical events, such as a radical Taiwanese declaration of independence, could shock the equilibrium as well. In the Taiwanese case, China may

value territorial integrity enough to raise its payoffs for dumping U.S. debt. Doing so could provoke a financial catastrophe, but for China, the alternative of losing a province could be worse. As a result of this and many other possible variations, the opportunities for further game theoretical research into the issue are plentiful. With more research, both countries might better understand the policy limitations of the other, and so avoid the accidental triggering of financial ruin.

References

Branegan, Jay, Alexander L. Taylor III and Frederick Ungeheuer. 1982. "The Wobbly World of Banking." *Time*, September 6. http://www.time.com/time/magazine/article/0,9171,949559,00.html (accessed January 10, 2011).

Davis, Bob. 2010. "Why Leaders Want to Rebalance Growth." *wsj.com*, November 8, sec. Economy.

http://online.wsj.com/article/SB10001424052748704049904575554322528218424.html (accessed January 10, 2011).

Dean, Jason, Andrew Browne, and Shai Oster. 2010. "China's 'State Capitalism' Sparks a Global Backlash." *wsj.com*, November 16, sec. World News. http://online.wsj.com/article/SB10001424052748703514904575602731006315198.html?mod=WSJ_article_MoreIn_Economy (accessed January 10, 2011).

Krugman, Paul. 2009. "China's Dollar Trap." *The New York Times*, April 3, sec. Opinion. http://www.nytimes.com/2009/04/03/opinion/03krugman.html?_r=2&ref=opinion (accessed January 10, 2011).

U.S. Department of Treasury. 2010. "Major Foreign Holders of Treasury Securities." Last modified December 15. http://www.treasury.gov/resource-center/data-chart-center/tic/Documents/mfh.txt (accessed March 4, 2011).

Wei, Lingling. 2010. "Developing Countries Keep Buying Dollars." *wsj.com*, December 31, sec. Markets Main. http://online.wsj.com/article/SB10001424052748703909904576051892942222126.html (accessed January 10, 2011).

4

Targeted Killings: Does Drone Warfare Violate International Law?

Rebecca Perlman

Targeted killing has been heralded as one of the most effective methods for reducing the terrorist threat in the Middle East, yet its legality remains a point of controversy. At issue is the question of whether the United States is, or even can be, at war with al-Qaeda, as a state's recourse to violence is severely restricted under international law in the absence of such a war. This paper analyzes the three main frameworks under which America's lethal actions have been evaluated: law enforcement, armed conflict, and self-defense. It concludes that while the United States has a legitimate claim to self-defense against these terrorist networks, targeted killing as currently practiced by the Obama administration cannot be justified under international law.

I. Introduction

On his inauguration day, President Barack Obama declared, "Our nation is at war against a far-reaching network of violence and hatred" (Obama 2009). Yet this "war" remains a point of controversy. Under international law, a war is recognized only under very specific circumstances, and it must be established that the United States is truly at war with al-Qaeda and associated Islamic jihadist organizations if U.S. military operations

Rebecca Perlman is a second year graduate student in the Master of Arts in Law and Diplomacy program at Tufts University's Fletcher School, where she focuses on security studies and international political economy. She can be reached at becca.perlman@gmail.com.

against them are to be declared legal. This is of particular relevance to the United States' strategy of targeted killing currently being employed in Afghanistan and surrounding regions.

International legal scholar Philip Alston defines targeted killing as "the intentional, premeditated and deliberate use of lethal force, by States or their agents acting under colour of law, or by an organized armed group in armed conflict, against a specific individual who is not in the physical custody of the perpetrator" (Alston 2010, 3). If performed outside of the law, targeted killing is considered murder, assassination, or an "extra-judicial execution" (Kretzmer 2005, 174), and various scholars contend that the use of targeted killing by the Obama administration in places like Yemen and Pakistan falls clearly into one of these latter three categories (see O'Connell 2010a).

Relevant laws on this matter, however, are rendered ambiguous by the fact that al-Qaeda and associated groups are not part of a single state. Beyond this, Alston argues that a terrorist organization fails even to meet the legal definition of a "party" to a conflict given that "al-Qaeda and other alleged 'associated' groups are often only loosely linked [to each other], if at all" (Alston 2010, 18). As a result, it is somewhat unclear which international rules apply to the United States' conduct of military operations abroad, as the nation seeks to protect itself against the threat of terrorism.

This article seeks to shed light on how targeted killing, as currently practiced by the Obama administration, might be evaluated under international law. The focus is on targeted killings that have taken place outside of any readily defined war zone, a feature that renders them particularly problematic. The three broad frameworks that have been most commonly used to evaluate U.S. drone strikes are analyzed within this context: law enforcement, armed conflict, and self-defense under Article 51 of the United Nations Charter.

This article examines both the appropriateness of these frameworks for addressing this issue and their implications for the legality of America's targeted killing strategy outside official theaters of war. It is argued that while self-defense under Article 51 provides some legal justification for targeted killings in countries including Yemen and Pakistan, the United States has overextended its claim to self-defense, thereby operating outside of what is permitted under international law. Such flouting of the international legal regime in this manner may prove harmful for a number of reasons. These are well articulated by University of Oxford Professor Adam Roberts:

> First, in all military operations, whether or not against terrorists, a perception that a state or a coalition of states is observing basic international standards may contribute to public support within the state or coalition; support, or at least tacit consent, from other states; and avoidance of disputes within and between coalition member states. Second, if the coalition were to violate *jus in bello* in a major way [...] that would help the cause of the adversary forces and even provide them with a justification for their resort to force under *jus ad bellum*. Third, in anti-terrorist campaigns in particular, a basis for engaging in military operations is often a perception that there is a definite moral distinction between the types of actions engaged in by terrorists and those engaged in by their adversaries. Observance of *jus in bello* can form a part of that moral distinction (Roberts 2002, 9).

In addition, one could argue that the violation of international legal norms encourages other nations to follow suit. For example, when Russian lawmakers in 2006 authorized their "security services to kill alleged terrorists overseas" (Alston 2010, 9), they "insisted that they were emulating Israeli and US actions in adopting a law allowing the use of military and special forces outside the country's borders against external threats" (Romero and Warren 2010). For all of these reasons, an understanding of the legal issues serves a very relevant purpose.

II. Background

Beginning in 2002 with the targeted killing in Yemen of Qaed Senyan al-Harithi, the alleged mastermind of the U.S.S. Cole bombing, both the Bush and Obama administrations have consistently used targeted killings in order to eliminate suspected members and affiliates of al-Qaeda. Yet, while these activities have generated significant attention, the United States is not the only country to utilize such methods. Israeli officials have officially acknowledged a policy of targeted killing against terrorists since 2000, claiming that such acts represent a legitimate means of self-defense under the laws of war, collectively known as international humanitarian law (IHL) (Alston 2010, 6). However, prior to September 11, 2001, the United States "routinely denounced" (Ofek 2010) Israel's policy. It is, therefore, somewhat ironic that in March of 2010 Harold Koh, the current Legal Advisor of the Department of State, used this very argument of self-defense to justify American targeted killings (Koh 2010).

Whereas such killings have traditionally been carried out by air or

ground military forces, new developments in robotics allow strikes to be performed by unmanned aerial vehicles, more commonly referred to as drones, which are operated remotely from a position of safety. These planes are so advanced that they can transmit a readable image of a license plate from a distance of two miles (Singer 2009, 33), and armed with Hellfire missiles, they can easily demolish the car to which such license plate is attached from the same distance, all while the pilot remains far removed from physical danger.

Drones have proven so invaluable in hunting down and eliminating alleged terrorists that CIA Director Leon Panetta has referred to them as "the only game in town" (Panetta 2009). Furthermore, intelligence reports have "revealed growing examples of Taliban fighters who are fearful of moving into higher-level command positions because of these lethal operations" (Cooper and Landler 2010).

Nevertheless, targeted killings raise serious concerns, especially when executed outside of demarcated war zones in which the international community has recognized the existence of an armed conflict. In this regard, it is important to highlight two distinct targeted killing programs. The first of these is run by the military, openly acknowledged by the U.S. government, and solely "targets enemies of U.S. troops stationed" (Mayer 2009) in Iraq and Afghanistan. The second program is a highly classified CIA operation "aimed at terror suspects around the world, including in countries where U.S. troops are not based" (Mayer 2009).

While the U.S. military's targeted killings within official war zones are not particularly controversial, the CIA's operations have been more contentious, as indicated by the sheer bulk of academic literature and media attention dedicated to the matter. No country is allowed to kill individuals at will and, outside of an official armed conflict, any resort to lethal means is strictly prohibited except in extreme cases, such as defense of life. This suggests that, barring the existence of a war, targeted killings transgress the individual right to life, as guaranteed by numerous bodies of international and domestic law, including the International Covenant on Civil and Political Rights (ICCPR).[1]

Moreover, the U.S. government's use of the CIA to execute these missions has been widely decried. Unlike members of the armed forces, CIA officials are not trained in the laws of war and do not bear the uniforms that serve to adequately distinguish them from civilians. In the eyes of many, this makes them unlawful combatants,[2] subject to attack "whenever and wherever they may be found, including Langley [VA]" (Solis 2010).

Finally, some have questioned the legitimacy of including those who

finance the Taliban on the list of permissible targets (Alston 2010, 19). Since in most cases these individuals are not directly participating in hostilities, they fall into the international legal category of "noncombatants" (Convention IV 1949, Art. 3.1), making their targeting almost certainly unlawful.

Members of the Obama administration nonetheless insist that they are "committed by word and deed to conducting ourselves in accordance with all applicable law" (Koh 2010). Yet an analysis of the competing claims surrounding this issue makes clear that one would be hard pressed to defend certain targeted killings in Pakistan and Yemen under international law. In addition, the use of the CIA to conduct these operations, as well as the inclusion of those not actively engaged in terrorism on targeted killing lists, further undermines claims of legitimacy.

III. Law Enforcement

In his widely cited book on targeted killing, Nils Melzer, the legal advisor for the International Committee of the Red Cross, argues that "the normative paradigm of law enforcement must—'by default' and regardless of temporal and territorial considerations—govern the international lawfulness of all State-sponsored targeted killings except those directed against a legitimate military target in a situation of armed conflict" (Melzer 2008, 223). In the case that no such conflict can exist between the United States and the terrorist groups in question, outside the borders of Afghanistan and Iraq targeted killings must be judged by their adherence to international law enforcement standards.

Law enforcement rules stipulate that resort to lethal force is legitimate only within very narrow parameters. Just as most domestic legal systems provide that a policeman may not intentionally kill an individual except in "defense of life" (Alston 2010, 22), law enforcement agents acting internationally are similarly restricted. These agents are also further bound by considerations of state sovereignty and the requirement not to disturb the peace between nations (Blum and Heymann 2010, 146).

The guarantees that ensure an individual cannot be intentionally killed unless he or she poses an immediate threat to another are found not only in most countries' domestic laws but, more importantly for the purposes of this analysis, in international human rights law. The principle has been incorporated into international law through a number of legal agreements including the ICCPR and the American Convention on Human Rights (ACHR), both of which guarantee "protection from 'arbitrary' deprivation of life" (Melzer 2008, 91-2).

Most scholars concur that as a result of these agreements, "[u]nder the international normative paradigm of law enforcement, the lawful use of lethal force may not exceed what is 'absolutely' or 'strictly' necessary to maintain, restore or otherwise impose law and order in the concrete circumstances" (Melzer 2008, 227-8). It follows from this that "it is never permissible for killing to be the *sole objective* of an operation" (Alston 2010, 11). Rather, killing can only be used to protect against a present and direct threat to life.

Considering that drones have executed the majority of targeted killings in Pakistan and Yemen, one would be challenged to argue that any such killings could be defended under claims of immediate necessity. Unlike a soldier or a policeman who might confront physical danger in his or her attempt to arrest a suspected terrorist, drones are operated from a "suburban redoubt" (Mayer 2009) far removed from harm's way, thereby rendering claims of personal self-defense unpersuasive. Should the law enforcement paradigm apply, therefore, these killings are likely "tantamount to extra-judicial execution or murder" (Blum and Heymann 2010, 146).

Furthermore, considering that "[a]s a general principle of international law, a country is strictly prohibited from engaging in law enforcement operations in the territory of another country" (Blum and Heymann 2010, 161), any such killings violate not only the right to life, but also the right to state sovereignty in cases where the nation in question has not given the United States permission to use force. In those cases where permission is granted, such as when the Yemeni government condoned the killing of al-Harithi, the United States would still be guilty of illegal execution. As Notre Dame law professor Mary Ellen O'Connell observed in her testimony before the U.S. Congress, "States cannot…give consent to a right they do not have" (O'Connell 2010a, 2).

However, viewing targeted killings through a law-enforcement lens may be somewhat misleading. Legal scholar David Kretzmer points out that "[t]he problem with the law-enforcement model in the context of transnational terror is that one of its fundamental premises is invalid: that the suspected perpetrator is within the jurisdiction of the law-enforcement authorities in the victim state, so that an arrest can be effected" (Kretzmer 2005, 179). In places like Yemen and Pakistan, the state authority clearly lacks the means, and in some cases the will, to locate and arrest the individuals who have been marked for killing. Lacking "an authority stable and strong enough to impose public security, law and order" (Melzer 2008, 88), the law enforcement paradigm seems to be an inadequate model for confronting the terrorist threat. In its 2003 analysis of Israel, The Human

Rights Committee did allow for a such a situation in which lethal methods could be employed, under the condition that "[b]efore resorting to the use of deadly force, all measures to arrest a person suspected of being in the process of committing acts of terror must be exhausted" (Concluding Observations of the Human Right Committee: Israel 2003). This suggests that in those cases where law enforcement measures fall short, one can look to a different legal framework.

IV. Armed Conflict

A second approach is to regard targeted killings not as law enforcement operations but as measures taken within the context of an armed conflict with the transnational terrorists. In this view, al-Qaeda's attacks against the United States, including but not limited to those carried out on September 11, 2001, are of such gravity as to create a situation of armed conflict. The existence of such a conflict means that all targeted killings against al-Qaeda and its associates should be evaluated under international humanitarian law (IHL).

IHL refers to the body of law that applies during times of armed conflict or war. These are based primarily on the Geneva Conventions of 1949 and the Additional Protocols of 1977, as well as customary international law. Because IHL only applies when a certain threshold of violence has been reached, these laws are far more permissive of force than is international human rights law; under IHL, individuals can be targeted solely on the basis of their status as "combatants" rather than as a result of the immediate threat they pose. It follows then that if a state is legally engaged in an armed conflict with al-Qaeda and its associates, any and all members of those groups are considered legitimate targets under international law.

Based on the above analysis, if it could be shown that the United States were in an armed conflict with the terrorist groups responsible for the September 11 attacks, targeted killings in Yemen and Pakistan might well be justified if they abide by the rules of IHL. Yet just as the law enforcement model is flawed for analyzing the current context, so too is the paradigm of armed conflict, for the reasons iterated below.

International human rights law acknowledges two types of armed conflicts—international and non-international—and applies different legal criteria to each. The three legitimate types of international armed conflict are laid out in Article 2 of the Geneva Conventions and have been widely interpreted as occurring only in conflicts between two states, a reading that was affirmed in *Prosecutor v. Tadic* at the International Criminal Tribunal for the Former Yugoslavia. Based on this understanding, the

conflict between the United States and al-Qaeda and its affiliates is not an international one.

The question then becomes whether it might be a non-international armed conflict. If so, it would trigger the application of Article 3 of the Geneva Conventions, which details the legal rules that apply "[i]n the case of armed conflict not of an international character occurring in the territory of one of the High Contracting Parties" (Convention (IV) relative to the Protection of Civilian Persons in Time of War). According to University of Texas Law Professor Derek Jinks, Common Article 3 ought to apply to "all 'armed conflicts' not covered by Common Article 2" (Jinks 2003, 41). In his view, so long as the September 11 attacks meet the threshold of an "armed conflict," which he argues they do, these acts fall under the classification of an "armed conflict not of an international character," thereby bringing into play Article 3 of the Geneva Conventions and the relevant humanitarian laws.

Harold Koh's remarks at the annual meeting of the American Society of International Law in March of 2010 suggest that the Obama administration is in full agreement with Professor Jinks. During his speech, Koh stated that "as a matter of international law, the United States is in an armed conflict with al-Qaeda, as well as the Taliban and associated forces, in response to the horrific 9/11 attacks, and may use force consistent with its inherent right to self-defense under international law" (Koh 2010).

Nevertheless, various scholars disagree with Harold Koh and Derek Jinks' interpretation of the circumstances under which Common Article 3 applies and whether September 11 triggered an armed conflict under international law. The first question in regard to Article 3 is whether the phrase "not of an international character occurring *in* the territory of one of the High Contracting Parties" denotes only conflict that occurs *within* rather than *between* states. If so, the U.S. war on terror when conducted overseas does not meet this requirement. This matter is addressed in Additional Protocol II, which reads that Article 3 "shall apply to all armed conflicts which are not covered by Article 1…which take place in the territory of a High Contracting Party…" (Protocol II, Art. 1.1). Conventional wisdom on this subject provides that when combined with Additional Protocol II, Common Article 3 does not exclusively apply to conflicts enclosed within state lines (Bassiouni 2002, 99).

A more serious deficiency in applying Article 3 to current U.S. operations is that terrorist groups may not meet the threshold necessary to be considered an official member of an armed conflict at all. Additional Protocol II specifies that the "dissident armed forces or other organized

armed groups" must be "under responsible command" and "exercise such control over a part of [the state's] territory as to enable them to carry out sustained and concerted military operations and to implement this Protocol" (Protocol II Art. 1.1). It is far from clear that this holds true in the situation under discussion.

Perhaps an even more compelling argument is that the September 11 attacks would not meet the requirements for initiating an armed conflict regardless of whether al-Qaeda and its associates constitute a group in the relevant sense. According to Marco Sassoli, a professor of international law at the University of Geneva, "terrorist acts by private groups…have not customarily been viewed as creating armed conflicts" (Sassoli 2004, 202). Protocol II actually specifies that "[t]his Protocol shall not apply to situations of internal disturbances and tensions, such as riots, isolated and sporadic acts of violence and other acts of a similar nature, as not being armed conflicts" (Protocol II, Art. 1.2). While authorities are divided on whether terrorist attacks against the United States classify as "sporadic," few contend that terrorism against U.S. interests represents the "sustained, persistent fighting" (Anderson 2010b) required to constitute an armed conflict. As a result, it is helpful to turn to the final framework of self-defense.

V. Self-Defense under Article 51

The self-defense framework for describing the current conflict with the terrorist groups takes the middle road between the law enforcement and armed conflict models. While it acknowledges the inadequacy of the law-enforcement model in accounting for the current situation, it avoids the assertion that the confrontation has reached the level of an armed conflict. Instead, this framework posits that self-defense provides the United States with a justification to which it can legally resort.

Article 51 of the United Nations Charter maintains that "[n]othing in the present Charter shall impair the inherent right of individual or collective self-defence if an armed attack occurs against a Member of the United Nations, until the Security Council has taken measures necessary to maintain international peace and security" (Charter of the UN Art. 51). This statement is not qualified by any threshold requirement of armed conflict. Consequently, a nation can claim this right without touching on the issue of armed conflict at all. Having done so, "a state's actions are subject to the requirements of necessity and proportionality" (Kretzmer 2005, 203).

The appeal of availing oneself of the self-defense argument is quite clear. Through it, the United States can sidestep the fraught debate over

whether a dispersed network of terrorist groups can be party to an armed conflict. Furthermore, the model accounts for a key shortcoming in the law enforcement paradigm by authorizing a more robust response to the September 11 attacks, particularly in the regions of Yemen and Pakistan where the rule of law is weak. Finally, it still falls squarely into a well defined section of international law for which there are clear rules and guidelines.

The UN Security Council has supported this position of self-defense in two separate resolutions passed in the immediate aftermath of the September 11 attacks. On September 12, 2001, under Resolution 1368, the Security Council recognized "the inherent right of individual or collective self-defense in accordance with the Charter" and "expresse[d] its readiness to take all necessary steps to respond to the terrorist attacks of 11 September 2001" (UN Security Council Resolution 1368 2001). On September 28, 2001, the Security Council reaffirmed this stance, adding that, "States shall…[t]ake the necessary steps to prevent the commission of terrorist acts…" (UN Security Council Resolution 1373 2001), effectively sanctioning the United States' claim to self-defense. The subsequent North Atlantic Treaty Organization (NATO) endorsement of the Security Council's position served to further validate U.S. strategy developed in the aftermath of the terrorist attacks (Paust 2010, 248).

Yet, even in the absence of U.N. Security Council and NATO endorsement, historical precedent supports the United States' right of self-defense. As George Washington University Law Professor Sean Murphy has pointed out, "[T]he destruction wrought [on September 11] was as dramatic as the Japanese attack on Pearl Harbor on December 7, 1941…[and] the death toll from the incidents was worse than Pearl Harbor" (Murphy 2002, 47). Since Pearl Harbor provided adequate provocation to elicit a U.S. declaration of war on Japan, by this standard the September 11 strikes constituted an "armed attack" under Article 51.

Also relevant is the fact that "the United States immediately perceived the incidents as akin to that of a military attack. President Bush declared a national emergency and called to active duty the reserves of the U.S. armed forces" (Murphy 2002, 47). The U.S. Congress, clearly in agreement with the President, authorized him:

> [T]o use all necessary and appropriate force against those nations, organizations, or persons he determines planned, authorized, committed, or aided the terrorist attacks that occurred on September 11, 2001, or harbored such organizations or persons, in order to prevent any future acts of international terrorism against the United States by such nations, organizations or persons (107th Congress 2001, Section 2).

Professor Mary Ellen O'Connell takes issue with this entire line of reasoning, arguing that "[a]n armed response to a terrorist attack will almost never meet [the] parameters for the lawful exercise of self-defense. Terrorist attacks are generally treated as criminal acts because they have all the hallmarks of crimes, not armed attacks that can give rise to the right of self-defense" (O'Connell 2010b, 14). However, as mentioned above, the law enforcement model is not appropriate considering the lack of adequate governance in the areas of Pakistan and Yemen where many of the perpetrators are located. In addition, O'Connell's arguments seem overly dogmatic given the UN Security Council's implied endorsement and NATO's full authorization for the United States to act in self-defense under Article 51.

Nevertheless, O'Connell maintains that "the Security Council has not authorized attacks, and the U.S. has no right on that basis to use drones. In the wake of the 9/11 attacks, the Security Council did find in Resolution 1368 that the attacks triggered Article 51 self-defense. The Council did not, however, authorize the use of force against any particular state" (O'Connell 2010b, 19).

Effectively, O'Connell is equating drone strikes against terrorists residing in a state with strikes against the state itself. However, it is important to note that Article 51 specifically says, "Nothing in the present charter shall impair the inherent right of individual or collective self-defence," (Charter of the UN Art. 51). This, therefore, overrides the protection of state sovereignty found in Article 2(4) of that same document, thereby refuting any portion of O'Connell's argument based on principles of state sovereignty.

Nonetheless, while hypothetically it does not violate sovereignty to use force against certain individuals within a state, this does not rule out the possibility that such force could be interpreted as an act against the state itself. Certainly, if the state in question cannot or will not eliminate the threat emanating from it, the United States' decision to take action into its own hands does not necessarily constitute force against that state (Alston 2010, 11-12). The distinction, however, is fine. Where the use of force produces numerous civilian casualties, it is difficult to argue that this does not constitute, to some degree, an attack on the nation itself. While this matter is revisited in the proceeding section, for now suffice it to acknowledge that as long as the United States avails itself of its right to self-defense in a "selective and proportionate manner merely against non-state actors that are perpetrating, aiding, or directing ongoing armed attacks" (Paust 2010, 258), this need not constitute an attack on the harboring state.

VI. Analysis of U.S. Targeted Killings under Article 51 of the UN Charter

If it is accepted that the United States has a legitimate recourse to force through Article 51 of the UN Charter, what are the legal guidelines under which it must operate? First, all actions must be genuine expressions of "self-defense." In order for a targeted killing to meet this requirement, it must be both "necessary" and "proportional," a condition derived from customary IHL.

For a strike to be considered necessary under international law, it must seek to prevent a future attack from occurring. In the current case, this means that "[t]he only acceptable justification for targeting suspected terrorists is protection of potential victims of terrorist acts" (Kretzmer 2005, 202). Judging whether this requirement has been fulfilled is not always simple. For example, one might ask: did the U.S. target al-Harithi because he was actually planning future attacks or simply based on his prior role in the *U.S.S. Cole* bombing? The answer weighs heavily on the legality of his killing. If al-Harithi was killed solely in reaction to an incident that had occurred two years earlier, "the Predator attack would be considered punitive rather than defensive, an act of reprisal that is judged to be illegal by the vast majority of states" (Downes 2004, 286).

By the same token, even if a particular targeted killing is undertaken for non-punitive reasons, international law still requires the expected threat from that individual to be so imminent that "the necessity of self defence, [is] instant, overwhelming, leaving no choice of means and no moment for deliberation" (Webster 1841). Short of this, the attack falls beyond the purview of legitimate preemptive self-defense and into preventive self-defense, which is not recognized as permissible under international law. While there still may be times that the requisite threshold is reached, it remains imperative that each attack be based on legitimate motives that seek to prevent the otherwise unavoidable loss of life.

Another consideration that must be factored into the necessity analysis is whether lethal action represents the only means of accomplishing the intended good. Nils Melzer describes this requirement as follows: "[T]here must be no non-lethal alternative which would entail a comparable military advantage without unreasonably increasing the risk to the operating forces or the civilian population" (Melzer 2008, 397). It follows from this condition that targeted killing is not legally justified in cases where arrest is possible.

Of course, there are cases in which it is not possible to arrest a suspected

terrorist; in those instances the proposed targeted killing may also be necessary as a true act of self-defense. When this theoretical condition is satisfied, the next question becomes whether the strike is proportionate. Harold Koh has defined the rule of proportionality as "prohibit[ing] attacks that may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, that would be excessive in relation to the concrete and direct military advantage anticipated" (Koh 2010). Since the direct military advantage of a targeted killing is simply preventing that individual from aiding in, carrying out, or planning a terrorist attack, the elimination of that person must be weighed against the collateral damage.

Compared to many of the lethal means available to the United States for the purpose of combat operations, drones are precision instruments. The popular Predator model is equipped with highly advanced video capabilities and precision lasers to guide its missiles to the correct target. Despite this, targeted killing operations have on various occasions produced substantial collateral damage. For example, a drone strike against Baitullah Meshud, a Taliban leader in Pakistan, resulted in the death of eleven other individuals, including Meshud's wife and mother-in-law. Even more concerning, in the numerous failed attempts to kill Meshud, "between two hundred and seven and three hundred and twenty-one additional people were killed" (Mayer 2009), some of whom appear to have been children and many of whom were innocent by law of war standards. These numbers raise doubts about whether such an attack was proportionate, even acknowledging that some of the casualties were surely al-Qaeda operatives.

If such a level of collateral damage does not seem unreasonable for a high priority target, it might still be concerning that President Obama has now expanded the target list to include drug lords known to finance the Taliban. Scholars, such as Harvard Law School Professors Gabriella Blum and Philip Heyman, have argued that drug lords, regardless of whom they finance, must be dealt with using law enforcement methods as opposed to lethal force (Blum and Heymann 2010, 148). Certainly, it does seem a stretch to claim that these drug traffickers pose an imminent danger to the United States, or that Hellfire missiles constitute a proportionate response. When one further considers that Predator strikes are usually accompanied by collateral damage, the argument against targeting these individuals is reinforced.

Another issue that arises in the discussion of collateral damage, as briefly discussed previously, is whether extensive civilian death outside of a war zone, even in pursuit of a high level target, constitutes a breach of

state sovereignty. Philip Alston points out that even if a country withholds consent to the use of force on its territory, "[s]tates may invoke the right to self-defence as justification for the extraterritorial use of force involving targeted killings…as long as that force is necessary and proportionate" (Alston 2010, 12). This argument potentially provides a legal foundation for using force within a sovereign and non-belligerent state, even if collateral damage occurs, so long as that damage is "proportionate" to the anticipated military advantage. Nonetheless, in the face of countless civilian deaths, countries might increasingly start to see America's use of force as impinging upon its sovereign rights, regardless of what international law allows.

A further question pertains to where, in a geographical sense, the United States' right to self-defense ends. Professor Kenneth Anderson of American University provided the following testimony on this issue to the U.S. Congress:

> [W]hat is justified in the ungoverned regions of Somalia or Yemen is a different matter applied to places under the rule of law such as our friends and allies. The United States is not going to undertake a targeted killing in London. The diplomatic fiction of the "sovereign equality" of states makes it difficult to say, as a matter of international law that, yes, Yemen is different from France, but of course that is true. (Anderson 2010a, 10)

This formulation is based on a policy that does not hold under international law, meaning that it would be difficult to defend if challenged. While there may be some truth to the logic, failure to provide guidelines as to where the United States will or will not engage in targeted killings opens up the possibility of an ever-expanding notion of self-defense.

Finally, there is the significant question of whether it is appropriate for the CIA to operate the drones used for these lethal operations. President Obama has attempted to move away from the Bush administration's terminology of a "war on terrorism," preferring instead to refer to the situation as an "Overseas Contingency Operation" (Kamen 2009). However, if the administration continues to assert that the United States is in an "armed conflict," should it not use America's uniformed military to fight that conflict? After all, as Professor of International Law Michael Schmitt has stated, "[t]here are but two categories of individuals in an armed conflict, combatants and civilians. Combatants include members of a belligerent's armed forces and others who are directly participating in a conflict…the latter [if they participate in hostilities] are labeled unlawful combatants

or unprivileged belligerents" (Schmitt 2004-2005, 522).

In addition to the problem of ambiguous combat status, the CIA is not subject to the same levels of transparency and oversight as the armed forces, nor are CIA officials instructed in the laws of war (O'Connell 2010b, 7). According to O'Connell, these features "may alone account for the high-unintended death rate" (O'Connell 2010b, 7) of CIA targeted killings. This high death rate arguably diminishes the effectiveness of the strategy itself, which raises further questions of proportionality and necessity. An op-ed published by the Council on Foreign Relations asserted that "Pakistani intelligence agencies have reported that refugees from Afghanistan have flocked to the Taliban by the hundreds to avenge the drones' killings of innocent civilians" (Zenko 2009). If true, this suggests that collateral damage is actually compromising the value of targeted killings. If these strikes are not accomplishing their intended objectives and are also inflicting massive collateral damage, this implies that legally they are neither proportionate nor necessary; they may not even be advisable.

In addition, the lack of transparency inherent in any covert action increases the risk of targeting individuals who may not meet the legal criteria for necessity. After all, "strikes are only as accurate as the intelligence that goes into them" (Mayer 2009); when there is no independent oversight to ensure that intelligence is adequate to justify the killing, mistakes become more likely. According to a disturbing account cited by Jane Mayer in *The New Yorker*, local informants often "say an enemy of theirs is Al-Qaeda because they just want to get rid of somebody. Or they ma[ke] crap up because they want[] to prove they [are] valuable, so that they [can] make money" (Mayer 2009). With no subsequent review of targeted killing operations, and therefore no real risk of consequences, it becomes far too easy to take these inaccurate tips at face value.

VII. Conclusion and Recommendations

As the United States continues to seek ways to protect itself from transnational terrorists, it will be forced to make difficult decisions. At present, the government hovers between operating within international law and setting that law aside in the belief that doing so is necessary to protect U.S. citizens. Ongoing claims that the U.S. is engaged in an "armed conflict" under international law with al-Qaeda and its affiliates remain unconvincing. Even if such claims were credible, the administration continues to undermine them by using the CIA to conduct its targeted killing operations instead of relying upon the military. Consequently, if President Obama wishes to bring American actions into accordance with international legal

standards, he will need to make some major changes.

First, the administration needs to reformulate its legal arguments. There may be a legitimate foundation for self-defense under Article 51, but that does not translate into the existence of an "armed conflict." The administration should abandon rhetoric assuming such a conflict and embrace the far more convincing claim that targeted killings are justified solely through the United States' right to defend itself in the face of an armed attack.

Second, the current target list must be brought in line with a stricter definition of "self-defense." While drug lords who give money to the Taliban may pose a veritable threat, ordering them to be executed by CIA-operated drones using Hellfire missiles stretches the concept of self-defense too far. Furthermore, the collateral damage that inevitably occurs in eliminating these individuals is not proportionate to the advantage accrued by their removal. Therefore, in order to abide by accepted understandings of self-defense and to avoid potential encroachment on state sovereignty, it is necessary to amend the target list. Individuals who do not qualify for targeting under the new guidelines must be dealt with using less lethal methods that more closely resemble law enforcement activities.

Third, there needs to be far more transparency and oversight within the U.S. government's targeting programs than currently exists. The United States does have a legal right to defend itself from terrorists, and that right does extend to some targeted killing operations outside of official war zones. However, it is vital that the U.S. government treat such operations with equally, if not more, exacting regulations than they would apply in a war zone. This means that under U.S. policy, only the military and not the CIA, should be operating the drones, and there ought to be a standard level of review and oversight. If the government worries that increased transparency of the program could impede the targeting of certain individuals, perhaps this is an indication that the executive has been overstepping his authority.

Finally, the United States ought to work with the global community to develop an international legal framework equipped to address the new political realities of a post-September 11 world. While terrorists and armed groups have existed throughout history, today they play a far more conspicuous role in the international security environment. In order to avoid the legal ambiguity that has engendered so much controversy about what the United States can and cannot do to defend itself against these actors, the United States should lead the effort to develop international laws that directly address this issue.

These laws would seek to answer such pressing questions as: when can a nation use force in another state to target individual terrorists if a war between the two states does not exist? What is the threshold that determines when a country may resort to its right of self-defense? And how closely linked must terrorist groups be in order to be viewed collectively as legitimate targets? Until international law confronts these questions directly, the boundaries of permissible conduct will remain undefined, leaving states free to interpret the law as suits their interests.

Notes

[1] For full text of ICCPR see: http://www2.ohchr.org/english/law/ccpr.htm.

[2] Although the term "unlawful combatants" is not actually employed by the Geneva Conventions, it is frequently referenced in academic works. According to a legal advisor at the International Committee of the Red Cross, the phrase is most commonly "understood as describing all persons taking a direct part in hostilities without being entitled to do so and who therefore cannot be classified as prisoners of war on falling into the power of the enemy" (Dormann 2003, 46).

References

107th Congress, "Authorization for Use of Military Force." September 18, 2001. http://news.findlaw.com/wp/docs/terrorism/sjres23.es.html (accessed October 3, 2010).

Alston, Philip. 2010. "Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Addendum, Study on Targeted Killings." UN Doc. A/HRC/14/24/Add.6, at 3, 54, 85-86. May 28.

Anderson, Kenneth. 2010a. "Drones II." Testimony before Congress of the United States, House of Representatives, Subcommittee on National Security and Foreign Affairs Hearing. April 28. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1619819 (accessed November 12, 2010).

_____. 2010. "Predators Over Pakistan." *The Weekly Standard* 15, no. 24. (accessed November 11, 2010b).

Bassiouni, M. Cherif. "Legal Control of International Terrorism: A Policy-Oriented Assessment Focus: September 11, 2001--Legal Response to Terror." *Harvard International Law Journal 43, no. 1: 83-103.*

Blum, Gabriella, and Philip Heymann. 2010. "Law and Policy of Targeted Killing." *Harvard National Security Journal* 1, no. 145: 145-170.

Charter of the United Nations. *United Nations.* 1945. http://www.un.org/en/documents/charter/ chapter7.shtml (accessed November 11, 2010).

Convention (IV) relative to the Protection of Civilian Persons in Time of War. 1949. *International Committee of the Red Cross*. August 12. http://www.icrc.org/ihl.nsf/385ec082b509e76c41256739003e636d/6756482d86146898c125641e004aa3c5 (accessed November 7, 2010).

Cooper, Hellen, and Mark Landler. 2010. "Targeted Killing is New U.S. Focus in Afghanistan." *The New York Times*, July 31. http://www.nytimes.com/2010/08/01/world/asia/01afghan.html (accessed November 21, 2010).

Dormann, Knut. 2003. "The legal situation of 'unlawful/unprivileged combatants." IRRC 85, no. 849: 45-74

Downes, Chris. 2004. "'Targeted Killings' in an Age of Terror: The Legality of the Yemen Strike." *Journal of Conflict & Security Law* 9, no. 2: 277-294.

Executive Order 12333. December 1, 1981. https://www.cia.gov/about-cia/eo12333.html#2.11 (accessed November 7, 2010).

Human Rights Committee. "Concluding Observations of the Human Rights Committee: Israel." CCPR/CO/78/ISR. (August 21, 2003).

International Committee of the Red Cross. 2004. "What is International Humanitarian Law?" *International Committee of the Red Cross*, July 31. http://www.icrc.org/Web/eng/siteeng0. nsf/html/humanitarian-law-factsheet (accessed November 7, 2010).

Jinks, Derek. 2003. "September 11 and the Laws of War." *Yale Journal of International Law* 28, no. 1: 1-50.

Kamen, Al. 2009. "End of the Global War on Terror." *Washington Post*, March 23. http://voices.washingtonpost.com (accessed October 16, 2010).

Koh, Harold Hongju, Legal Adviser, U.S. Department of State. 2010. "The Obama Administration and International Law." Speech at the Annual meeting of the American Society of International Law. March 25. http://www.state.gov/s/l/releases/remarks/139119.htm (accessed October 3, 2010).

Kretzmer, David. 2005. "Targeted Killing of Suspected Terrorists: Extra-Judicial Executions or Legitimate Means of Defence?" *European Journal of International Law* 16, no. 2: 171-212.

Mayer, Jane. 2009. "The Predator War." *The New Yorker*, October 26. http://www.newyorker.com/reporting/2009/10/26/091026fa_fact_mayer (accessed October 10, 2010).

Melzer, Nils. 2008. *Targeted Killing in International Law*. New York: Oxford University Press.

Murphy, Sean. 2002. "Terrorism and the Concept of 'Armed Attack' in Article 51 of the U.N. Charter." *Harvard International Law Journal* 43, no. 1: 41-51.

Obama, Barack. 2009. "President Barack Obama's Inaugural Address." The White House. January 21. http://www.whitehouse.gov/blog/inaugural-address (accessed January 5, 2011).

O'Connell, Mary Ellen. 2010a. "Lawful Use of Combat Drones." Testimony before Congress of the United States, House of Representatives, Subcommittee on National Security and Foreign Affairs Hearing. April 28.

_____. 2010b. "Unlawful Killing with Combat Drones: A Case Study of Pakistan, 2004-2009." Notre Dame Law School Legal Studies Research Paper, no. 09-43. July.

Ofek, Hillel. 2010. "The Tortured Logic of Obama's Drone War." *New Atlantis*, no. 27. http://www.thenewatlantis.com/publications/the-tortured-logic-of-obamas-drone-war (accessed November 11, 2010).

Panetta, Leon. 2009. "Director's Remarks at the Pacific Council on International Policy." Central Intelligence Agency, May 18. https://www.cia.gov/news-information/speeches-testimony/ directors-remarks-at-pacific-council.html (accessed January 5, 2011).

Paust, Jordan J. 2010. "Self-defense Targeting of non-state Actors and Permissibility of U.S. Use of Drones in Pakistan." Draft: April 2. Forthcoming in *Journal of Transnational Law and Policy* 19, no. 2.

Prosecutor v Tädic. [Judgment]. Case No. IT-94-1-A. 15 July 1999.

Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I). 1977. *International Committee of the Red Cross*. June 8. http://www.icrc.org/ihl.nsf/NORM/0A9E03F0F2EE757CC1256402003FB6D2?OpenDocument (accessed November 11, 2010).

Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of Non-International Armed Conflicts (Protocol II). 1977. *International Committee of the Red Cross*, June 8. http://www.icrc.org/ihl.nsf/FULL/475?OpenDocument (accessed November 11, 2010).

Roberts, Adam. 2002. "Counter-terrorism, Armed Force and the Laws of War." *Survival* 44. iss. 1. pp. 7-32.

Romero, Anthony and Vincent Warren. "Sentenced to Death – Without Trial," T*he Washington Post*, September 3, 2010, http://www.lexisnexis.com.ezproxy.library.tufts.edu/hottopics/lnacademic (accessed November 17, 2010).

Sassoli, Marco. 2004. "Use and Abuse of the Laws of War in the 'War on Terrorism.'" *Law and Inequality: A Journal of Theory and Practice* 22, no. 2: 195-222.

Schmitt, Michael. 2004-2005. "Humanitarian Law and Direct Participation in Hostilities by Private Contractors or Civilian Employees." *Chicago Journal of International Law* 5, no. 2: 511-546.

Singer, P.W. 2009. *Wired for War*. New York: The Penguin Press.

Solis, Gary. 2010. "CIA drone attacks produce America's own unlawful combatants." *Washington Post*, March 12. http://www.washingtonpost.com/wp-dyn/content/article/2010/03/11/AR2010031103653.html (accessed October 16, 2010).

United Nations Security Council Resolution 1368. *United Nations.* September 12, 2001, http://www.un.org/terrorism/sc-res.shtml (accessed November 24, 2010).

United Nations Security Council Resolution 1373, *United Nations.* September 28, 2001. http://www.un.org/terrorism/sc-res.shtml (accessed November 24, 2010).

Webster, Daniel. *The Caroline Case.* 1841. "Daniel Webster to Lord Ashburton," April 24. Yale Law School:

The Avalon Project. http://avalon.law.yale.edu/19th_century/br-1842d.asp#web1 (accessed November 12, 2010).

Zenko, Micah. 2009. "Pakistan Strikes are not the Answer." *Council on Foreign Relations*. March 19. http://www.cfr.org/publication/18899/pakistan_strikes_are_not_the_answer.html (accessed October 16, 2010).

5

Understanding State Resistance to International Regulation of Private Military and Security Companies

Adam Ross

Private military and security companies (PMSCs) have exploded into public consciousness in recent years. PMSC involvement in warfare has expanded amidst an absence of regulations governing appropriate conduct, resulting in a climate of impunity. This unprecedented privatization of force also raises concerns about the implications of the "PMSC revolution" on state sovereignty. Despite widespread acknowledgement that international regulation is vital to addressing the PMSC problem, most states have been reluctant to embrace this course of action, thus little progress has been made. This paper argues that an international regulatory regime has not emerged because the client states of PMSCs benefit from an unregulated industry, which lowers the political costs of warfare and provides lucrative opportunities for national elites. Furthermore, the powerful positions of the main client states within the global system ensure that international regulation does not develop without their support. The benefits of an unregulated environment for client states, and the lack

Adam Ross is a graduate of the MSc Global Politics program at the London School of Economics, where he focused on conflict and global governance. He can be reached at sadamross@gmail.com.

of political will to formulate an international regime, lead to the conclusion that the prospects for international regulation of PMSCs are rather bleak.

I. Introduction

Private military and security companies (PMSCs) have come under much public scrutiny with the wars in Afghanistan and Iraq. Much of the media attention and public concern over PMSCs stems from the lack of accountability these actors face for human rights violations, criminal acts, and aggressive behavior in the line of duty. Judicial and accountability structures for military personnel have not adapted to the expanding role of PMSCs in conflict, which has generated interest among academics and legal scholars. This unprecedented privatization of force has also raised concerns amongst political scientists and international relations theorists about its implications for state sovereignty.

The PMSC phenomenon is more than a passing trend. PMSCs are a critical element of states' military and security apparatuses, and a wide range of non-state actors increasingly depend on them. Recognizing their value and intransience, observers have begun to examine how best to regulate PMSCs. The transnational character of many of these firms and the limitations of domestic regulation have led many experts to conclude that the international arena is the most suitable environment for industry regulation (Avant 2005; Leander 2005; Percy 2006). Despite widespread acknowledgement that regulation at the international level is vital to addressing the PMSC problem, most countries have been reluctant to embrace this course of action; as a result, little progress has been made. Yet little attention has been paid to why states are so unwilling.

This paper seeks to explain states' reluctance to embrace international regulation of PMSCs, based on the idea that state inaction is the primary reason why an international regime has not yet emerged. It argues that major client states of PMSCs enjoy a broad range of benefits from an unregulated industry, which would come under threat in a climate of international regulation. The powerful positions of these states in the international system ensure that international regulation cannot develop without their support. Since the negative effects of PMSCs are felt mainly by host states (states in which PMSCs conduct their missions) and not by their clients, there is little incentive for client states to rectify these problems. The majority of states do not feel the direct impact of PMSC activity

and therefore have little incentive to bear the constraints and enforcement costs of an international regulatory regime. Host states are often embroiled in, or recovering from, conflict, and have little sway in the international system; as such, they lack the capacity to effectively regulate PMSCs domestically and cannot champion regulation at the international level. Further undermining efforts to advocate for international regulation are the lucrative opportunities for host state elites to profit from the industry by regulating it domestically. However ineffective domestic regulation may be, it provides an opportunity for host states to assert their sovereignty and enables local elites to take control over the industry.

This paper begins with an overview of the private market for force and the context in which PMSCs have emerged. It then discusses the current regulatory mechanisms in place at the domestic and international levels, and follows with an analysis of the forces that have inhibited the emergence of an international regulatory regime. The paper concludes by assessing the prospects for international regulation and makes recommendations on how best to approach the regulatory problem.

Defining PMSCs

The range of private actors providing military and security services is highly diverse, as are their clients and the environments in which they operate. Here, PMSCs refer to legal corporate entities providing non-combat military and security services in conflict zones. Companies providing military goods are not included in the present discussion, nor are domestic security firms unless they are operating within conflict situations. Although some observers differentiate between private military firms and private security firms, no such distinction is made here because of the significant overlap in the services provided by individual firms and a general blurring of roles in conflict situations.

Although PMSCs are often equated with "corporate mercenaries," such a simplistic comparison fails to take into account fundamental differences that are important to a thorough analysis of existing legal mechanisms governing PMSC's use and conduct. Article 47 of the first protocol to the Geneva Conventions, which provides the most agreed upon definition of *mercenary*, excludes PMSC personnel on a number of grounds (Avant 2005). The corporate structure and transnational operations of PMSCs and their abstention from direct combat activities are key factors that differentiate PMSCs from mercenaries, necessitating a separate approach to defining and regulating this unique segment of the private market for force.

II. Background

The private market for force is not new (Avant 2005). Prior to the 1648 Peace of Westphalia, organized force was allocated solely through the market. Only since the formation of the modern state has legitimate control over organized violence become the prerogative of governments (Tilly 1990). This understanding has shaped the contemporary international system, and is reflected in the laws and norms shaping interstate behavior. The Geneva Conventions governing modern warfare, for instance, deem the use of violence by non-state actors unlawful (De Nevers 2009). As the Westphalian system has evolved, however, so has the private market for force.

In the contemporary interstate system, states have continued to embrace the utility of private force despite its incompatibility with interstate norms and international legal obligations. During the colonial period, European states delegated control of force to chartered companies, privateers, and other non-state actors in order to expand and protect their empires. The Cold War powers used mercenaries in many of their covert interventions and proxy wars, particularly in Africa (Musah and Fayemi 2000, 17). Mercenaries were also involved in a number of coups, rebellions, and assassinations during decolonization, causing uproar among former colonial governments and populations. Mercenaries came to be seen as "agent[s] of the colonial powers ... Symbol[s] of racism and opposition to self-determination" (Shearer 1998, 15). As independence movements progressed, a number of international and regional frameworks were created with a view to prohibiting and criminalizing mercenary activity. Yet despite the growing public and institutional opposition to mercenaries, the end of the Cold War led to a vast expansion of the market for private force.

The end of the Cold War marked a shift away from the nationalization of organized violence, with structural and market conditions combining to usher in a new era of privatized force. Rapid military downsizing created vast pools of unemployed soldiers and cheap weaponry, and the eruption of conflicts in former "third world" states provided the necessary demand. The decisive triumph of free-market ideology also began to dilute the norms legitimizing the state's monopoly on organized violence. As governments outsourced more functions it became less outlandish to consider doing so in the military domain (Singer 2003, 66).

Emphasizing their corporate structure, professionalism, and legitimacy (i.e. only working for internationally recognized governments), the PMSCs that emerged in the wake of the Cold War sought to distance themselves

from the "rogue, individual, free-floating, and ultimately unreliable contemporary mercenaries" of the Cold War era (Leander 2005, 608). Initial reactions to these companies were largely positive, with many observers highlighting their potential for peacekeeping (Bures 2005), cost savings (Barber 2000), and even conflict resolution (Shearer 1998). Yet government and public opinion soured in the late 1990s with scandals such as the "Arms to Africa" debacle and the "Sandline Affair." As the provision of private combat services once again lost its legitimacy with governments and the public, PMSCs concentrated on other core competencies.

Though combat services are no longer readily available on the market, supply and demand for other services have grown. At the turn of the millennium, PMSCs were active in nearly every conflict and peacekeeping mission worldwide. Demand for PMSCs has further intensified in the context of the "War on Terror," as states have used them to buttress overstretched militaries in Afghanistan and Iraq. Although reliable data on the size of the industry does not exist, by all accounts it has continued to experience rapid growth. With the downscaling of foreign forces in Afghanistan and Iraq, PMSCs are likely to replace military personnel; this process is indeed already underway, with PMSC personnel now outnumbering military personnel in both Afghanistan and Iraq (Scahill 2010; HSRP 2010).

There has also been a notable expansion in the breadth of PMSC services and in the clientele that hire them. Contemporary PMSCs provide a wide range of services from logistical and operational support to military training to humanitarian assistance and other non-military services. During the Cold War, demand for combat services was driven predominantly by governments of weak states and groups looking to overthrow them (Musah and Fayemi 2000). Demand for support services was exclusive to western industrialized states, which were the only countries with the capacity and political will to contract nonessential services to the private sector. Today, not only do both weak and strong states feature among PMSCs' major clients, but a range of non-state actors have turned to the private sector to meet their security needs. Humanitarian organizations such as World Vision and CARE International employ PMSCs to provide security for their personnel and to deliver humanitarian aid (Singer 2006, 8). PMSCs are also heavily employed by the private sector, as companies look to the market to meet security needs that cannot, or will not, be met by the state. Perhaps the largest non-state client of all is the UN; despite institutional opposition to PMSCs, the UN has used them in every peacekeeping mission since 1990 (Avant 2005, 7).

III. The Current Regulatory Environment

There is a prevailing misconception that PMSCs operate in a legal and regulatory vacuum (Percy 2006, 57). This stems partially from the fact that no international legal mechanisms specifically address PMSCs, and partly from states' failures to hold PMSCs accountable under existing domestic and international laws. The few states that have attempted to regulate PMSCs domestically have done so with limited success. The transnational nature of the industry and the extraterritorial operations of PMSCs pose formidable challenges to regulation at the domestic level. Internationally, regulation has been impeded by ambiguity surrounding the application of existing laws to PMSCs, and by states' disagreement over the appropriate scope of additional regulatory mechanisms.

Domestic Regulation

Regulation at the state level has a number of strengths. Domestic legislation has the greatest degree of enforceability, as even international law must be implemented at the domestic level. Domestic regulation reaffirms the centrality of the state to international security issues and strengthens state control over organized force. It may also serve political interests by pacifying public opposition to an unregulated PMSC industry. Yet on its own, domestic regulation is problematic because of the transnational nature of the industry and its firms. Inconsistent domestic regulations allow PMSCs to relocate to states with less cumbersome regulatory environments, and may encourage some states to create regulatory havens to attract foreign investment (Gaston 2008, 241). The extraterritorial operations of PMSCs also pose difficulties for enforcement of domestic laws and regulations, both in terms of jurisdiction and the practical challenges of monitoring and conducting investigations in far-off conflict zones. The following sections examine the regulatory environments in the United States, the UK, Iraq, and Afghanistan, providing examples of the various approaches to regulation as well as the challenges faced by host states with limited regulatory capacity.

United States – In the United States, domestic legislation was developed by extending laws on arms exports to the export of military services. Under the Arms Export Control Act of 1976, the State Department administers and enforces a licensing system through the International Traffic in Arms Regulation (ITAR). Under ITAR, certain states are prohibited from soliciting PMSC services (Percy 2006, 26). Critics note that the system suffers from procedural defects and a lack of resources. Multiple offices grant

licenses with minimal coordination, and contract administrators within the State Department — who were in short supply before the Afghanistan and Iraq wars — have not been expanded despite the additional workload (Holmqvist 2005, 51; Percy 2006, 27). The oversight mechanisms under ITAR are also weak. Contracts of $50 million or more are subject to congressional review, but this rule has been circumvented by drafting multiple lower-valued contracts for larger deals (Avant 2005, 151). Once a license is granted for a particular contract, there are no follow-up mechanisms in place to ensure appropriate delivery. Furthermore, ITAR only applies to contracts between American firms and foreign clients, and not to contracts with the U.S. government.

Another piece of legislation, the Military Extraterritorial Jurisdiction Act of 2000 and MEJA Expansion and Enforcement Act of 2007 (MEJA, collectively) applies U.S. criminal law to PMSC personnel under contract with U.S. government agencies abroad. However, MEJA's jurisdiction may not extend to subcontracted personnel, thereby excluding a great number of cases. Military justice may also apply to PMSC personnel. Simon Chesterman suggests that in certain cases the amended Uniform Code of Military Justice (UCMJ) extends to civilian contractors, as it applies to all military personnel in conflict situations (Chesterman 2008, 41). Yet PMSC personnel do not fall under UCMJ jurisdiction unless they are officially integrated into the armed forces, which they rarely are. Other federal legislation, including the U.S. War Crimes Act of 1996 and the anti-torture statute, may also apply to PMSC employees, but under limited circumstances.

Despite this web of applicable legislation, the challenges and costs of conducting investigations extraterritorially mean that only the most egregious and highly publicized cases have been prosecuted. These available legal mechanisms for regulating PMSCs have gone almost entirely unimplemented. The 2006 conviction of CIA contractor David Passaro for beating to death a detainee at Iraq's Abu Ghraib prison is, to date, the only successful prosecution of a PMSC employee for abuses in the line of duty.

United Kingdom – In contrast to the United States, there are currently no formal oversight mechanisms to regulate contracts between UK PMSCs and other states or organizations (Percy 2009, 70), and there are no prosecutions or criminal investigations on record. Following the "Sandline Affair" and the "Arms to Africa" scandals, the British government began to consider domestic legislation and regulatory options. In 2002, the Foreign and Commonwealth Office (FCO) produced a Green Paper,

Private Military Companies: Options for Regulation, which recommended adopting a regulatory system of general and contractual licensing similar to ITAR, as well as an independent body to monitor UK firms. Despite a positive response from the government and British PMSC industry, the recommendations have not been adopted or subjected to parliamentary debate. Were they to be adopted, each of the proposed options would face oversight and control problems similar to those that plague the American regulatory mechanisms (Holmqvist 2005, 54).

Iraq and Afghanistan – Host states face different challenges in regulating PMSCs within their borders. They are often embroiled in, or recovering from, conflict and lack the capacity to develop or enforce laws and regulations for PMSCs. This limited capacity is often further undermined by the policies and behavior of client states.

The experience of the Iraqi government is demonstrative of the impotence of host states to effectively control international PMSCs. Following the 2007 Nisoor Square incident, where Blackwater personnel killed seventeen Iraqi civilians during a convoy protection mission, the Iraqi government attempted to prosecute those involved. It was unable to do so because of Coalition Provisional Authority (CPA) Order 17, which granted PMSCs and their personnel immunity from Iraqi prosecution. Order 17 was repealed in one of the first acts of the Iraqi Parliament in 2009, and the Ministry of Interior used its expanded authority to refuse Blackwater an operating license. Despite decisive action by the Iraqi authorities, the company (now called Xe Services) continues to operate in Iraq and is still the main provider of personal security services for the U.S. State Department (Partlow 2007).

The problems experienced by the Iraqi authorities are not unique. PMSC employees are often immune from local prosecution under Status of Forces Agreements (SoFAs). To date, SoFAs have not been designed with PMSCs in mind. Rather, they are intended to safeguard military personnel from "problematic prosecution in states with weak or nonexistent judicial systems" (Percy 2009, 60). When applied to PMSC personnel, this immunity has not been matched by robust legal mechanisms within the United States, creating a significant accountability gap. As Sarah Percy notes, "the wisdom of making contractors immune from local prosecution was never questioned, even in the absence of instruments that would allow them to be prosecuted in the US" (2009, 61).

Where domestic regulation exists, it tends to be ad hoc, uncoordinated, and incomprehensive, with significant gaps in the legal frameworks and implementation measures. Norms are inconsistent, as states are more

inclined to facilitate and tolerate PMSC activity abroad than within their own borders. The inconsistencies of domestic regulation demonstrate the lack of convergence among states in determining which rules should apply to the PMSC industry.

International Law

Though PMSC personnel are covered under international humanitarian law (IHL) and international human rights law (IHRL), there have been no prosecutions of PMSC personnel for violations of IHL or IHRL, and their rights and obligations remain ambiguous. A significant step toward international regulation was taken recently with the drafting of a United Nations (UN) convention to regulate PMSCs, but states have been unable to agree upon key elements of the treaty. Without drastic changes it is unlikely to be codified into law.

While there is some consensus that PMSC personnel do not qualify as lawful combatants under IHL, their status as civilians is disputed (Montreux Document 2008, 14). Civilians cannot lawfully take part in combat and, in turn, are guaranteed protection. Combatants, on the other hand, are considered valid military targets and may participate in combat, but are obliged to follow a code of conduct. The ambiguous nature of PMSC personnel's participation in conflicts is problematic in determining which laws apply. Clarifying the status of PMSC personnel is vital to evaluating the legality of their actions.

Despite this ambiguity, if IHL violations do occur, states have an obligation to exercise due diligence to "prevent violations, to investigate violations and punish perpetrators, and to provide victims with access to justice and effective remedies" (Tonkin 2009, 785). This obligation extends to private actors but is hindered when links between PMSCs and the state are limited, including cases of sub-contracting and of contracts with non-state clients. The often-tenuous links between PMSCs and state militaries, and governments' predilection to avoid tacit admission of responsibility for IHL violations mean abuses by PMSC personnel go unprosecuted.

PMSCs are also obliged to uphold the human rights standards enshrined in the treaties to which their home state is party, though, as with IHL, responsibility ultimately lies with the state (De Nevers 2009, 186). One international enforcement mechanism that can be applied to PMSCs directly is the Rome Statute and the International Criminal Court (ICC). The ICC has jurisdiction to prosecute PMSC personnel accused of genocide, war crimes, and crimes against humanity. However, most violations of international law do not qualify. Moreover, the ICC only has jurisdiction

in states party to the Rome Statute (which excludes the United States), and can only prosecute individuals, not corporate entities.

Despite the problems of the mercenary regime and its inapplicability to PMSCs, the UN has addressed PMSCs and mercenaries as two sides of the same coin. The name of the body mandated to address PMSCs, the Working Group on the use of mercenaries as a means of violating human rights and impeding the exercise of the right of peoples to self-determination, provides an indication of the lens through which the UN sees the industry and the context in which it was created. The Working Group recently admitted that the mercenary definition does not apply to most PMSC personnel and has softened its prohibitionist stance. With a view toward legal clarification, greater accountability, and reinforcing the state monopoly on force, in 2009 the Working Group proposed a *draft International Convention on the regulation of activities of Private Military and Security Companies*.

The draft convention is the outcome of a consultative process involving a range of experts, industry representatives and states, and has been heralded as a major step toward international regulation. The practical elements of the draft include a compulsory licensing system, an international registry of PMSCs, and an independent monitoring body. While these measures enjoy the support of many states, the implementation costs are a matter of concern (UN 2010). Other elements are more problematic for many states. The convention reaffirms state responsibility for the activities of PMSCs in their jurisdiction, making states liable for violations of international law by PMSC personnel; this is an issue of particular concern for states that use PMSCs in military operations abroad. States would also be required to amend domestic law to meet the prescriptions of the convention, impinging upon their internal sovereignty. Another contentious issue has been the prohibition of "outsourcing of inherently state functions," which would impose severe restrictions on states' use of PMSCs (UN 2010).

States are reluctant to accept external restrictions of their security options. In addition, there is little agreement over what are "inherently state functions." For the Working Group, these functions include: direct participation in hostilities, combat operations, arrest and detention, lawmaking, espionage, intelligence, interrogation, knowledge transfer with military, security and policing application, and activities related to weapons of mass destruction (UN 2010). In the United States, a 2010 defense funding bill included a provision barring the Department of Defense (DoD) from contracting out interrogation services, which it cited as an "inherently governmental function" (De Young 2009). The provision was pulled

after President Obama threatened to veto the bill. It is unlikely that the Working Group definition enjoys consensus within the UN itself. Many departments, including UN Peacekeeping, use PMSCs extensively and may perceive the definition as overly restrictive. As the debate over "inherently state functions" shows, the international community is a long way from forming the common principles and norms necessary for an international regime governing use of PMSCs.

IV. State Resistance to International Regulation

The absence of a regulatory regime for PMSCs can be explained in part by the value of an unregulated industry to their main client states, and by the positions of power that these states hold in the international system. The majority of states, with the exception of host states, do not see an unregulated industry as a security threat and are thus unconcerned with improving regulation. The few states that *are* adversely impacted by PMSC activity lack the power to propel international cooperation and the resources to support it. Moreover, links between the industry and government elites in such states suggest there may be less political will for international regulation than is publicly professed.

For client states of PMSCs, the existing regulatory environment provides strategic manoeuvrability that could be threatened by an international regulatory regime. Using PMSCs in an unregulated environment enables states to get around political and resource constraints that would otherwise limit the use of force (Gaston 2008). Regulation would also increase administrative costs and threaten the informal networks that exist between PMSCs and state elites.

By contracting to PMSCs, the state is able to bypass democratic oversight and decision-making mechanisms usually required for military and security matters. The War Powers Clause of the U.S. Constitution and the 1973 War Powers Resolution are intended as democratic balancing mechanisms in situations that risk the involvement of the United States in war. Under the War Powers Resolution, the president must obtain the consent of congress when sending U.S. armed forces into hostilities, and must regularly report on developments of such operations. Using PMSCs provides the executive with greater autonomy and manoeuvrability, as disclosure requirements apply only to the armed forces. None of the 105 "war powers" reports submitted to the U.S. Congress in the last twenty years have disclosed any information regarding the significant involvement of PMSCs in military operations (Grimmett 2010). While military and

security budgets require congressional authorization, funds earmarked for PMSC contracts do not. U.S. intelligence agencies have also been able to expand operations without congressional approval by employing the services of PMSCs (Smith 2009b).

By using PMSCs, the state can make war more palatable to its electorate. Military casualties often bear significant political costs, particularly in conflicts that are not perceived as necessary to the national interest. The political costs of "unnecessary" casualties are widely documented. For instance, following the highly publicized killings of American soldiers during the U.S. intervention in Somalia, President Clinton's approval ratings fell to the lowest of his presidency (Carr 2003, 15). PMSC casualties do not rouse the same media attention and public reaction as military casualties. The official military death toll for the war in Iraq is currently 4,422 (DoD 2011), but this figure does not include an estimated 1,487 contractor deaths (Schooner 2010). In an unregulated environment, the use of PMSCs enables the state to keep casualty figures artificially low, obscuring the true human cost of the war and positively impacting public opinion. Furthermore, the extensive use of PMSCs has enabled the military to avoid additional expansion, which tends to be unpopular among voters.

At the international level, an unregulated PMSC industry endows states with foreign policy options that would otherwise be unavailable. In an interstate system predicated on the norm of non-interference, states cannot openly intervene in the affairs of other sovereign states without consequence. PMSCs provide some flexibility in this regard, as the connections between clients and contractors are often obscured. This was the case with U.S. support for the Croatian military during the first Balkan conflict. To avoid jeopardizing its role as mediator by offering direct military support to the Croats, the United States contracted a little-known PMSC to train the Croat forces. Some observers argue that this move changed the course of the war and paved the way for the Dayton Peace Agreement (Chesterman 2008, 40). Covert use of PMSCs may run considerably deeper than this. The CIA recently confirmed that it had outsourced a "kill or capture" programme in Pakistan to Blackwater, although it was never carried out (Smith 2009a). Such controversial activities would likely be proscribed by international regulation.

The legal ambiguity surrounding PMSCs and their personnel allows states to shirk certain international obligations. Current practices of subcontracting and the lack of clarity over command and control structures mean that PMSCs are rarely considered part of a state's armed forces under IHL (Montreux Document 2008). This enables governments to

claim ignorance of violations and avoid responsibility for abuses by PMSC personnel. As E.L. Gaston notes, "states therefore have few incentives to develop accountability and control mechanisms that would prevent or redress the type of misconduct and international law violations that have been associated with private military and security companies in Iraq and Afghanistan," (Gaston 2008, 222).

On the financial front, an unregulated industry is likely to be less costly than a regulated one. Such factors as independent oversight mechanisms, burdensome contractual procedures, and licensing administration would increase the costs of government use of PMSCs. The fact that the number of ITAR contract administrators did not increase despite the wars in Afghanistan and Iraq demonstrates the reluctance within the U.S. government to increase administrative costs.

A regulated environment may also include safeguards against conflicts of interest and facilitate more transparent competition between PMSCs, which would constrain the informal network between the industry and the state. Peter Singer brings to attention the "revolving door syndrome" between the state and the military and security sector (Singer 2003, 154). A 2008 Government Accountability Office (GAO) investigation found that over a two-year period, 2,435 Pentagon officials had taken positions with PMSCs under contract with the DoD (GAO 2008). The GAO recently reversed a Pentagon decision to award a $1 billion contract to Xe/Blackwater to train the Afghan police after a rival firm complained that the process excluded others from bidding (Warrick 2010). There were 611 such protests in 2008 alone, to which the DoD's response was to advocate curbing the protest rights of contractors (Capaccio 2009). Although some mechanisms are in place to ensure competitive bidding, there is a clear pattern of partisanship when it comes to Pentagon contracting procedures (Hedgepeth 2008).

Many Afghan and Iraqi politicians are vocal critics of PMSC involvement in their countries, but this may stem more from a lack of control and ownership of foreign PMSCs than from objections to private force *per se*. There will be considerable demand for PMSCs in Afghanistan and Iraq for some time; it is in the interest of these governments to ensure they benefit from the industry. Elites in other post-conflict states have profited significantly from the private provision of security. Afghan and Iraqi elites have recognized this potential and are creating barriers for international PMSCs in favour of domestic firms. In 2010, President Karzai passed a presidential decree banning foreign PMSCs from operating in the country, vastly increasing demand for firms controlled by the Afghan elite (Scahill

2010). Though law prohibits high-ranking government officials and their relatives from owning PMSCs, a license was recently granted to a firm owned by the son of the current Defence Minister, and to another firm owned in part by cousins of President Karzai (Roston 2009).

States that cannot depend upon their own military and police forces to consolidate control and ensure their own survival may also benefit from an unregulated PMSC industry. Despite their ostensible opposition to private force, a number of African governments have employed PMSCs to suppress insurgencies or other internal threats (Gumedze 2009). Though a norm exists against private provision of combat services, under an international regime it would be formally prohibited and thus officially unavailable to threatened governments.

An unregulated PMSC industry does have certain drawbacks for states. Without adequate oversight mechanisms PMSCs may undermine national interest. While some PMSCs align themselves with client state interests in order to obtain government contracts and maintain the *status quo*, historically this is not always the case. The "Arms to Africa" scandal occurred because Sandline International had not aligned itself with British interests, causing the UK to unwittingly breach a UN arms embargo in Sierra Leone. While the UK has not taken concrete steps to prevent similar conflicts of interest from occurring, the U.S. ITAR licensing system was designed with such strategic alignment in mind.

Concerns have also been raised over PMSCs undermining military strategy and effectiveness, in light of experiences in Afghanistan and Iraq. U.S. Secretary of Defense Robert Gates expressed concern that the methods used by PMSCs in providing security details often "work at cross purposes to our larger mission" (DoD 2007). Short-term, outcome-oriented contracts include no provisions to deter PMSCs from employing heavy-handed tactics in carrying out their missions. Contractor abuses not only engender ill will toward PMSCs but also toward the military at large, as local populations often do not distinguish between PMSC and military personnel (Fainaru 2007). Such abusive or aggressive behavior generates antipathy among the local population, undermining the Pentagon's broader "hearts and minds" counterinsurgency strategy.

Additionally, military commanders do not have direct control over PMSC personnel operating alongside them. The U.S. military has "no specifically identified force structure nor detailed policy on how to establish contractor management oversight" (U.S. Army, 2003: 1/24). Contractors are rarely required to inform the military of their actions or positions, or to share other pertinent information with military command or among

each other. The rules of engagement for PMSCs are unclear and often poorly understood by personnel (Avant 2006, 337). PMSCs' concern with limiting operating costs may lead to cutting corners (e.g. inadequate training and vetting of personnel), thus exacerbating the problem.

Where PMSCs are a significant part of a state's security strategy, as in the United States, domestic mechanisms have been developed to mitigate potential threats to state authority and national interest. Potential drawbacks, such as those described above, are a small price to pay for the capacity to enter into those wars in the first place. Now that these risks to military effectiveness have been identified, the United States is acting to remedy them. It remains in the interest of client states to improve the effectiveness of PMSCs rather than imposing additional constraints upon their use. For client states, preventing abuses by PMSC personnel is an auxiliary concern, as the negative effects are felt disproportionately by host states and their populations.

The United States reaps the greatest benefits from an unregulated industry and would bear the greatest costs of international regulation. Like any other state, the United States will resist an international regime contrary to its interests. An international PMSC regime depends upon American participation because of the market dominance of U.S. PMSCs and the magnitude of U.S. government contracts. Regime enforcement and monitoring would be impossible without American backing, as the majority of both supply *and* demand for PMSC services comes from the United States.

In contrast to client states, for which PMSCs are a key strategic asset, host states may see PMSCs as a source of insecurity. In Afghanistan, this is confirmed by the negative perceptions held by both the local population and the state (Schmeidl 2007; Scahill 2010). The Iraqi and Afghan governments have attempted to impose more stringent domestic regulation upon PMSCs, but have been unable to do so effectively. Given the detrimental effects of unregulated PMSC activity in these states and their inability to effectively implement domestic regulations, it may be in their interest to pursue international regulation. However, their inability to exert control over their domestic environments is matched by their weak positions within the international system. Even if these states wished to champion an international regulatory regime, they would lack the necessary leverage to do so.

Only when states cannot address a transnational problem through independent policies will they consider an international regime (Keohane 1983, 141-171). Both client states and host states prefer domestic solutions

to the problems of unregulated PMSC activity, as domestic solutions do not jeopardize sovereign authority or place unwanted limits on the political and economic benefits of PMSCs for national elites. Demonstrating the flexibility of domestic regulation, the U.S. government recently fined Xe/Blackwater $42 million for breach of ITAR restrictions. By fining the firm instead of initiating criminal proceedings, the government is able to retain the firm's services and award it future contracts (Hodge 2010). Even in the face of limited enforcement capacity, as with Afghanistan and Iraq, states have demonstrated a preference for the manoeuvrability and sovereignty afforded by domestic actions.

Most states do not perceive PMSC regulation as a collective problem. The majority of states are not impacted by an unregulated industry; PMSCs are not a key part of their security strategies, nor do PMSCs operate within their borders. This is reflected in the scant attention domestic lawmakers have paid to the industry. International regulation will not provide these states with security gains, relative or absolute, and for them the *status quo* is acceptable.

States that foresee relying upon PMSCs for their own security are unlikely to limit their options by adopting the constraints of an international regime. This applies not only to weak or threatened states, but also to stronger states with small militaries and limited security infrastructure. Enforcement costs associated with the creation of an international regime also act as a deterrent to participation.

While non-state actors can be influential players in regime formation, they are unlikely to be a driving force behind a PMSC regime. Civil society advocates have been instrumental in generating public concern about PMSCs, but they have historically had little influence when it comes to security regimes (Meierding 2005). Industry players have also demonstrated limited influence on regulation, as evidenced by their unsuccessful calls for greater government oversight in the UK (Bearpark 2007). While the UN has thus far been central to regulatory discussions, it depends upon the support of member states, which are the signatories and enforcers of international law.

V. Conclusions and Recommendations

State reliance upon PMSCs will continue to grow. The U.S. military and security strategy is highly dependent upon contractors and could not function without them. If the United States were to desist from using PMSCs, massive military expansion would be necessary to maintain current force capabilities. Many other states that have embraced privatization of military

functions would also face political and resource constraints in a climate of international regulation. For the United States and its allies, military disengagement from Afghanistan and Iraq depends upon a civilian force capable of stabilizing and rebuilding two collapsed states, or at the very least preventing a return to outright war. Domestic PMSCs are quickly evolving to fill the "security gaps" in both countries, but transnational PMSCs are likely to remain a dominant presence in the years to come. Likewise, U.S. dominance of the industry is likely to continue.

If effective PMSC regulation is to evolve, the United States must be central to the process. As the largest supplier and source of demand, the United States has the ability to reshape the industry, whether through domestic actions or leading international regulation. Given the clear preference of the United States for domestic solutions, advocates of greater PMSC regulation would be prudent to make domestic regulation a priority and lobby the American government to lead on best practices for industry regulation. This could start by improving upon existing legislation, such as closing the loophole in ITAR that allows for contracts of less than $50 million to avoid congressional review, or by investing in greater resources for monitoring and information sharing.

Although abuses by PMSC personnel raise serious human rights concerns, states are primarily concerned with the negative impact PMSC abuses have on their security interests. As Emily Meierding observes, for states "moral arguments…appear irrelevant and easy to dismiss when confronted with core security concerns" (Meierding 2005, 10). Current steps toward international regulation have come from within the UN human rights framework and do not adequately address security issues. Normative appeals to join a regime are unlikely to work if states do not perceive security gains. Only if the PMSC problem is addressed as a security issue will questions regarding the role of private actors as providers of organized force and the nature of "inherently state functions" move to the forefront of the debate.

In order to be viable, international regulation needs the support of third party states with the capacity to bear enforcement costs and increase costs of non-compliance for powerful client states. One way to garner the attention of third party states may be through labor and migration issues, as PMSCs now draw personnel from states all over the world. Among the security personnel in Iraq and Afghanistan are nationals of third party states such as Chile, Fiji, and Uganda, who are attractive to PMSCs because of their extensive training and relatively low wage requirements. For states whose nationals are increasingly involved in the industry, clarifying the

international legal standing of PMSC personnel in conflict and ensuring adequate standards of employment are likely to become issues of growing concern.

Another means of appealing to third party states may be through emphasizing the risk PMSCs pose to the legitimacy of international law. As Gaston notes, the inability of states to adapt the laws of war to reflect the expanding role of PMSCs threatens to undermine the effectiveness and validity of international laws and norms governing the use of force (Gaston 2008, 248). The status of PMSC personnel under IHL is in desperate need of clarification. The International Committee of the Red Cross (ICRC) has already taken steps toward this in drafting the Montreux document, but further clarification is still required. By championing further discussions the UN could move away from the current human rights approach and reassert itself as a central actor in future discussions on the private use of force.

For advocates of international PMSC regulation, the prospects are somewhat bleak. Despite the transnational character of many PMSCs, most states do not experience any ill effects from an unregulated industry. Client states have strong disincentives to developing an international regulatory regime. The few states that would benefit from such a regime lack the collective power to implement it and may see greater returns if they are able to control the PMSC problem domestically. An international regime would require ceding a degree of control and constraining their freedom to make decisions about their own security. Although it is clear that unconstrained PMSC activity can have harmful consequences for states, under an internationally regulated environment, PMSCs are likely to lose the Machiavellian maneuverability that has made them such a popular tool of governments and a target of political activists (Percy 2009, 71). A pragmatic approach to regulation, focusing on domestic policy and the revision of international law to accommodate these new actors, would address states' concerns over security and sovereignty and would be a move toward ending the impunity that has characterized PMSC involvement in conflict.

References

Avant, Deborah. 2005. *The Market for Force: The Consequences of Privatizing Security*. Cambridge: Cambridge University Press.

—. 2006. The Privatization of Security: Lessons from Iraq, *Orbis*, Spring 2006: 327-342.

Barber, Simon. 2000. Bring Executive Outcomes Back to Fight in Sierra Leone, *Business Day*, 12/05. http://allafrica.com/stories/200005100283.html (accessed August 10, 2010).

Bearpark, A. 2007. The Regulation of the Private Security Industry and the Future of the Market, in Chesterman S. and Lehnardt, C. eds., *From Mercenaries to Market: The Rise and Regulation of Private Military Companies*, Oxford: Oxford University Press.

Bures, Oldrich. 2005. Private Military Companies: A Second Best Peacekeeping Option? *International Peacekeeping,* 12(4): 533-546.

Capaccio, Tom. 2009. Pentagon Reviews Contract Protests After 24% Spike, *Bloomberg*, 17/11, http://www.bloomberg.com/apps/news?pid=newsarchive &sid=afQlx05IYgdI (accessed August 20, 2010).

Carr, Damian P. 2003. Military Intervention during the Clinton Administration: A Critical Comparison, *US Army War College,*
http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf &AD=ADA414933
(accessed August 26, 2010).

Chesterman, Simon. 2008. Leashing the Dogs of War: The Rise of Private Military and Security Companies, *Public Law & Legal Theory Working Paper No. 08-24*. New York: New York University School of Law.

De Nevers, Renée. 2009. Private Security and the Laws of War, *Security Dialogue,* 40(2): 169-190.

De Young, Karen. 2009. Administration Objects to Bar on Using Contractors to Conduct Interrogations, *Washington Post*, 16/07,
http://www.washingtonpost.com/wp-dyn/content/article/2009/07/15/AR2009071503415.html
(accessed August 16, 2010).

Department of Defense (DoD). 2007. *News Transcript, 18/10*,
www.defenselink.mil/transcripts/transcript.aspx?transcriptid=4064 (accessed 14/07/2010).

—. 2011. *Casualty Status Report, 14/01/2011*. http://www.defense.gov/news/casualty.pdf (accessed January 15, 2011).

Fainaru, Steve. 2007. Where Military Rules Don't Apply: Blackwater's Security Force in Iraq Give Wide Latitude by State Dept, *Washington Post*, 20/09,
http://www.washingtonpost.com/wp-dyn/content/article/2007/09/19/AR2007091902503.html
(accessed August 10, 2010).

Foreign and Commonwealth Office (FCO). 2002. *HC 577 Private Military Companies: Options for Regulation*, London: The Stationary Office.

Gaston, E.L. 2008. Mercenarism 2.0? The Rise of the Modern Private Security Industry and its Implications for International Humanitarian Law Enforcement, *Harvard International Law Journal,* 49(1): 221-248.

Government Accountability Office (GAO). 2008. *Defense Contracting: Post-Government Employment of Former DOD Officials Needs Greater Transparency*, http://www.gao.gov/new.items/d08485.pdf (accessed August 18, 2010).

Grimmett, Richard F. 2010. The War Powers Resolution: After Thirty-Six Years, *Congressional Research Service*, 22/04, http://www.fas.org/sgp/crs/natsec/R41199.pdf (accessed August 20, 2010).

Gumedze, Sabelo. 2009. Addressing the use of private security and military companies at the international level, *Institute for Security Studies paper 206.*

Hedgepeth, Dana. 2008. Pentagon Auditors Pressured To Favor Contractors, GAO Says, *Washington Post*, 24/07,

http://www.washingtonpost.com/wp-dyn/content/article/2008/07/23/AR2008072301437.html

(accessed August 20, 2010).

Holmqvist, Caroline. 2005. *Private Security Companies: The Case for Regulation*. SIPRI Policy Paper No. 9.

Human Security Report Project (HSRP). 2010. *Afghanistan Conflict Monitor*, http://www.afghanconflictmonitor.org/securityforces.html (accessed July 10, 2010).

Keohane, Robert O. 1983. The demand for international regimes, in Krasner, S.D. ed. *International Regimes*, Cambridge: Cornell University Press.

Leander, Anna. 2005. The Market for Force and Public Security: The Destabilizing Consequences of Private Military Companies, *Journal of Peace Research,* 42(5): 605-622.

Meierding, Emily. 2005. Missing the Target: Light Weapons and the Limits of Global Governance*, 2005 annual meeting of the International Studies Association, Honolulu*, http://www.allacademic.com/meta/p70438_index.html (accessed August 25, 2010).

Montreux Document on Pertinent International Legal Obligations and Good Practices for States Related to Operations of Private Military and Security Companies During Armed Conflict. 2008.

Musah, Abdel-Fatau and Kayode Fayemi. 2000. *Mercenaries: An African Security Dilemma*, London: Pluto.

Partlow, Joshua. 2007. Iraq Bans Security Contractor, *Washington Post*, 18/09, http://www.washingtonpost.com/wp-dyn/content/article/2007/09/17/AR2007091700238.html (accessed August 14, 2010).

Percy, Sarah. 2006. *Adelphi Paper 384: Regulating the Private Security Industry* Oxon*:* Routledge.

— . 2009. Private Security Companies and Civil Wars, *Civil Wars*, 11(1): 57-74.

Roston, Aram. 2009. How the US funds the Taliban. *The Nation.* 11/11. www.thenation.com/doc/20091130/roston (accessed August 27, 2010).

Scahill, Jeremy. 2010. Iraq Withdrawal? Obama and Clinton Expanding US Paramilitary Force in Iraq. *The Nation*. 22/07. http://www.thenation.com/

blog/37877/iraq-withdrawal-obama-and-clinton-expanding-us -paramilitary-force-iraq (accessed August 12, 2010).

Schmeidl, Susanne. 2007. Case study Afghanistan – Who guards the guards? *Private Security Companies and Local Populations: An Exploratory Study of Afghanistan and Angola*, Swisspeace report, 1155(331): 14-45.

Schooner, Steven and Colin D. Swan. 2010. Contractors and the Ultimate Sacrifice. *Service Contractor*, September, http://www.pscouncil.org/Content/NavigationMenu/PublicationsServiceContractorMagazine/SC_SEPT2010_Web.pdf (accessed March 10, 2011).

Shearer, David. 1998. *Adelphi Paper 316: Private Armies and Military Intervention.* Oxford: Oxford University Press.

Singer, Peter W. 2003. *Corporate Warriors: The Rise of the Privatized Military Industry.* Ithaca: Cornell University Press.

—. 2006. Humanitarian principles, private military agents: implications of the privatized military industry for the humanitarian community, in Wheeler, V. and Harmer A. eds., *Resetting the Rules of Engagement: Trends and Issues in Military–Humanitarian Relations*. London: ODI.

Smith, Jeffrey. 2009a. Sources Say CIA Hired Blackwater for Assassin Program, *Washington Post*, 20/08, http://www.washingtonpost.com/wp-dyn/content/article/2009/08/19 AR2009081904315.html?sid=ST2009082004397 (accessed August 18, 2010).

—. 2009b. Disclosure of 'Targeted Killing' Program Comes at Bad Time for CIA, *Washington Post*, 21/08, http://www.washingtonpost.com/wpdyn/content/article/2009/08/20/AR2009082004064.html

(accessed August 24, 2010).

Tilly, Charles. 1990. *Coercion, Capital, and European States, AD 990-1992*. Malden: Blackwell Publishers.

Tonkin, Hannah. 2009. Common Article I: A Minimum Yardstick for Regulating Private Military and Security Companies, *Leiden Journal of International Law*, 22(2009): 779-799.

UN. 2010. *A/HRC/15/25: Report of the Working Group on the use of mercenaries as a means of violating human rights and impeding the right of peoples to self-determination.*

U.S. Army. 2003. *Field Manual No.3-100.21: Contractors on the Battlefield*, Washington: Department of the Army.

Warrick, Joby. 2010. GAO blocks contract to firm formerly known as Blackwater to train Afghan police, *Washington Post*, 16/03, http://www.washingtonpost.com/wp-dyn/content/article/2010/03/15/AR2010031503289.html (accessed August 20, 2010).

6

Homegrown Islamist Terrorism: Assessing the Threat

Wes Heinkel and Alexandra Mace

This paper examines the relatively new phenomenon of homegrown Islamist terrorism in the United States. Although numerous U.S. policy makers and law enforcement officials have declared homegrown Islamist terrorism to be a grave and growing threat to the United States, there is a dearth of statistics to confirm or refute these claims. The purpose of this paper is to determine how many incidents of homegrown Islamist terrorism have been plotted and perpetrated in the United States since 9/11, how many Muslim-Americans have been involved in these plots, what their motivations were, and whether any similarities exist among the plots or the individuals who orchestrated them. Based on those findings, this paper provides an assessment of the magnitude of the threat posed by homegrown Islamist terrorism to the United States and proposes policy recommendations that could help prevent the occurrence of future attacks.

Wesley Craig Heinkel graduated from the Graduate School of Public and International Affairs at the University of Pittsburgh in December 2010 with a major in Security and Intelligence, where he focused on national security and foreign policy. He can be reached at wesheinkel@yahoo.com.

Alexandra Lindsay Mace graduated from the Graduate School of Public and International Affairs at the University of Pittsburgh in December 2010 with a major in Security and Intelligence, where she focused on diplomacy and foreign policy. She can be reached at alexandra.l.mace@gmail.com.

I. Introduction

Ever since the devastating terrorist attacks in New York and Washington on September 11, 2001 (9/11), Americans have lived in fear of another terrorist attack against the United States by al-Qaeda or another foreign Islamist terrorist group with a violent jihadist ideology. Over the past decade, the American public, policy makers, and law enforcement officials have grown increasingly concerned about terrorist "sleeper cells" infiltrating American society and awaiting orders to execute a terrorist attack. More recently, the focus of this concern has shifted to Muslim-Americans who become radicalized and engage in violent jihad on American soil. Both of these scenarios fall under the umbrella of a relatively new phenomenon that has come to be known as "homegrown Islamist terrorism."[1]

Nearly every month over the past two years, American law enforcement officials seem to announce that they have foiled yet another homegrown Islamist terrorist plot, each one purportedly more deadly than the previous plot. The incessant and often sensational media coverage of each thwarted attack has led to the impression that America is under siege from within by an endless army of violent Islamist jihadists, and that another attack of the scale of 9/11 – or worse – is virtually inevitable. Over the past year in particular, many U.S. policy makers and law enforcement officials have perpetuated this idea.

During a U.S. Senate Homeland Security and Governmental Affairs Committee hearing in September 2010, Committee Chairman Senator Joseph Lieberman stated that "an astoundingly high number of (Muslim) American citizens… have attacked or planned to attack their own country" (Associated Press 2010). In December 2010, U.S. Attorney General Eric Holder stated in an interview on ABC's "Good Morning America" that the threat of homegrown Islamist terrorism" is of great concern" and "keeps me up at night" (Cloherty and Thomas 2010). In February 2011, U.S. Attorney for the Eastern District of North Carolina George Holding stated that homegrown Islamist terrorism "is the greatest threat we're facing right now in the homeland" (Associated Press 2011). That same month, Homeland Security Secretary Janet Napolitano told the U.S. House of Representatives Committee on Homeland Security that due to the threat of homegrown Islamist terrorism, "the terrorist threat to the homeland is, in many ways, at its most heightened state since 9/11" (Napolitano 2011, 12). And in March 2011, Representative Peter King chaired the first of a series of Congressional hearings on "the critical issue of the radicalization of Muslim-Americans," which he characterized as "a real and dangerous threat to the safety and security of the citizens of the United States" (King 2011).

Despite these alarming proclamations, few facts or figures have been given to back up these statements, and little academic research has been performed to determine how many acts of homegrown Islamist terrorism have actually been plotted – let alone perpetrated – against the United States. Given the hype surrounding homegrown Islamist terrorism as well as the shortage of solid statistics regarding this phenomenon, the goal of this paper is to separate fact from fiction. It will examine the frequency as well as the severity of past attacks to determine the true magnitude of the threat posed to the United States by homegrown Islamist terrorism, determine how many incidents of homegrown Islamist terrorism have actually occurred in the United States since 9/11 through the end of 2010, how many attacks have been plotted, and what the nature of each of those plots was. This paper will also determine how many Muslim-Americans have been involved in these plots, what their motivations were, and whether any similarities exist among the plots or the individuals who orchestrated them. Based on those findings, this paper will conclude with an assessment of the nature of the threat posed by homegrown Islamist terrorism to the United States, and will propose policy recommendations that could help prevent the occurrence of future homegrown Islamist terrorist attacks.

II. Methodology

For the purpose of this paper, "homegrown Islamist terrorism" is defined as a deliberate act of violence directed against people or places within the United States that was plotted or perpetrated by at least one Muslim-American citizen (either U.S.-born, naturalized, or a legal permanent resident) for ideological and/or political reasons).[2] By these criteria, twenty-seven homegrown Islamist terrorist plots involving sixty individuals were either perpetrated or interrupted by law enforcement between September 11, 2001 and December 31, 2010. An analysis of legal documents and media reports associated with each of these plots led to the following key findings.

III. Key Findings

Individuals

There Is No Typical Homegrown Islamist Terrorist

Although there are some parallels among the sixty individuals involved in the twenty-seven plots, it is clear that there is no generic homegrown Islamist terrorist profile. While all but one of the individuals were male and the majority of them were in their twenties, other demographics varied widely, which defies the long-held assumption that individuals who com-

mit terrorism tend to be single, uneducated, and poor.

The individuals involved in these plots ranged from eighteen to sixty-three years of age with an average age of thirty, although the majority of individuals (thirty-six) were between the ages of eighteen and twenty-nine. As best as could be determined, the individuals involved were of various races: twenty-one were of African descent, twenty-one were of Middle Eastern or South Asian descent, sixteen were Caucasian, and two were Hispanic. Twenty-nine of the individuals were born in the United States, and thirty-one were born abroad in fourteen different countries, the most common of which was Pakistan (seven individuals were born there). The individuals resided in eighteen states, the most common of which was New York (fifteen individuals resided there).

Although the individuals' marital status was not always reported, at least twenty were single, at least thirteen were married, at least three were divorced, and at least eleven had children. The individuals' educational attainment also varied: some were high school drop outs, such as Najibullah Zazi, who planned a suicide attack on the New York City subway system (Wilson 2009), while others earned advanced degrees, such as Fort Hood shooter Nidal Hasan, who had a doctorate in psychiatry (Flaherty, Wan, and Davenport 2009). The individuals were not linked by a common career; their jobs ranged from fry cook (Michael Finton, who attempted to bomb the Paul Findley Federal Building in Springfield, Illinois) (Johnson 2009) to meteorological technician (Paul Rockwood Jr., who created a list of more than twenty Americans he planned to assassinate) (Rubenstein 2010). Several individuals worked in retail positions, many owned their own businesses, at least five drove taxis, limousines or shuttles, and at least four worked in Islamic schools or specialty stores. The individuals' economic statuses also varied: some, such as Tarek Mehanna, who plotted to attack shopping malls, came from very wealthy backgrounds (Ford and Kennedy 2009), while others such as Russell Defreitas, the alleged ringleader of a plot to blow up the jet fuel pipeline at John F. Kennedy International Airport in New York City, struggled to make ends meet (Fernandez 2007).

Many Homegrown Islamist Terrorists Had Experienced Serious Hardships

Although it was not possible to find detailed information regarding the personal lives of all sixty individuals, at least fifteen of the twenty-seven plots involved at least one individual who had recently experienced some type of serious hardship. At least three of the individuals were recently

divorced or separated, and at least two individuals' parents had recently divorced. At least four individuals had recently experienced the death of a parent or child, such as Daniel Boyd, whose youngest son was killed in a car accident less than two years before Boyd and six other men were arrested for plotting to attack Quantico Marine Corps Base in Virginia (Johnson and Hsu 2009). Other individuals were reported to have had alcohol and/or drug problems, chronic mental and physical illnesses, and bouts of homelessness, and at least one had been raised in foster care. Significant debt, bankruptcy, and other serious financial problems also plagued at least six individuals, including Times Square bomber Faisal Shahzad, whose house had gone into foreclosure shortly before his attempted attack (Barron and Tavernise 2010).

Many of these individuals had endured multiple hardships. James Elshafay, one of two individuals who plotted to bomb the Herald Square subway station in New York City, had recently dropped out of high school after failing the ninth grade three times. His parents were divorced, he had been sexually abused by a male relative, he abused drugs and alcohol, he was being treated for depression and schizophrenia, and he had recently spent time in a psychiatric ward (Rashbaum 2006). Iyman Faris, who was directed by al-Qaeda to cut the cables on the Brooklyn Bridge, had recently been released from a mental hospital after attempting suicide. He was going through a divorce, and his father had recently died (Thomas, Walsh, and Ryan 2003). And Laguerre Payen, one of four individuals accused of plotting to blow up the jet fuel pipeline at John F. Kennedy International Airport in New York City, was an illiterate, paranoid schizophrenic illegal immigrant who had recently been released from prison after serving fifteen months for assault and whom a judge had deemed too "insane" to deport to Haiti (Melago, Gendar, Kemp, and Lysiak 2009).

Several Homegrown Islamist Terrorists Had Previous Criminal Records

Although the criminal history of all sixty individuals is not definitively known, at least seven of the twenty-seven plots involved one or more individuals with previous criminal records, at least thirteen of whom had served time in prison. For example, Jose Padilla, who was directed by al-Qaeda to bomb apartment buildings, had served time in prison both as a juvenile and as an adult on assault and weapons-related charges (Wilgoren and Thomas 2002).

At least four of the plots involved at least one individual who had converted to Islam while in prison, the most notable of which was the plot led

by Kevin James. James was a California state prison inmate who founded a radical Muslim group and recruited dozens of fellow inmates as followers, some of whom James directed to commit terrorist attacks on the group's behalf upon their release from prison (Hamm 2008). This case is often cited as an example of the threat posed by prisoner radicalization, which is a growing concern among law enforcement officials and policy makers given that approximately 10 percent of all inmates have converted to Islam while in prison (Kapralos 2009). However, not all inmates who convert to Islam have become radicalized: a 2008 study funded by the National Institute of Justice found that "only a very small percentage of (Islamic) converts turn radical beliefs into terrorist action" and that "in most cases, the (Islamic) conversion experience makes a meaningful contribution to prisoner rehabilitation" (Hamm 2008).

Plots

Only Six Homegrown Islamist Terrorist Attacks Have Occurred Since 9/11

Of the twenty-seven planned homegrown Islamist terrorist attacks against the United States, only six plots were actually perpetrated.[3] On July 4, 2002, Hesham Hadayet shot and killed two people at the El Al Airlines ticket counter at Los Angeles International Airport before being killed by an El Al security guard (Dixon, Leonard, and Connell 2002). On March 3, 2006, Mohammed Taheri-Azar drove a Jeep Grand Cherokee through a crowded courtyard on the University of North Carolina's Chapel Hill campus, injuring nine people (Associated Press 2006). On July 28, 2006, Naveed Haq shot and killed one person and wounded five others at the Jewish Federation of Greater Seattle (Superior Court of Washington for King County 2008). On June 1, 2009, Carlos Bledsoe shot two soldiers at a U.S. Army recruiting station in Little Rock, Arkansas, killing one soldier and seriously wounding the other (Federal Bureau of Investigation Little Rock 2009). On November 5, 2009, Army Major Nidal Hasan shot and killed thirteen people and wounded thirty others at the Soldier Readiness Center on the Fort Hood U.S. Army Base in Killeen, Texas (McFadden 2009). And on May 1, 2010, Faisal Shahzad attempted to detonate an SUV filled with explosives in Times Square in New York City, but no one was injured (Mazzetti, Tavernise, and Healy 2010). A total of seventeen people were killed and forty-seven people were injured in all six of these attacks (not including the perpetrators).

Most Homegrown Islamist Terrorists are Lone Wolves

More than half (fourteen) of the twenty-seven homegrown Islamist terrorist plots against the United States were planned and/or perpetrated by one individual. From Michael Finton's plot to bomb a federal building in Springfield, Illinois (Johnson 2009) to Farooque Ahmed's plot to bomb Northern Virginia Metro stations (Hsu 2010), the majority of these plots were planned by so-called lone wolves. Of the other thirteen plots, six involved two people, three involved four people, two involved seven people, one involved three people, and one involved six people. All six of the homegrown Islamist terrorist attacks that have occurred in the United States since 9/11 were committed by a single individual.[4]

Half of the Individuals Who Committed Homegrown Islamist Terrorist Attacks Had Previously Aroused Law Enforcement Suspicion

The individuals responsible for three of the six homegrown Islamist terrorist attacks – Carlos Bledsoe, Nidal Hasan, and Faisal Shahzad – had previously come to the attention of law enforcement officials before they perpetrated their attacks, but all three were determined not to have posed a threat. Fort Hood shooter Nidal Hasan came to the attention of the Federal Bureau of Investigation (FBI) after sending emails to Anwar al-Awlaki, a radical cleric and propagandist for al-Qaeda in the Arabian Peninsula whose email account was being monitored by the FBI (Gibbs 2009). After Hasan's correspondence with al-Awlaki was reviewed by both the Washington-area Joint Terrorism Task Force (led by the FBI) and the Pentagon's Defense Criminal Investigative Service, it was determined to be innocuous (Gibbs 2009). Little Rock recruiting center shooter Carlos Bledsoe was interviewed by the FBI after being arrested and imprisoned in Yemen over a visa violation and for allegedly possessing fake Somali identification papers and was interviewed again by the FBI upon his return to the United States, but was determined not to have any connections to terrorism (Dao 2010). And Times Square bomber Faisal Shahzad had been placed on the Traveler Enforcement Compliance System (TECS), a Department of Homeland Security (DHS) watch list, for bringing $80,000 in cash or instruments into the United States, but had been removed from the list in 2008 for unknown reasons (CBS News 2010).

Only Four Homegrown Islamist Terrorist Plots Were Directed by Foreign Terrorist Organizations

Foreign terrorist organizations were not directly involved with the vast

majority of the plots: only four of the twenty-seven plots were ordered and/or funded by foreign terrorist groups. After receiving weapons and explosives training at an al-Qaeda training camp in Afghanistan in 2000, Jose Padilla met al-Qaeda operative Khalid Shaykh Mohammad in Pakistan in 2002, who directed Padilla to return to the United States to bomb several apartment buildings and provided him with thousands of dollars in cash to initiate the plan (U.S. Deputy Attorney General 2004, 6). After attending an al-Qaeda training camp in Afghanistan in late 2000 and providing logistical support for al-Qaeda operations in Pakistan, Iyman Faris returned to the United States in 2002 on orders from al-Qaeda to destroy the Brooklyn Bridge (U.S. District Court for the Eastern District of Virginia 2003). While in Pakistan in late 2008, Najibullah Zazi was recruited by al-Qaeda and received weapons and explosives training before being directed to return to the United States to conduct a suicide attack on the New York City subway system (Federal Bureau of Investigation New York 2010). And after receiving weapons and explosives training at an Tehrik-e-Taliban Pakistan training camp in 2009, Faisal Shahzad was given several thousand dollars in cash and directed to return to the United States to detonate a car bomb in New York's Times Square in 2010 (Elliott 2010).

The Majority of Homegrown Islamist Terrorist Attacks Have Been Shootings

Of the six homegrown Islamist terrorist attacks, four were shootings, one involved explosives, and one used a vehicle as a weapon. The types of weapons that were most commonly intended to be used in all twenty-seven plots were explosives (twelve plots) and guns (ten plots). None of the twenty-seven plots involved the use of chemical, biological, radiological, or nuclear (CBRN) weapons (Bergen and Hoffman 2010, 4).[5]

The guns, explosives, and other weapons that were to be used to carry out these plots had only been acquired in twelve of the twenty-seven plots (including the six attacks that actually occurred) before the plots were perpetrated or interrupted by law enforcement. At the time of their arrests, the individuals involved in the other fifteen plots either had not yet acquired any weapons (ten plots), or the FBI had supplied them with fake weapons (five plots). For example, the FBI provided Mohamed Mohamud with inert explosives as part of Mohamud's plot to detonate a car bomb at a Christmas tree lighting ceremony in Portland, Oregon (U.S. District Court of Oregon 2010, 5).

Only One Homegrown Islamist Terrorist Plot was Intended to be a Suicide Attack

With the exception of the al-Qaeda-directed plot to bomb several New York City subway stations to be carried out by Najibullah Zazi and two of his high school classmates, none of the other twenty-six plots were specifically intended to be suicide operations.[6] In fact, several individuals specifically expressed their unwillingness to participate in so-called martyrdom operations. For example, Shahawar Siraj, one of two individuals who plotted to bomb the Herald Square subway station in New York City, refused to place the bombs inside of trash cans in the subway station as had been previously discussed, and stated to his co-conspirator, "I am not ready to die" (Horowitz 2005, 3).

U.S. Military Members Account for the Majority of Casualties and are the Most Common Target of Homegrown Islamist Terrorist Plots

Thirteen of the seventeen casualties of the six homegrown Islamist terrorist attacks were members of the U.S. military, twelve of whom were killed during Nidal Hasan's shooting rampage at Fort Hood U.S. Army Base. U.S. military personnel and/or installations were the most common target of the twenty-seven plots, as they were specifically targeted in ten of them, including Carlos Bledsoe's shooting at a U.S. Army recruiting center in Little Rock, during which one soldier was killed. The targeting of U.S. military personnel in these ten plots was deliberate. Mohamed Shnewer, who planned to attack Fort Dix U.S. Army Base in New Jersey along with five other men, stated that "my intent is to hit a heavy concentration of soldiers" (U.S. District Court of New Jersey 2007, 11). Ahmed Abu Ali, who plotted to assassinate President George W. Bush after he was unable to travel to Afghanistan to fight U.S. troops, stated that "[I] had a dream that I was shooting American soldiers, and I wanted to make it happen" and "if we got any opportunity to shoot American soldiers, it's good. Even if you kill two of them, it would be good" (Federal Bureau of Investigation 2003, 1–2).

Israeli/Jewish Landmarks and Businesses are the Second Most Common Target of Homegrown Islamist Terrorist Plots

Synagogues, Jewish centers, and other buildings and businesses with ties to Israel were specifically targeted in five of the twenty-seven plots, including two of the six plots that were actually perpetrated (Hesham Hadayet's

shooting at the El Al counter at Los Angeles International Airport, and Naveed Haq's shooting at the Jewish Federation of Greater Seattle). The El Al counter at Los Angeles International Airport was targeted again in a 2004 plot by Kevin James and three other men, who also targeted the Israeli consulate in Los Angeles as well as synagogues in Los Angeles and Sacramento, California (U.S. Department of Justice 2005). Synagogues and temples were also targeted in two other plots.

Individuals Involved in One-Third of Homegrown Islamist Terrorist Plots Had Accessed Violent Jihadist Propaganda on the Internet

Individuals involved in at least nine of the twenty-seven plots had accessed and/or engaged in various forms of jihadist propaganda on the Internet. Tarek Mehanna, who plotted to attack shopping malls, watched, translated, and re-posted violent jihadist videos and writings on the Internet and considered himself to be the "media wing" of al-Qaeda in Iraq (U.S. District Court of Massachusetts 2009, 5). Antonio Martinez, who attempted to detonate a car bomb outside of the Armed Forces Career Center in Catonsville, Maryland, viewed martyrdom videos and messages from Osama bin Laden, posted links to videos promoting violent jihad on his Facebook profile and was "friends" with several pro-jihad groups on Facebook (U.S. District Court of Maryland 2010, 6, 9–10). And Portland plotter Mohamed Mohamud contributed several articles to the online pro-jihad publication "Jihad Recollections" (U.S. District Court of Oregon 2010, 3).

Individuals in seven of these plots were specifically influenced by Anwar al-Awlaki, an American-born radical Muslim cleric now based in Yemen and affiliated with al-Qaeda in the Arabian Peninsula, whose pro-jihadist sermons and propaganda videos are frequently posted on the Internet. Nidal Hasan exchanged eighteen emails with al-Awlaki in the eleven months prior to his shooting rampage at Fort Hood (Shane and Mazzetti 2010). Times Square bomber Faisal Shahzad told investigators that he was inspired by al-Awlaki's online sermons (Shane and Mazzetti 2010). And Paul Rockwood Jr., who plotted to assassinate twenty Americans, had been a strict adherent of al-Awlaki's radical Islamist ideology for more than four years (U.S. District Court of Alaska 2010, 7). While it is impossible to prove causality between al-Awlaki's influence and the perpetration of any of these plots, the United States believes al-Awlaki poses such a serious threat that in April 2010, President Obama took the unprecedented step of authorizing the targeted killing of al-Awlaki (Shane 2010), and, at the

urging of American and British lawmakers, YouTube removed hundreds of videos featuring al-Awlaki in November 2010 (Burns and Helft 2010).

Many of the Individuals Involved in Homegrown Islamist Terrorist Plots Had Recently Traveled Abroad

In eleven of the twenty-seven plots, at least one individual involved in plotting or perpetrating the attack had recently traveled abroad prior to the planned date of attack, the most common destination being Pakistan (seven plots). For example, Raja Khan, who plotted to blow up a sports stadium, traveled to Pakistan in 2008 and 2009 and was planning a third trip before he was arrested in 2010 (U.S. District Court for the Northern District of Illinois 2010). In five of the plots in which at least one individual involved had recently traveled abroad, that individual had done so in order to attend a terrorist training camp. One such individual was Times Square bomber Faisal Shahzad, who traveled to Pakistan in 2009, attended a Tehrik-e-Taliban Pakistan training camp, and received instruction in bomb-making before returning to the United States (Elliott 2010). In two of these plots, at least one individual attempted to attend a terrorist training camp but was rejected. For example, Ahmad Abousamra, who conspired to attack shopping malls with Tarek Mehanna, traveled to Pakistan twice in 2002 and to Yemen in 2004 seeking terrorist training but was rejected by both Lashkar-e-Taiba and the Taliban (Hsu 2009).

The Majority of Homegrown Islamist Terrorist Plots Involved the Use of Informants and/or Undercover FBI Agents

Informants and/or undercover FBI agents played a role in fifteen of the twenty-seven plots, three of which involved the use of both. Many of the informants were cooperating with law enforcement agencies in order to avoid jail time, deportation, or both. For example, Shahed Hussain, an illegal immigrant from Pakistan, agreed to cooperate with the FBI to avoid being sent to prison on fraud charges and then deported (Rayman 2009). Hussain served as an informant in two of the most controversial homegrown Islamist terrorism plots. He was involved in the 2004 arrests of Yassin Aref and Mohammed Hossain in Albany, New York for agreeing to launder money for a fictional plot made up by Hussain to assassinate Pakistani Prime Minister Pervez Musharraf (Rayman 2009). He also played a role in the 2009 arrests of James Cromitie, Laguerre Payen, David Williams, and Onta Williams for plotting attacks on the Riverdale Jewish Center and Riverdale Temple in the Bronx, New York and Stewart Air National Guard Base in Newburgh, New York (Fahim 2010). In both cases, Hussain

claimed fictional ties to terrorist groups, suggested which targets to attack, and provided money (and allegedly, in one case, drugs) to the individuals involved in order to encourage them to cooperate (Rayman 2009).

Informants who participated in other plots were often paid large sums of money in return for their cooperation. Osama Eldawoody, a naturalized American citizen from Egypt who began working as a confidential informant with the New York City Police Department in 2003, earned more than $100,000 for his role in the 2004 arrests of James Elshafay and Shahawar Siraj, who plotted to bomb the Herald Square subway station in New York City (Shulman 2007). As with the plots in which Hussain was involved, Eldawoody claimed that he had connections to a terrorist group, suggested the target, and told Elshafay and Siraj that he could obtain explosives. In another plot, the FBI paid an informant and dismissed his previous arrests for assault and drug possession in exchange for his participation in the 2006 arrests of Narseal Batiste and six other men in Miami, Florida for plotting to bomb the Sears Tower in Chicago, Illinois (Pincus 2006). When that informant failed to provide the FBI with evidence of a terrorist plot, the FBI brought in another informant, whom they also paid and approved for a petition for political asylum in the United States (Pincus 2006). This second informant provided Batiste and the six other men with vehicles, a meeting space, cell phones and video cameras, instructed them to conduct surveillance of specific targets, and asked them to take oaths of allegiance to al-Qaeda, which was the key evidence used to convict the men of terrorism (Pincus 2006).

Motives

Most Homegrown Islamist Terrorists Believe that the United States is at War with Islam

The belief that the United States is at war with Islam is a commonly held perception among many of the individuals involved in these twenty-seven plots. For example, Nidal Hasan gave a presentation to his fellow graduate students titled "Why the War on Terror is a War on Islam" less than two years before the Fort Hood shootings (Shane and Dao 2009). During his plea hearing, Times Square bomber Faisal Shahzad repeatedly referred to the United States' "war against Allah" and "war with Muslims" and stated that "(President) Bush had made already clear when he started the war on us, on Muslims, he said, you are either with us or against us" (U.S. District Court for the Southern District of New York 2010, 1–2). And Antonio Martinez, who attempted to bomb a military recruitment center in Maryland, stated that "each and every Muslim in this country… knows that

America is at war with Islam" (U.S. District Court of Maryland 2010, 6).

Fourteen of the Homegrown Islamist Terrorist Plots Appear to Have Been Motivated Primarily by a Desire to Engage in Violent Jihad

In fourteen of the twenty-seven plots, the individuals involved did not mention a particular grievance against the United States, but they specifically expressed a desire and/or sense of obligation to engage in violent jihad in order to defend the Islamist ideology to which they prescribed. For example, during the sentencing hearing for his plot to attack a shopping mall in Rockford, Illinois, Derrick Shareef told the court that his "intention was to bring victory to Islam" (Vrsansky 2008). Paul Rockwood Jr., who planned to assassinate twenty people throughout the United States, believed that it was his "personal responsibility to exact revenge by death on anyone who desecrated Islam" (U.S. District Court of Alaska 2010, 8). And Kevin James, the alleged ringleader of a plot to attack several targets throughout California, stated that the purpose of the attacks was to "defend and propagate traditional Islam in its purity" (Federal Bureau of Investigation Los Angeles 2009).

Eleven of the Homegrown Islamist Terrorist Plots Appear to Have Been Motivated Primarily by Anger over U.S. Military Action in the Middle East

In eleven of the twenty-seven plots, at least one individual involved expressed anger over U.S. military action in the Middle East and desire to exact revenge. For example, after pleading guilty to attempting to detonate a car bomb in Times Square, Faisal Shahzad stated that "until the hour the U.S. pulls its forces from Iraq and Afghanistan, and stops the drone strikes in Somalia and Yemen and in Pakistan, and stops the occupation of Muslim lands, and stops killing the Muslims, and stops reporting the Muslims to its government, we will be attacking (the) U.S" (Elliott 2010). Najibullah Zazi stated that the attack that he and his two co-conspirators were planning on the New York City subway system was intended "to bring attention to what the United States military was doing to civilians in Afghanistan" (Sulzberger and Rashbaum 2010). Tarek Mehanna stated that it is "unfathomable" that the United States has military bases in the "heart of the Muslim world" and that the "land of Mohammad… is being used as a military base to attack Muslims" (U.S. District Court of Massachusetts 2009, 35). And Carlos Bledsoe stated that his shooting of

U.S. Army recruiters in Little Rock "was an attack of retaliation" against the U.S. military because "U.S. soldiers are killing innocent Muslim men and women," and stated that he also wanted revenge against U.S. service members for allegedly desecrating copies of the Quran and raping Muslim women (Associated Press 2009).

Two Homegrown Islamist Terrorist Plots Appear to Have Been Motivated Primarily by a Hatred of Israel

Hesham Hadayet's shooting at the El Al ticket counter at Los Angeles International Airport and Naveed Haq's shooting at the Jewish Federation of Greater Seattle both seem to have been primarily motivated by their hatred of Israel. One of Hadayet's former employees stated that Hadayet had expressed hatred of Israel and blamed Israel for the turmoil in the Middle East (Dixon, Leonard, and Connell 2002). And Naveed Haq told a 911 operator during the shooting at the Jewish Federation of Greater Seattle that he was upset that the United States was "sending bombs to Israel" and that "the Jews are running the country" and stated that he was "tired of our people (Muslims) getting pushed around" by Israel (Superior Court of Washington for King County 2008, 6).

Individuals involved in five other plots also specifically expressed their hatred of Israel and/or Jewish people, even if they did not specifically target Israeli/Jewish facilities. For example, Russell Defreitas was angry that the United States was supplying weapons to Israel, which he believed would be used to kill Muslims, and stated that he "wanted to do something to get those b******s" (U.S. District Court Eastern District of New York 2007, 15). And James Cromitie expressed his desire to bomb a synagogue, adding, "I hate those… f****** Jewish b******s" (U.S. District Court for the Southern District of New York 2009, 7).

IV. Assessment

The magnitude of the threat posed by homegrown Islamist terrorism has been greatly exaggerated. As Charles Kurzman, Director of Graduate Studies at the University of North Carolina Chapel Hill's Sociology Department concluded in a February 2011 report, "out of the thousands of acts of violence that occur in the United States each year, an efficient system of government prosecution and media coverage brings Muslim-American terrorism suspects to national attention, creating the impression – perhaps unintentionally – that Muslim-American terrorism is more prevalent than it really is" (Kurzman 2011, 7).

As previously noted, only six homegrown Islamist terrorist attacks have actually occurred in the United States since the 9/11 terrorist attacks nearly ten years ago, which have resulted in the deaths of seventeen Americans. To put that in perspective, more than twenty times as many Americans have been killed by lightning (Cooper 2007) and nearly forty times as many Americans have been killed by bee stings since 9/11 (Zhou 2009) as have been killed in homegrown Islamist terrorist attacks. Over the same time period, more than 360,000 Americans were killed in motor vehicle accidents (National Highway Traffic Safety Administration 2010; National Highway Traffic Safety Administration 2011), and more than six million Americans died due to heart disease (Centers for Disease Control and Prevention 2011).

Many of the key findings raise questions as to the severity of the threat posed by some of these homegrown Islamist terrorist plots. As previously mentioned, the weapons that were to be used in these plots were only actually acquired in fewer than half of the plots, and more than half of the plots involved the sometimes questionable participation of informants and/or undercover FBI agents. It is not clear that the twenty-one plots that were interrupted by law enforcement before they could be perpetrated – one of which Deputy FBI Director John Pistole admitted was "more aspirational than operational" (Associated Press 2006) – would have ever actually been perpetrated were it not for the intervention of law enforcement. Additionally, despite fears and predictions that suicide attacks might begin taking place within the United States as they have overseas (Bergen and Hoffman 2010, 27), only one of the twenty-seven plots was specifically intended to be a suicide operation.

Furthermore, two-thirds of the attacks that actually occurred were shootings, and the majority of the plots – including all six attacks which actually occurred – were planned and/or perpetrated by individuals acting alone. As *Time* Deputy Managing Editor Romesh Ratnesar wrote in January 2011, "a lone wolf has little chance of pulling off the kind of mass-casualty strike that counterterrorism experts worry about most" (Ratnesar 2011). The findings that none of the plots involved the use of CBRN weapons and that only four of the twenty-seven plots were directed by foreign terrorist organizations suggest that not only are these types of mass-casualty attacks unlikely, but that al-Qaeda and other foreign Islamist terrorist organizations' operational abilities have been greatly diminished since 9/11.

The recent increase in arrests of Muslim-Americans allegedly involved in homegrown Islamist terrorism can be attributed to increased vigilance on

the part of law enforcement, as there is no evidence to support Congressman Peter King's recent claim that "a growing number of young (Muslim) Americans (are) being radicalized" by "al-Qaeda and its adherents" (King 2011, 1–2). Although al-Qaeda propagandist Anwar al-Awlaki's YouTube videos have been viewed more than 3.5 million times (Meek and Bazinet 2010), al-Awlaki only appears to have influenced seven of the twenty-seven plots. While more than half of the plots seem to have been primarily motivated by the type of radical ideology that al-Qaeda and other foreign Islamist terrorist groups promulgate, many of those plots involved individuals who had experienced serious hardships and/or had previous criminal records, either of which could have predisposed them to commit acts of violence.

Other research has also failed to find evidence that the Muslim-American population is becoming increasingly radicalized. A 2010 study published by RAND noted that out of the more than three million Muslims in the United States, fewer than one hundred have been involved in homegrown Islamist terrorism since 9/11, which suggests that a significant majority of the Muslim-American population does not prescribe to radical Islamist ideology (Jenkins 2010, vii). And a 2010 study published by Duke University and the University of North Carolina found that "Muslim-American communities strongly reject radical jihadist ideology, are eager to contribute to the national counterterrorism effort, and are fiercely committed to integration within the mainstream of American social and economic life" and "are taking a variety of positive steps that help prevent radicalization within their communities" (Schanzer, Kurzman, and Moosa 2010, 40).

While the threat of homegrown terrorism carried out by Muslim-Americans has been vastly overstated, it still must be taken seriously. However, it is equally important to keep the magnitude of the purported threat of homegrown Islamist terrorism in perspective. As noted in a September 2010 Bipartisan Policy Center assessment of the threat posed by radicalized Muslim-Americans, "it must be emphasized that the number of U.S. citizens and residents affected or influenced in this manner remains extremely small" (Bergen and Hoffman 2010, 30). As such, U.S. policy makers, law enforcement officials, and media members should refrain from making generalizations about the entire Muslim-American population based on the actions of a few individuals, as doing so will only decrease Muslim-Americans' trust of and cooperation with law enforcement and further alienate and ostracize the Muslim-American community, which in turn would increase the vulnerability of Muslim-Americans to radicalization.

In conclusion, it seems that the February 2010 Annual Threat Assessment

by former Director of National Intelligence Dennis Blair most accurately describes the current threat posed to the United States by homegrown Islamist terrorism.

> It is clear… that a sophisticated, organized threat from radicalized individuals and groups in the United States comparable to traditional homegrown threats in other countries has not emerged… Violence from homegrown jihadists probably will persist, but will be sporadic. A handful of individuals and small, discrete cells will seek to mount attacks each year, with only a small portion of that activity materializing into violence against the Homeland. (Blair 2010, 11)

Based on the above conclusions, this paper proposes several policy recommendations that U.S. policy makers should consider.

V. Policy Recommendations

1. The U.S. Government should not attempt to counter the radical Islamist narrative on its own.

Given that most U.S. policy makers do not possess even basic Islamic literacy, any official attempt to condemn the radical Islamist narrative perpetuated by al-Qaeda and other foreign terrorist organizations might be misinterpreted by Muslims-Americans as an attack on Islam, which would further alienate the Muslim-American community and could inadvertently incite the very sort of violence that it is intended to prevent. Given how few Muslim-Americans actually prescribe to radical Islamist ideologies, it is not worth the risk. Instead, U.S. policy makers should seek to empower mainstream Muslim-American leaders and fund mainstream Muslim organizations whose efforts to counter the violent jihadist narrative are likely to be more credible to and effective among the Muslim-American population.

2. U.S. policy makers should recognize that current U.S. foreign policy in the Middle East increases the risk of homegrown Islamist terrorism, but should not necessarily alter it.

Although anger over the U.S. military presence in Afghanistan and Iraq, as well as hatred of Israel, appear to have motivated almost half of the plots that have been planned or perpetrated in the United States since 9/11, the overall lack of severity of the threat does not necessitate a change of course on foreign policy. An immediate withdrawal of U.S. troops from the Middle East and an abrupt shift in relations with Israel in order to prevent future

homegrown Islamist terrorist attacks from occurring in the United States would not be warranted. However, the threat is still significant enough that it should be included in any risk-benefit analysis performed by U.S. policy makers when considering Middle East foreign policy. Based on past homegrown Islamist terrorist plots, it is likely that as long as the U.S. military has a presence in the Middle East and as long as Israel refuses to recognize Palestinian statehood, homegrown Islamist terrorism is likely to continue to occur – albeit sporadically – in the United States.

3. The U.S. should encourage and support the fledgling democratic movements in the Middle East.

The recent uprisings throughout the Middle East have weakened the radical Islamist narrative propagated by al-Qaeda and other foreign terrorist organizations by proving that change can be effected in the Middle East through non-violent, secular, and democratic means. As Paul Pillar, Director of Graduate Studies at Georgetown University's Center for Peace and Security Studies recently stated, "…democracy is bad news for terrorists. The more peaceful channels people have to express grievances and pursue their goals, the less likely they are to turn to violence" (Shane 2011). The United States should provide political, logistical, and financial support to countries such as Egypt and Tunisia to help them transition to democratic societies, and encourage populist uprisings in countries such as Yemen, where the radical Islamist narrative is more widely accepted. However, U.S. policy makers should be mindful that, as is the case with any U.S. military action in a Muslim country, providing military assistance to facilitate these revolutions (such as in Libya) – even without actual troops on the ground – may anger some Muslims, and could possibly even provoke homegrown Islamist terrorist attacks.

4. U.S. law enforcement agencies, particularly on the local level, should do more to engage and establish relationships with Muslim communities within their jurisdictions.

In particular, they should be responsive to instances of bias or hate crimes committed against members of their local Muslim community. As Muslim-Americans begin to view law enforcement officers as allies rather than enemies, they will be more likely to speak up if they feel someone in their community may be becoming radicalized, which will reduce the need to rely on informants. Examples of programs worth emulating are the Los Angeles County Sheriff's Department's Muslim Community Affairs Unit (Los Angeles County Sheriff's Department 2011), the Seattle Police

Department's Arab, Muslim, and Sikh Advisory Council (Seattle Police Department 2011) and the FBI's Specialized Community Outreach Teams (Hovington 2010), all of which seek to foster trust and enhance ties with their local Muslim communities.

5. Individuals who frequently travel to countries known to be associated with terrorism should be required to undergo secondary security procedures at air and seaports upon arriving from, departing to, or traveling through those countries.[7]
Such procedures could include a more thorough inspection of their luggage or verification of their whereabouts in that country. Suspicious findings, such as a large amount of cash, radical ideological materials, or addresses that cannot be verified should trigger further examination. Since the majority of individuals who will be required to undergo these procedures will likely be traveling for innocuous purposes, perhaps some type of trusted traveler system such as the CLEAR program, which requires the disclosure of extensive background information to enroll and relies on biometric identification techniques to identify the traveler (Stone 2010), could be put in place so that frequent travelers to these countries whose identities have been verified can opt out of secondary screening measures.

6. The U.S. government should establish an office within the National Counterterrorism Center (NCTC) with the sole purpose of monitoring violent jihadist propaganda on the Internet.[8]
Currently, the task of monitoring the myriad violent Islamist websites on the Internet is primarily handled by private security firms such as the SITE Intelligence Group (SITE Intelligence Group 2011) and Griffith Colson Intelligence Service (Griffith Colson Intelligence Service 2011), which distribute translations of significant postings to intelligence and law enforcement agencies that subscribe to their services. However, an office should be established within the NCTC in order to perform these sensitive tasks internally. As suggested in a 2008 report by the New America Foundation, this office should "monitor the few jihadist websites with high intelligence value" (Bergen and Footer 2008) and should also coordinate with social media sites such as Facebook and YouTube to ensure that violent jihadist propaganda videos which violate these sites' terms of conduct are removed as soon as they are discovered.

Notes

[1] Although the phrase "homegrown Islamic terrorism" is the term most frequently used in the media to describe terrorist plots by Muslim-Americans, it is inaccurate in that it implies that the individuals involved in these plots are adherents of mainstream Islam. Although much scholarly debate exists as to what the proper term to describe Muslims who prescribe to a violent jihadist ideology is, for the purpose of this paper, the individuals involved in homegrown terrorist plots have been termed as Islamists using the definition of Dr. Jeffrey M. Bale, Director of the Monterey Terrorism Research and Education Program at the Monterey Institute of International Studies. Dr. Bale defines Islamism as "an intrinsically radical and anti-democratic extreme right-wing political ideology" (Bale 2009, 73) which is "based upon an exceptionally intolerant and puritanical interpretation of Islamic scriptures and Islamic law" (Bale 2009, 79). Therefore, a more accurate term for the type of terrorism examined in this paper is "homegrown Islamist terrorism."

[2] This figure does not include plots by American Muslims against domestic targets that were planned prior to the 9/11 terrorist attacks; plots which were perpetrated or planned by American Muslims against foreign targets; plots in which American Muslims traveled or attempted to travel abroad with the goal of engaging in violent jihad and/or joining a foreign terrorist organization; plots in which American Muslims attempted to provide classified material to foreign terrorist groups; or plots against domestic targets in which the sole perpetrator or plotter was Muslim but was not a U.S. citizen.

[3] In a February 2011 report published by the University of North Carolina that used a methodology similar to the one used in this paper, sociologist Charles Kurzman concluded that ten homegrown Islamist terrorist attacks have occurred in the United States since 9/11, resulting in thirty-three deaths (Kurzman 2011, 4–5). However, we did not find there to be sufficient evidence in four of those attacks to indicate that they were perpetrated for ideological and/or political reasons.

[4] Although Faisal Shahzad's unsuccessful detonation of a car bomb in Times Square was directed and funded by Tehrik-e-Taliban Pakistan, Shahzad was the only individual involved in the planning and perpetration of the plot.

[5] Although it was initially alleged that al-Qaeda had directed Jose Padilla to detonate a radiological device in the United States, "the allegation was dropped when the case went to trial" (Bergen and Hoffman 2010, 4).

[6] It has not been definitively determined whether Nidal Hasan intended to be killed during his shooting rampage at Fort Hood.

[7] Since the Transportation Security Agency (TSA) does not disclose what criteria it

uses to select travelers for secondary screenings, it is not known if this recommendation is already being implemented.

[8] As with the previous recommendation, the NCTC will not disclose whether they have such capabilities, so it is not known if this recommendation is already being implemented.

References

Associated Press, The. 2011. Man Charged as N.C. Terror Ringleader Pleads Guilty. *The Associated Press*, February 9. http://www.news-record.com/content/2011/02/09/article/man_charged_as_nc_terror_ringleader_due_in_court (accessed May 17, 2011).

———. 2010. Terrorists Trying to Recruit U.S. Residents Is Latest Challenge, U.S. Says, September 23. http://www.foxnews.com/politics/2010/09/22/homegrown-terrorism-evolving-officials-tell-congress/ (accessed May 17, 2011).

———. 2009. Recruitment Shooting Suspect Doesn't Think Killing Was Murder, June 9. http://www.foxnews.com/story/0,2933,525584,00.html (accessed December 15, 2010).

———. 2006. FBI Says U.S. Suspects Sought Al-Qaida Help, June 24. http://www.msnbc.msn.com/id/13524120/ns/us_news-security/t/fbi-says-us-suspects-sought-al-qaida-help/ (accessed December 15, 2010).

———. 2006. UNC Graduate Charged with Attempted Murder, March 4. http://www.usatoday.com/news/nation/2006-03-04-suv-crash_x.htm (accessed December 15, 2010).

Bale, Jeffrey M. 2009. Islamism and Totalitarianism. *Totalitarian Movements and Political Religions* 10(2): 73-96. http://www.miis.edu/media/view/18961/original/baleislamismandtotalitarianism.pdf (accessed May 17, 2011).

Barron, James, and Sabrina Tavernise. 2010. Money Woes, Long Silences and a Zeal for Islam. *New York Times*, May 5. http://www.nytimes.com/2010/05/06/nyregion/06profile.html (accessed December 15, 2010).

Bergen, Peter, and Bruce Hoffman. 2010. Assessing the Terrorist Threat: A Report of the Bipartisan Policy Center's National Security Preparedness Group. *Bipartisan Policy Center*, September 10. http://bipartisanpolicy.org/sites/default/files/NSPG%20Final%20Threat%20Assessment.pdf (accessed May 17, 2011).

Bergen, Peter, and Laurence Footer. 2008. Defeating the Attempted Global Jihadist Insurgency: Forty Steps for the Next President to Pursue against al Qaeda, Like-Minded Groups, Unhelpful State Actors, and Radicalized Sympathizers. *The Annals of the American Academy of Political and Social Science*, 618. http://counterterrorism.newamerica.net/node/8924 (accessed May 17, 2011).

Blair, Dennis C. 2010. Annual Threat Assessment of the U.S. Intelligence Community for the Senate Select Committee on Intelligence, February 2. http://www.dni.gov/testimonies/20100202_testimony.pdf (accessed December 15, 2010).

Burns, John F. and Miguel Helft. 2010. YouTube Withdraws Cleric's Videos. *New York Times*, November 4. http://www.nytimes.com/2010/11/05/world/05britain.html (accessed December 15, 2010).

CBS News. 2010. Shahzad on U.S. Travel Security List Since 1999, May 10. http://www.cbsnews.com/8301-31727_162-20004263-10391695.html (accessed December 15, 2010).

Centers for Disease Control and Prevention. 2011. Leading Causes of Death. http://www.cdc.gov/nchs/fastats/lcod.htm (accessed May 17, 2011).

Cloherty, Jack, and Pierre Thomas. 2010. Attorney General's Blunt Warning on Terror Attacks. *ABC News*, December 21. http://abcnews.go.com/Politics/attorney-general-eric-holders-blunt-warning-terror-attacks/story?id=12444727 (accessed May 17, 2011).

Cooper, Mary Ann. 2007. Medical Aspects of Lightning. National Weather Service. http://www.lightningsafety.noaa.gov/medical.htm (accessed December 15, 2010).

Dao, James. 2010. A Muslim Son, a Murder Trial and Many Questions. *New York Times*, February 16. http://www.nytimes.com/2010/02/17/us/17convert.html (accessed December 15, 2010).

Dixon, Robyn, Jack Leonard and Rich Connell. 2002. Those Who Knew LAX Killer Say Personal Agenda Died With Him. *Los Angeles Times*, July 14. http://articles.latimes.com/2002/jul/14/local/me-hadayet14 (accessed May 17, 2011).

Elliott, Andrea. 2010. Militant's Path from Pakistan to Times Square. *New York Times*, June 22. http://www.nytimes.com/2010/06/23/world/23terror.html (accessed December 15, 2010).

Fahim, Kareem. 2010. Tapes in Bomb-Plot Trial Show Suspect Wavering as an Informer Prods Him to Act. *New York Times*, August 31. http://www.nytimes.com/2010/09/01/nyregion/01plot.html (accessed December 15, 2010).

Federal Bureau of Investigation. 2003. Interview of Ahmed Omar Abu Ali, June 15. http://www.investigativeproject.org/documents/case_docs/288.pdf (accessed May 17, 2011).

Federal Bureau of Investigation – Little Rock. 2009. Arrest of Abdulhakim Mujahid Muhammed, June 2. http://www.fbi.gov/littlerock/press-releases/2009/lr060209.htm Arrest of Abdulhakim Mujahid MuhammedArrest of Abdulhakim Mujahid Muhammed (accessed December 15, 2010).

Federal Bureau of Investigation – Los Angeles. 2009. Man who Formed Terrorist Group that Plotted Attacks on Military and Jewish Facilities Sentenced to 16 Years in Federal Prison, March 6. http://www.fbi.gov/losangeles/press-releases/2009/la030609ausa.htm (accessed December 15, 2010).

Federal Bureau of Investigation – New York. 2010. Najibullah Zazi Pleads Guilty to Conspiracy to Use Explosives Against Persons or Property in U.S., Conspiracy to Murder Abroad, and Providing Material Support to al Qaeda, February 22. http://www.fbi.gov/newyork/press-releases/2010/nyfo022210.htm (accessed December 15, 2010).

Fernandez, Manny. 2007. One Had Stature and One Didn't, but Neither Seemed an Extremist. *New York Times*, June 4. http://www.nytimes.com/2007/06/04/nyregion/04suspects.html (accessed December 15, 2010).

Flaherty, Mary Pat, William Wan and Christian Davenport. 2009. Suspect, Devout Muslim from Va., Wanted Army Discharge, Aunt Said. *Washington Post*, November 6. http://www.washingtonpost.com/wp-dyn/content/article/2009/11/05/AR2009110505216.html (accessed December 15, 2010).

Ford, Beverly, and Helen Kennedy. 2009. Feds Arrest Mass. Man for Alleged Terror Plot on U.S. Malls. *New York Daily News*, October 21. http://articles.nydailynews.com/2009-10-21/news/17935542_1_tarek-mehanna-terrorism-probe-iraq (accessed December 15, 2010).

Gibbs, Nancy. 2009. The Fort Hood Killer: Terrified... or Terrorist? *Time*, November 11. http://www.time.com/time/magazine/article/0,9171,1938698,00.html (accessed December 15, 2010).

Griffith Colson Intelligence Service. 2011. http://gcis.us/ (accessed May 17, 2011).

Hamm, Mark S. 2008. Prisoner Radicalization: Assessing the Threat in U.S. Correctional Institutions. *National Institute of Justice Journal* 261, October 27. http://www.nij.gov/journals/261/prisoner-radicalization.htm (accessed December 15, 2010).

Horowitz, Craig. 2005. Anatomy of a Foiled Plot. *New York*, May 21. http://nymag.com/nymetro/news/features/10559/ (accessed December 15, 2010).

Hovington, Brett. 2010. Statement before the House Committee on Homeland Security, Subcommittee on Intelligence, Information Sharing, and Terrorism Risk Assessment, March 17. http://www.fbi.gov/news/testimony/working-with-communities-to-disrupt-terror-plots (accessed May 17, 2011).

Hsu, Spencer. 2010. Suspect in Metro Bomb Plot Pleads Not Guilty; Searches Find Weapons, Lectures by Radical Cleric Aulaqi. *Washington Post*, November 9. http://www.washingtonpost.com/wp-dyn/content/article/2010/11/09/AR2010110904805.html (accessed May 17, 2011).

———. 2009. FBI Says Massachusetts Man Sought Terrorist Training, Plotted Attacks. *Washington Post*, October 22. http://www.washingtonpost.com/wp-dyn/content/article/2009/10/21/AR2009102101088.html (accessed December 15, 2010).

Jenkins, Brian Michael. 2010. Would-Be Warriors: Incidents of Jihadist Terrorist Radicalization in the United States since September 11, 2001. RAND Corporation. http://www.rand.org/pubs/occasional_papers/2010/RAND_OP292.pdf (accessed December 15, 2010).

Johnson, Carrie, and Spencer S. Hsu. 2009. Terror Suspect Daniel Boyd Seemed to Have Typical Suburban Life. *Washington Post*, July 29. http://www.washingtonpost.com/wp-dyn/content/article/2009/07/28/AR2009072803193.html (accessed December 15, 2010).

Johnson, Dirk. 2009. Suspect in Illinois Bomb Plot 'Didn't Like America Very Much.' *New York Times*, September 27. http://www.nytimes.com/2009/09/28/us/28springfield.html (accessed December 15, 2010).

Kapralos, Krista J. 2009. One in 10 Inmates Behind Bars Turns to Islam. *The Herald*, February 1. http://www.heraldnet.com/article/20090201/NEWS01/702019905&news01ad=1 (accessed December 15, 2010).

King, Peter. 2011. Letter to Ranking Member Thompson Reaffirming Islamic Radicalization Hearings, February 8. http://homeland.house.gov/letter/king-letter-ranking-member-thompson-reaffirming-islamic-radicalization-hearings (accessed May 17, 2011).

———. 2011. Opening Statement: Hearing of the House Committee of Homeland Security: 'The Extent of Radicalization in the American Muslim Community and that Community's Response,' March 10. http://homeland.house.gov/sites/homeland.house.gov/files/03-10-11%20Final%20King%20Opening%20Statement_0.pdf (accessed May 17, 2011).

Kurzman, Charles. 2011. Muslim-American Terrorism since 9/11: An Accounting. University of North Carolina, Chapel Hill, February 2. http://sanford.duke.edu/centers/tcths/about/documents/Kurzman_Muslim-American_Terrorism_Since_911_An_Accounting.pdf (accessed May 17, 2011).

Los Angeles County Sheriff's Department. 2011. "Muslim Community Affairs Unit." http://la-sheriff.org/sites/muslimoutreach_new/index.html (accessed May 17, 2011).

Mazzetti, Mark, Sabrina Tavernise and Mark Healy. 2010. Suspect, Charged, Said to Admit to Role in Plot. *New York Times*, May 4. http://www.nytimes.com/2010/05/05/nyregion/05bomb.html (accessed December 15, 2010).

McFadden, Robert. 2009. Army Doctor Held in Ft. Hood Rampage. *New York Times*, November 5. http://www.nytimes.com/2009/11/06/us/06forthood.html (accessed December 15, 2010).

Meek, James Gordon, and Kenneth R. Bazinet. 2010. YouTube's Got to Gag Jihad Mouthpiece Anwar al-Awlaki: Rep. Anthony Weiner. *New York Daily News*, October 24. http://articles.nydailynews.com/2010-10-24/news/27079074_1_anwar-al-awlaki-adam-gadahn-umar-farouk-abdulmutallab (accessed May 17, 2011).

Melago, Carrie, Alison Gendar, Joe Kemp and Matt Lysiak. 2009. Newburgh Neighbors Stunned as Cops See Enough Hate to Kill in Terror Fiends. *New York Daily News*, May 21. http://www.nydailynews.com/news/ny_crime/2009/05/22/2009-05-22_meet_the_dangerous_terrorist_dummies.html (accessed December 15, 2010).

Napolitano, Janet. 2011. Understanding the Homeland Threat Landscape – Considerations for the 112th Congress, February 9. http://homeland.house.gov/sites/homeland.house.gov/files/02.09.11%20Sec.%20Napolitano%20Testimony.pdf (accessed May 17, 2011).

National Highway Traffic Safety Administration. 2011. Early Estimate of Motor Vehicle Traffic

Fatalities in 2010. http://www-nrd.nhtsa.dot.gov/Pubs/811451.PDF (accessed May 17, 2011).

———. 2010. Fatality Analysis Reporting System Encyclopedia. http://www-fars.nhtsa.dot.gov/Main/index.aspx (accessed May 17, 2011).

Pincus, Walter. 2006. FBI Role in Terror Probe Questioned. *Washington Post*, September 2. http://www.washingtonpost.com/wp-dyn/content/article/2006/09/01/AR2006090101764.html (accessed May 17, 2011).

Rashbaum, William. 2006. Staten Island Man Describes Shattered Life, Then a Plot to Bomb a Subway Station. *New York Times*, May 10. http://www.nytimes.com/2006/05/10/nyregion/10herald.html (accessed December 15, 2010).

Ratnesar, Romesh. 2011. The Myth of Homegrown Islamic Terrorism. *Time*. January 24. http://www.time.com/time/world/article/0,8599,2044047,00.html (accessed May 17, 2011).

Rayman, Graham. 2009. The Alarming Record of the F.B.I.'s Informant in the Bronx Bomb Plot. *Village Voice*, July 8. http://www.villagevoice.com/content/printVersion/1246457/ (accessed December 15, 2010).

Rubenstein, Andrew. 2010. King Salmon Couple Pleads Guilty in Terrorism Case. *Alaska Dispatch*, July 2. http://www.alaskadispatch.com/article/king-salmon-couple-pleads-guilty-terrorism-case (accessed December 15, 2010).

Schanzer, David, Charles Kurzman and Ebrahim Moosa. 2010. Anti-Terror Lessons of Muslim-Americans. Duke University, January 6. http://www.sanford.duke.edu/news/Schanzer_Kurzman_Moosa_Anti-Terror_Lessons.pdf (accessed December 15, 2010).

Seattle Police Department. 2011. "Muslim, Sikh, and Arab Advisory Council." http://www.seattle.gov/police/programs/advisory/MSA/default.htm (accessed May 17, 2011).

Shane, Scott. 2011. As Regimes Fall in Arab World, Al Qaeda Sees History Fly By. *New York Times*, February 27. http://www.nytimes.com/2011/02/28/world/middleeast/28qaeda.html (accessed May 17, 2011).

———. 2010. U.S. Approves Targeted Killing of American Cleric. *New York Times*, April 6. http://www.nytimes.com/2010/04/07/world/middleeast/07yemen.html (accessed December 15, 2010).

Shane, Scott, and James Dao. 2009. Investigators Study Tangle of Clues on Fort Hood Suspect. *New York* Times, November 14. http://www.nytimes.com/2009/11/15/us/15hasan.html (accessed December 15, 2010).

Shane, Scott, and Mark Mazzetti. 2010. Times Sq. Bomb Suspect Is Linked to Militant Cleric. *New York Times*, May 6. http://www.nytimes.com/2010/05/07/world/middleeast/07awlaki-.html (accessed December 15, 2010).

Shulman, Robin. 2007. The Informer: Behind the Scenes, or Setting the Stage? *Washington Post*, May 29. http://www.washingtonpost.com/wp-dyn/content/article/2007/05/28/AR2007052801401.html (accessed December 15, 2010).

SITE Intelligence Group. 2011. http://news.siteintelgroup.com/services (accessed May 17, 2011).

Stone, Brad. 2010. A Rapid Security Check Could Be Revived at Airports. *New York Times*, May 3. http://www.nytimes.com/2010/05/04/technology/04secure.html (accessed May 17, 2011).

Sulzberger, A.G., and William K. Rashbaum. 2010. Guilty Plea Made in Plot to Bomb New York Subway. *New York Times*, February 22. http://www.nytimes.com/2010/02/23/nyregion/23terror.html (accessed December 15, 2010).

Superior Court of Washington for King County. 2008. *State of Washington vs. Naveed Haq*, State's Trial Memorandum, February 21. http://www.investigativeproject.org/documents/case_docs/1150.pdf (accessed May 17, 2011).

Thomas, Pierre, Mary Walsh and Jason Ryan. 2003. Officials Search for Terrorist Next Door. *ABC News*, September 8. http://abcnews.go.com/WNT/story?id=129090 (accessed December 15, 2010).

United States Department of Justice, 2005. Four Men Indicted on Terrorism Charges Related to Conspiracy to Attack Military Facilities, Other Targets, August 31. http://www.justice.gov/opa/pr/2005/August/05_crm_453.html (accessed December 15, 2010).

United States Deputy Attorney General. 2004. Summary of Jose Padilla's Activities with Al Qaeda, May 28. http://fl1.findlaw.com/news.findlaw.com/cnn/docs/padilla/dod2doj52805sum.pdf (accessed December 15, 2010).

United States District Court for the District of Alaska. 2010. *United States of America vs. Paul Gene Rockwood, Jr.*, Sentencing Memorandum, August 16. http://www.investigativeproject.org/documents/case_docs/1362.pdf (accessed December 15, 2010).

United States District Court for the District of Maryland. 2010. *United States of America vs. Antonio Martinez*, Criminal Complaint, December 8. http://www.investigativeproject.org/documents/case_docs/1437.pdf (accessed December 15, 2010).

United States District Court for the District of Massachusetts. 2009. *United States of America vs. Tarek Mehanna and Ahmad Abousamra*, Superseding Indictment, November 5. http://www.investigativeproject.org/documents/case_docs/1121.pdf (accessed December 15, 2010).

———. 2009. Search and Seizure Warrant, Cases No. 09-120-LTS and 09-122-LTS, October 20. http://www.investigativeproject.org/documents/case_docs/1101.pdf (accessed December 15, 2010).

United States District Court for the District of New Jersey. 2007. *United States of America vs. Mohamad Ibrahim Shnewer*, Criminal Complaint, May 7. http://www.investigativeproject.org/documents/case_docs/386.pdf (accessed May 17, 2011).

United States District Court for the District of Oregon. 2010. *United States of America vs. Mohamed Osman Mohamud*, Criminal Complaint, November 29. http://www.investigativeproject.org/documents/case_docs/1426.pdf (accessed December 15, 2010).

United States District Court for the Eastern District of New York. 2007. U*nited States of America vs. Russell Defreitas, Kareem Ibrahim, Abdul Kadir and Abdel Nur*, Criminal Complaint, June 1. http://www.investigativeproject.org/documents/case_docs/419.pdf (accessed December 15, 2010).

United States District Court for the Eastern District of Virginia, Alexandria Division. 2003. *United States of America vs. Iyman Faris*, Statement of Facts, May 1. http://fl1.findlaw.com/news.findlaw.com/cnn/docs/faris/usfaris603sof.pdf (accessed December 15, 2010).

United States District Court for the Northern District of Illinois. 2010. *United States of America vs. Raja Lahrasib Khan*, Criminal Complaint, March 25. http://www.investigativeproject.org/documents/case_docs/1204.pdf (accessed December 15, 2010).

United States District Court for the Southern District of New York. 2010. *United States of America vs. Faisal Shahzad,* Plea, October 5. http://www.nypost.com/p/news/local/manhattan/read_the_faisal_shahzad_transcript_zDoUXlGEMoqZMwzsIRrlkM#ixzz14eZAo376 (accessed December 15, 2010).

United States District Court for the Southern District of New York. 2009. *United States of America vs. James Cromitie, David Williams, Onta Williams and Laguerre Payen*, Criminal Complaint, May 19. http://www.investigativeproject.org/documents/case_docs/952.pdf (accessed May 17, 2011).

Vrsansky, Nicole. 2008. 24 Year Old Derrick Shareef Sentenced to 35 Years in Prison. WIFR, September 30. http://www.wifr.com/news/headlines/29981564.html (accessed December 15, 2010).

Wilgoren, Jodi, and Jo Thomas. 2002. Traces of Terror: The Bomb Suspect; From Chicago Gang to Possible Al Qaeda Ties. *New York Times*, June 11. http://www.nytimes.com/2002/06/11/us/traces-of-terror-the-bomb-suspect-from-chicago-gang-to-possible-al-qaeda-ties.html (accessed December 15, 2010).

Wilson, Michael. 2009. From Smiling Coffee Vendor to Terror Suspect. *New York Times*, September 25. http://www.nytimes.com/2009/09/26/nyregion/26profile.html (accessed December 15, 2010).

Zhou, Momo. 2009. When One Bee Sting Is Your Last. *ABC News*, July 23. http://abcnews.go.com/Health/AllergiesNews/story?id=8148229 (accessed May 17, 2011).

7

Exchange Rate Volatility and Intra-Regional Trade in the East African Community

Mary Yang

The East African Community (EAC) has the opportunity to create a monetary union that would boost economic development for the region, and among the many objectives that the EAC hopes to achieve, increase intra-regional trade. A key component that would reduce transactions costs would be the creation of a single currency; the exchange rates for each country would be pegged to one another, such that volatility of exchange rate movements would no longer be an issue. Recent literature has indicated that exchange rate volatility has a negative impact on bilateral trade relationships. Following such literature, this paper estimates the impact that exchange rate volatility has had on intra-regional trade within the EAC using the gravity model. The main results of this paper indicate a fairly robust negative relationship between exchange rate volatility and trade, with impact estimates of a 13 percent or greater reduction in trade for a one standard deviation increase in volatility. Given these results, there is some evidence to suggest that the creation of a single currency would advance recent efforts toward fostering better trade relationships and integration between the members of the East African Community.

Mary Yang is a 2011 graduate of the Master in Public Affairs program at Princeton University's Woodrow Wilson School of Public and International Affairs, where she concentrated in economics and public policy. She can be contacted at myang@alumni.princeton.edu.

I. Introduction

The East African Community (EAC) has the opportunity to create a monetary union within the next few years in order to boost economic development, bring about enhanced global competitiveness, increase regional integration, and establish a single currency within the region. Consisting of Kenya, Tanzania, Uganda, Burundi, and Rwanda, the EAC has already made ardent strides towards this objective, with the re-establishment of the community in 2000 and the July 2010 launch of the Common Market Protocol, which expanded upon the region's existing customs union from 2005.

Despite advances made by the customs union and Common Market Protocol in promoting trade within the EAC, such as instituting duty-free trade between member countries and enacting a common external tariff on imports from third countries, intra-regional trade remains relatively low. Current exports from the three larger economies, Kenya, Tanzania, and Uganda, to the other EAC countries make up only 5-20 percent of total exports for each country.[1] Furthermore, research has shown that the EAC may not be the most suitable unit for trade integration, given that the five economies are not particularly complementary or competitive. Other obstacles remain, including under-investment in public goods such as regional infrastructure (especially in transport and power) and lack of regional cooperation in security and legal matters, all of which make it especially difficult for export sectors to grow.

Another transactions cost that may serve as a barrier to trade is the volatility of exchange rate movements between the five countries, each of which holds a different currency whose exchange values span a wide range. Recent literature has found that exchange rate movements have increased in magnitude due to shocks in the world economy, greater liberalization of capital flows, and an increase in the number of cross-border transactions.[2] These movements have been especially evident for countries with insufficiently developed capital markets lacking stable economic policies. While there has not been evidence of a negative relationship between volatility and trade at an aggregate worldwide level, some evidence of a negative relationship has been found on a bilateral level.[3]

With the prospect of the formation of an East African Monetary Union (EAMU), a key component that would reduce transactions costs would be the creation of a single currency. Similar to other regional currency unions, the exchange rates for each country would be pegged to one another, such that volatility of exchange rate movements would no longer be an issue.

This paper uses a model for bilateral trade to estimate the extent to which exchange rate volatility between the EAC countries has impacted intra-regional bilateral trade. If there has been a considerable negative impact to trade as a result of this volatility, then the EAC can hope to achieve significant regional trade gains once a currency union is in place.

II. Methodology and Data

The methodology of this paper follows the gravity model for estimating bilateral trade flows using a panel data approach. The gravity framework is drawn directly from the study of physics and postulates that trade between two countries varies as a function of the economic size (the countries' gross domestic products (GDP)) and the geographical distance between them. Since the model was first introduced by Tinbergen (1962) and Poyhonen (1963) to estimate bilateral trade flows among European countries, it has been widely used in empirical studies on international trade and is known for its strong fit to cross-country trade data. Moreover, the model's predictive ability has been justified in theoretical studies by Helpman and Krugman (1985), who derived a version of the gravity equation using sectors with homogeneous products that produced constant returns and sectors with differentiated products that produced increasing returns, and Bergstrand (1985, 1989), who developed microeconomic foundations of the model with various alternative assumptions, among others.

Since the early studies, empirical specifications of the gravity model have expanded to include a wide array of factors that may either increase or reduce trade. Among them are variables that affect transactions costs between two countries, such as cultural background, geographical factors, historical links, and preferential trading agreements. It is also acknowledged that the gravity equation can be useful for assessing the impacts of regional trade agreements. Studies such as Aitkens (1973) have included a dummy variable for intra-regional trade to capture trade between members, while others have tried using a second binary variable to capture the effects of extra-regional trade with non-members.

The focus of this study is the introduction of various measures of exchange rate variability within the gravity framework to examine the extent to which this may affect transactions costs and trade between the East African countries. If this volatility has had a significant negative impact on trade in the EAC, then the prospect of an East African Monetary Union and a single currency should increase trade flows among the five countries.[4]

The dataset is comprised of a panel of twenty bilateral country pairs

from 1970 to 2009. However, due to data availability, relatively few of the observations in the regressions come from the 1970s.

The benchmark specification is as follows:

$$\ln(X_{ij,t}) = \beta_0 + \beta_1 \ln(D_{ij}) + \beta_2 \ln(Y_{it}) + \beta_3 \ln(Y_{jt}) + \beta_4 Border_{ij} + \beta_5 Lang_{ij} + \beta_6 Col_{ij} + \beta_7 Rel_{ij} + \beta_8 V(e)_{ij,t} + \beta_{ij,t},$$

where *i,j* denotes the exporter and importer, respectively, *t* denotes time, and the variables are defined as:

Figure 1: Variable Definitions	
$X_{ij,t}$	The real value of exports from country *i* to country *j* at time *t*,[5] deflated by U.S. consumer price index (CPI) as exports are measured in U.S. dollars.[6]
D_{ij}	The geographical distance between country *i* and country *j*.[7]
Y_{it} and Y_{jt}	The real GDP of country *i* and country *j* at time *t*, respectively.[8]
Border	A binary variable equal to one if the two countries share a common land border.[9]
$Lang_{ij}$	A binary variable equal to one if the two countries share a common official language.
Col_{ij}	A binary variable equal to one if the two countries share a common colonizer.
Rel_{ij}	A binary variable equal to one if more than 70 percent of the two countries' populations share the same religion.
$V(e)_{ij,t}$	The volatility of the bilateral nominal or real monthly exchange rate, taken as the standard deviation in either one to five years prior to *t*.[10]

This specification takes into account various geographic, economic, cultural, and historical factors that may impact trade. Distance between the two countries would be expected to have a negative effect, as greater distance would give way to greater transactions costs such as transportation. Countries sharing a border should also boost trade and allow for increased cross-border transactions. Of the five EAC countries, only Tanzania shares a land border with each of its four neighbors. The GDP for each economy should have positive coefficients, since greater economic development is typically correlated with greater export supply and import demand.

The other explanatory binary variables describe the cultural and

historical closeness of the two countries, and all are expected to have a positive effect on trade. For official language, only Rwanda shares a common language with each of the other countries, as its official languages include both English and French. Tanzania, Kenya, and Uganda are all former British colonies whose official languages are English and Swahili, while Rwanda and Burundi are both former Belgian colonies that share French as an official language, among others. Countries sharing the same languages and former colonial rulers may have better established cultural connections and legal regulations for importing and exporting products. Another commonly used variable is common legal origin, though in this case, the legal origins for the EAC countries are derived from their former colonial rulers. The three former British colonies use a legal system based on English common law, while the two former Belgian colonies' legal systems are based on German and Belgian civil codes and customary law (CIA *World Factbook* 2011). The religious composition of these five countries vary, with Roman Catholicism, Protestantism, Islam, and indigenous religions making up the majority. Countries for which more than 70 percent of the population share the same religion as one of its neighbors are Kenya with Uganda and Rwanda with Uganda.

The variable of interest, $V(e)_{ij,t}$, is the volatility of the bilateral exchange rate, calculated as the standard deviation of the first-difference of the monthly natural log of the bilateral exchange rate in the years preceding year *t*. The long-run measure of the volatility calculates the standard deviation over the previous five years, whereas the short-run measure calculates the standard deviation over the previous year. The benchmark specification uses the official monthly end-of-period real exchange rates from the International Monetary Fund's *International Financial Statistics*, deflated by the consumer price index. To check the robustness of results, alternative measures of volatility are used, including the nominal official long and short-run exchange rate volatility.

Furthermore, parallel market rates, data for which come from Reinhart and Rogoff (2002), are used to calculate nominal and real, long and short-run exchange rate volatility. All of the East African countries maintained a tightly controlled exchange rate until the mid-to-late 1990s, when exchange rate liberalization occurred.[11] During that time, the parallel exchange rate market dominated price changes. In Tanzania, for example, the parallel rate was dominant from the late 1970s to 1985, and the parallel rate premium gradually tapered off thereafter and disappeared by 1993 (Rutasitara 2004, 2). The dataset is unfortunately limited, does not contain Rwanda in its sample, and only includes data until 1998, thus excluding the possibility

of using any country pair that includes Rwanda and reducing the overall sample size of the regressions. However, the time span of the dataset does cover the period in which parallel exchange rates were in effect in East Africa.

The benchmark specification used is pooled ordinary least squares (OLS). To control for unobservable trade barriers for each country, country-specific fixed effects (FE) are used for both exporter and importer countries. Time effects are also included in order to control for time-specific factors such as fuel shocks and food price shocks, both of which have significantly impacted the East African economies. A specification using both time and country fixed effects is also included.

Recent trade literature has called attention to time-varying country fixed effects to control for time-varying multilateral resistance.[12] However, this is not included because it may lead to overcorrection, since the specification would not be able to capture certain effects. This may occur if, for example, an unexpected increase in money supply led to an increase in the bilateral exchange rate volatility of that country. Even if this caused all bilateral trade to be reduced, the time-varying country fixed effects specification would still not be able to capture the exchange rate volatility's negative effect.

III. Possible Endogeneity of Exchange Rate Volatility

Up to this point, the baseline model has assumed that exchange rate volatility and trade are exogenous. However, this may not necessarily be the case, to the extent that countries may implement exchange rate policies that lower exchange rate volatility in order to increase trade. Although most of the countries now maintain a *de jure* floating exchange rate policy, in some cases the *de facto* classification has shown that the exchange rate is not completely a free float.[13] If such policies have been enacted, then the baseline model would suffer from endogeneity bias.

The possibility of endogeneity bias has been addressed in the economic literature using three different instrumental variable (IV) approaches. The first, proposed by Frankel and Wei (1993), uses the volatility in the relative quantity of money as an instrumental variable for exchange rate volatility.

Another approach has been proposed by Tenreyro (2003), who uses the sharing of a common exchange rate anchor as a proxy for exchange rate volatility. However, this method cannot be used in the context of the East African countries, since the anchor for all (when the exchange rates were under a controlled regime) was the U.S. dollar.[14]

A final approach, proposed by Devereux and Lane (2003), uses the fac-

tors explaining the theory of the optimal currency area such as labor and capital mobility as well as financial linkages such as balance sheet effects and the presence of external debt denominated in foreign currency to explain bilateral exchange rate volatility between trading partners. Unfortunately, this method is very data intensive and unable to be used here due to data constraints.

Since the Frankel and Wei (1993) approach is the only one possible, this paper made an attempt to follow its method, using as an instrument the relative money volatility, calculated analogously to the exchange rate volatility. However, monthly money data was scarce for most countries (and unavailable for Rwanda) and only available for several years in the recent past.[15] Whether due to data scarcity or other reasons, the first-stage least squares (1SLS) equation showed that relative money volatility was a poor proxy for determining exchange rate volatility. Thus, the two-stage least squares (2SLS) equations yielded unreasonable results and are excluded here.[16]

IV. Results

Tables I and II show the results of the gravity equation using official exchange rate volatility over five years and one year, respectively. Tables III and IV show the results using parallel exchange rate volatility over the long and short run.

Using official exchange rate volatility, the coefficient on distance is negative and statistically significant across all variations of the model and ranges from -0.25 to -1.7, which is consistent with previous empirical estimates. The coefficient on the exporter's real GDP is positive and significant, indicating that overall economic health is positively correlated with greater export supply. The coefficient on the importer's real GDP varies across the different regressions and seems to show that overall economic health does not necessarily correlate with greater import demand in the EAC countries.

The two variables which yield surprising results are border and official language. Both are slightly negative and significant in this case, suggesting that sharing a common land border and common official language does not necessarily lead to increased trade between those countries. Sharing a common land border may not help to increase trade in this case due to the lack of transport infrastructure in the EAC countries, and cross-border roads between the countries are currently very limited. Furthermore, some trade that occurs across borders may be informal and not captured by official data. The common language variable captures only common official

languages. Because there are many local and indigenous languages spoken, many of which are not designated as official languages of the country, the variable may not capture the full effect of sharing a common language.

The variables for common colony and common religion take on the expected positive and significant results for the most part, and sharing the same former colonial ruler especially seems to play an important role in increasing trade as it is positive and significant across all specifications. These results suggest that common rules and regulations, better historically established ties, and greater cultural closeness all play a role in augmenting trade between the East African countries.

The results using parallel exchange rates suffer from being drawn from a significantly smaller sample, especially those using long-run exchange rate volatility as an explanatory variable. Many of the results are similar to those in the regressions using official exchange rate volatility. However, the coefficients on GDP for both importer and exporter countries are positive and significant, suggesting greater import demand as well as export supply when GDP is higher. The border dummy is positive and significant in a number of specifications in Table III, but fail to show up in Table IV. Due to data scarcity, the variables for common colony and common religion were omitted from several of the regressions from lack of variation in the available data.

V. Effect of Exchange Rate Volatility on Trade

The results using official exchange rate volatility show a significant and negative impact of exchange rate volatility. The significant coefficients for long-run volatility range from -2.7 to -15.9, while they are generally lower on average for short-term volatility, ranging from -2.4 to -3.8. This can be interpreted as the effect of increasing volatility by one standard deviation around the mean. This impact can be calculated by multiplying the coefficient in the regression equation by one standard deviation of the volatility measure.[17] Using an average of all the coefficients, the impact of an increase in one standard deviation of long-run real volatility is a 14.3 percent reduction in trade flow, while the impact for long-run nominal volatility is 25.5 percent. For short-run measures, the reduction for a standard deviation increase in volatility is estimated to be 13.38 percent for real volatility and 14.29 percent for nominal volatility.[18]

These results suggest that the long-run impact of volatility on reducing trade is slightly larger on average than the short-run impact. Previous baseline estimates for this impact using a worldwide sample implied a trade reduction of 9 percent (Clark et al. 2004, 75). This would indicate

that exchange rate volatility has a greater cost to trade for the East African countries than it does, on average, for the world as a whole. The results using nominal versus real exchange rate volatility imply a greater cost to trade from nominal volatility, though the results are fairly similar on the whole. Indeed, the correlation between the nominal and real volatility measures are quite high, at 0.96 for the two long-run measures and 0.94 for the two short-run measures.

Turning to the results using parallel exchange rate data, there is evidence of a negative impact of volatility on trade, though this finding is not robust to the choice of estimation technique. In particular, positive (though insignificant) coefficients are found for long-run nominal volatility as well as short-run nominal and real volatility. As with the official rates, there is a high correlation between nominal and real parallel exchange rate volatility; the correlation between the two long-run measures are 0.99, and the correlation between the two short-run measures is 0.99.

Using the pooled OLS coefficient in column (1) of Table IV, the estimated impact of a reduction on trade from one standard deviation of exchange rate volatility is 25.28 percent.[19] Comparing this to the 13.38 percent impact found using official rates, this would suggest that parallel exchange rate volatility had a greater influence in lowering trade. However, one cannot make a direct comparison in this case, as the time period covered in the two samples varies greatly, with the official volatility measure inclusive of many more recent years' movements in exchange rate.

Moreover, it should be noted that the long-run volatility of parallel exchange rate estimations shown in Table III suffer from a very small sample size. The real exchange rate volatility estimations only have 74 observations in the sample, due to a combination of lack of monthly historical exchange rate and CPI data for some countries and the fact that standard deviations are taken over a much longer time period than the short-run volatility measures. The reduction impact for these estimates are not calculated or compared to the official rate volatility measures for this reason.

VI. Conclusion and Policy Implications

Motivated by the prospect of an East African Monetary Union in the near future and the creation of a single currency among its five members, this paper estimated the impact of exchange rate volatility on intra-regional trade in the EAC using the gravity model. Data was collected from 1970 to 2009 on intra-regional exports in a bilateral panel dataset, with information on distance, level of economic development, historical, and cultural linkages between each pair of countries serving as control variables and

determinants of bilateral trade. Monthly official and parallel exchange rates were used to calculate various measures of volatility.

Specifications of the gravity model used were pooled OLS, export and import country fixed effects, time fixed effects, and both country and time fixed effects. An attempt was made at an instrumental variable estimation using relative money volatility as an instrument for exchange rate volatility, though data constraints made the estimation difficult. This research could be improved upon given better data availability and quality. Other possible specifications which can be used would be to include time-varying fixed effects (though as mentioned previously, this may lead to overcorrection), as well as using country-pair fixed effects, which would allow one to control for unobserved cultural, historical, and geographical factors which are specific for a particular pair of countries.[20] Assuming data availability, a similar exercise could also be conducted on disaggregated trade data for particular sectors to see which export sectors were impacted most by exchange rate volatility.

The main results of this paper using official exchange rate volatility indicated a fairly robust negative relationship between exchange rate volatility and trade, with impact estimates of a 13 percent or greater reduction in trade. This negative coefficient was not, however, found to be significant in all specifications. The impact estimates were somewhat higher for long-term volatility compared to short-term volatility, and similar results were found between nominal and real exchange rate volatility. The estimated reduction in trade found here is slightly higher than those found in studies conducted on a worldwide sample, suggesting a higher cost to trade due to exchange rate volatility in the EAC countries than in other countries of the world on average.

Results for exchange rate volatility calculated from parallel market rates suffered from a small sample size, and the negative effect on volatility was not robust to all specifications. Results for other bilateral controls indicated a negative relationship for distance and trade, as well as a positive relationship for the export country's GDP and trade. Sharing a common colonial ruler and having a similar religious composition tended to increase trade between two countries. Surprisingly, sharing a land border and a common official language did not seem to indicate an improvement in intra-regional trade between two countries, and certain specifications suggested it may actually be a hindrance.

If it is the case that EAC countries implement exchange rate policies that lower exchange rate volatility to increase trade, then endogeneity may still be a concern due to the difficulty in finding a proper instrument. However,

given the results on exchange rate volatility in this study, there is some evidence to suggest that the creation of a single currency would further advance recent efforts at fostering better intra-regional trade relationships and integration between the members of the East African Community.

Moving forward, several policy implications can be drawn from these results. First, because intra-regional trade may be increased by reducing bilateral exchange rate fluctuations, the EAC countries should consider aligning and stabilizing the five currencies to move within a narrow band of each other in preparation for establishing a single currency. Similar to the European Exchange Rate Mechanism, the EAC countries could maintain a system in which bilateral rates are pegged within a narrow band in order to reduce exchange rate variability. This mechanism should facilitate a smoother transition to a single currency as well as reduce transactions costs in trade even prior to the full formation of a monetary union.

Secondly, the EAC countries would benefit from more research into the factors that affect intra-regional trade and bilateral exchange rate volatility. As such, enhanced data collection for the region would support these research efforts, and the EAC should consider adopting a standard mechanism across the five countries to improve consistency in data collection. In addition, these efforts should try to incorporate methods for capturing informal trade between countries in order to reduce measurement errors in further research.

Finally, in order to foster greater regional trade over the long term, the region would benefit immensely from more integrated transport networks and improved infrastructure between countries. Developing more roads and railways between countries would provide more direct routes for trading and reduce current time and costs of transport. More access to all areas of the region would also enhance diversification of products traded and boost intra-regional trade by augmenting complementarity in the regional economy. Moreover, a more integrated transport infrastructure would also allow the landlocked countries (Uganda, Rwanda, and Burundi) easier access to extra-regional trade networks.

The EAC has recently taken steps to cultivate greater integration and trade within the five economies by establishing a common market and customs union. Although hindered by data limitations, the evidence in this paper suggests that further gains in intra-regional trade can be made in the future by reducing exchange rate volatility with the formation of a single currency.

Table I. Long-Run Official Exchange Rate Volatility

	Real Exchange Rate				Nominal Exchange Rate			
	(1) Pooled OLS	(2) Country FE	(3) Time FE	(4) Country & Time FE	(5) Pooled OLS	(6) Country FE	(7) Time FE	(8) Country & Time FE
Log of distance	-0.3047	-1.2592***	-1.7062***	-1.3032***	-0.2553	-1.0530***	-1.5713***	-1.0583***
	(0.286)	(0.299)	(0.305)	(0.291)	(0.233)	(0.268)	(0.249)	(0.258)
Log of exporter's real GDP	1.5790***	2.7620***	2.2964***	3.9393***	1.4915***	2.4701***	2.2185***	3.4415***
	(0.127)	(0.390)	(0.139)	(0.606)	(0.098)	(0.221)	(0.122)	(0.373)
Log of importer's real GDP	0.0136	-2.1309***	0.7225***	-0.8941*	-0.0768	-1.6774***	0.6565***	-0.7357**
	(0.120)	(0.432)	(0.132)	(0.528)	(0.088)	(0.265)	(0.115)	(0.366)
Common border	-1.8059***	-6.6431***	-0.6971**	-0.5918**	-1.9575***		-0.9389***	
	(0.250)	(2.099)	(0.273)	(0.260)	(0.224)		(0.239)	
Common official language	-0.3643		-1.3971***	-11.3211***	-0.1925		-1.1435***	
	(0.266)		(0.293)	(3.093)	(0.234)		(0.242)	
Common colony	1.8643***	3.4221***	0.7040**	5.9926***	2.0504***	0.3968*	0.9591***	0.3563*
	(0.290)	(1.136)	(0.283)	(1.521)	(0.223)	(0.220)	(0.239)	(0.204)
Common religion	1.0740***	6.4908***	0.2250	0.5379	1.1041***	-0.2523	0.1376	-0.1607
	(0.310)	(2.064)	(0.329)	(0.391)	(0.245)	(0.256)	(0.262)	(0.244)
Exchange rate volatility	-7.1025	-15.9913***	-3.0133	-3.2357	-3.8916***	-6.2165***	-2.7157**	-3.3162**
	(4.466)	(4.293)	(6.231)	(5.617)	(1.045)	(1.063)	(1.327)	(1.381)
Constant	-16.9527***	15.1746*	-38.3500***	-34.6272*	-13.5951***	6.4677	-34.0887***	-31.5570**
	(3.578)	(8.405)	(3.954)	(18.217)	(2.454)	(5.618)	(3.361)	(12.961)
Observations	394	394	394	394	489	489	489	489
R-squared	0.6283	0.7481	0.7083	0.8019	0.6288	0.7373	0.6964	0.7858

Robust standard errors in parentheses

*** significant at 1%, ** significant at 5%, * significant at 10%

Table II. Short-Run Official Exchange Rate Volatility

	Real Exchange Rate				Nominal Exchange Rate			
	(1) Pooled OLS	(2) Country FE	(3) Time FE	(4) Country & Time FE	(5) Pooled OLS	(6) Country FE	(7) Time FE	(8) Country & Time FE
Log of distance	-0.4033*	-0.8674***	-1.5770***	-1.0132***	-0.3339	-0.7900***	-1.5505***	-0.8423***
	(0.232)	(0.284)	(0.234)	(0.267)	(0.219)	(0.262)	(0.237)	(0.250)
Log of exporter's real GDP	1.6203***	2.4427***	2.2447***	3.2978***	1.5707***	2.4134***	2.2222***	3.4791***
	(0.100)	(0.260)	(0.112)	(0.400)	(0.090)	(0.205)	(0.112)	(0.314)
Log of importer's real GDP	-0.0062	-1.2458***	0.6257***	-0.4738	-0.0654	-1.2566***	0.5961***	-0.3079
	(0.094)	(0.291)	(0.110)	(0.365)	(0.084)	(0.233)	(0.110)	(0.291)
Common border	-1.7789***		-0.7440***		-1.8385***		-0.8833***	
	(0.222)		(0.238)		(0.214)		(0.229)	
Common official language	-0.2313		-1.0540***		-0.2171		-1.0132***	
	(0.228)		(0.239)		(0.223)		(0.234)	
Common colony	1.7297***	0.4593*	0.6957***	0.3139	1.9264***	0.6234***	0.8795***	0.5311***
	(0.232)	(0.244)	(0.231)	(0.221)	(0.209)	(0.215)	(0.223)	(0.199)
Common religion	0.8360***	-0.2966	0.0614	-0.1285	0.7961***	-0.3760	-0.0958	-0.2278
	(0.253)	(0.267)	(0.247)	(0.246)	(0.212)	(0.247)	(0.229)	(0.232)
Exchange rate volatility	-1.3886	-2.4646**	-0.0252	-0.0576	-3.0008***	-3.7983***	-1.3886	-0.9771
	(1.306)	(1.129)	(1.673)	(1.387)	(1.154)	(0.951)	(1.628)	(1.230)
Constant	-17.1034***	-3.8936	-33.6100***	-34.4940***	-15.1367***	-3.5224	-33.3409***	-43.1041***
	(2.630)	(5.741)	(3.080)	(12.476)	(2.246)	(4.605)	(3.043)	(9.819)
Observations	481	481	481	481	524	524	524	524
R-squared	0.6283	0.7286	0.6970	0.7820	0.6323	0.7347	0.6921	0.7825

Robust standard errors in parentheses

*** significant at 1%, ** significant at 5%, * significant at 10%

Table III. Long-Run Parallel Exchange Rate Volatility

	Real Exchange Rate				Nominal Exchange Rate			
	(1) Pooled OLS	(2) Country FE	(3) Time FE	(4) Country & Time FE	(5) Pooled OLS	(6) Country FE	(7) Time FE	(8) Country & Time FE
Log of distance	-0.8710	0.6921	-0.9441	0.4674	-1.1785*	0.6244	-1.0209*	0.7220
	(0.864)	(1.159)	(0.793)	(0.949)	(0.608)	(0.664)	(0.600)	(0.617)
Log of exporter's real GDP	2.7825***	2.2091	3.9212***	4.3351	2.3037***	0.5966	2.6411***	2.5363
	(0.470)	(2.418)	(0.518)	(2.844)	(0.295)	(1.196)	(0.319)	(2.167)
Log of importer's real GDP	1.8818***	3.4854*	3.0300***	4.8253***	1.0478***	0.6421	1.3981***	2.2619*
	(0.508)	(1.749)	(0.542)	(1.364)	(0.376)	(0.909)	(0.351)	(1.266)
Common border	0.0808	0.9939	1.0761*	1.4794**	-0.1412	1.1213**	-0.1851	0.9186
	(0.555)	(0.703)	(0.579)	(0.614)	(0.579)	(0.553)	(0.582)	(0.582)
Common official language	-3.3198***		-6.3979***		-1.8631**	3.1589	-2.5245***	-1.7577
	(1.006)		(1.290)		(0.720)	(3.116)	(0.680)	(5.610)
Common colony	0.0000				0.0000			
	(0.000)				(0.000)			
Common religion	0.1525		-0.2648		0.5239		0.3358	
	(0.976)		(1.016)		(0.530)		(0.588)	
Exchange rate volatility	-11.2297	-12.5260	-57.8724***	-59.1522***	2.8627	-1.7576	4.1828	4.9426
	(15.615)	(14.766)	(18.414)	(19.548)	(3.355)	(2.544)	(4.445)	(4.105)
Constant	-79.4832***	-113.3557**	125.6112**	-180.7905**	-49.9549***	-16.8783	-65.7295***	-94.2947
	(24.486)	(49.875)	(24.040)	(69.105)	(16.784)	(20.888)	(16.387)	(62.691)
Observations	74	74	74	74	110	110	110	110
R-squared	0.5925	0.7036	0.6951	0.8104	0.6098	0.7779	0.6503	0.8113

Robust standard errors in parentheses

*** significant at 1%, ** significant at 5%, * significant at 10%

Table IV. Short-Run Parallel Exchange Rate Volatility

	Real Exchange Rate				Nominal Exchange Rate			
	(1) Pooled OLS	(2) Country FE	(3) Time FE	(4) Country & Time FE	(5) Pooled OLS	(6) Country FE	(7) Time FE	(8) Country & Time FE
Log of distance	-1.0968*	0.5442	-1.1621**	0.2742	-1.6790***	0.3151	-1.3876***	0.2411
	(0.571)	(0.606)	(0.578)	(0.598)	(0.583)	(0.589)	(0.518)	(0.528)
Log of exporter's real GDP	2.2647***	0.8427	2.5907***	3.5713*	1.8474***	0.8874	2.5725***	3.7849**
	(0.272)	(1.274)	(0.264)	(1.835)	(0.259)	(1.040)	(0.230)	(1.567)
Log of importer's real GDP	1.1190***	0.3722	1.4767***	2.3554**	0.5416*	-0.3975	1.3148***	2.0307**
	(0.336)	(1.156)	(0.305)	(0.975)	(0.323)	(0.968)	(0.270)	(0.839)
Common border	-0.6074	0.3972	-0.2929	0.6556	-0.5255		-0.3153	
	(0.532)	(0.576)	(0.544)	(0.585)	(0.542)		(0.526)	
Common official language	-1.5201**	2.9916	-2.2445***	-4.1979	-0.6317	2.7782	-2.1257***	-4.7731
	(0.657)	(3.127)	(0.628)	(4.690)	(0.661)	(2.572)	(0.586)	(4.035)
Common colony	0.0000				0.0000			
	(0.000)				(0.000)			
Common religion	0.1091		0.3455		0.0582	0.3237	0.1006	0.5975
	(0.581)		(0.602)		(0.517)	(0.550)	(0.521)	(0.551)
Exchange rate volatility	-0.5428*	-0.4113	0.3503	0.3137	-0.4452	-0.3495	0.3461	0.3265
	(0.316)	(0.371)	(0.610)	(0.461)	(0.322)	(0.355)	(0.590)	(0.441)
Constant	-50.7308***	-15.4176	-63.7356***	-113.0636**	-25.2413*	2.4061	-58.1801***	-109.5197***
	(14.794)	(22.895)	(13.759)	(47.631)	(14.002)	(13.473)	(11.523)	(39.399)
Observations	115	115	115	115	132	132	132	132
R-squared	0.5955	0.7128	0.6836	0.7894	0.5858	0.7429	0.6728	0.8123

Robust standard errors in parentheses

*** significant at 1%, ** significant at 5%, * significant at 10%

Notes

[1] Data from the International Monetary Fund's *International Financial Statistics.*

[2] See, for example, Prasad et al. (2003).

[3] See, for example, Clark et al. (2004).

[4] Similar studies have been done on the ASEAN free trade area showing that its formation has resulted in significant trade creation among its members. See, for example, Nguyen (2009).

[5] Data from the International Monetary Fund's *Direction of Trade Statistics.*

[6] Data from the World Bank's *World Development Indicators.*

[7] Data from the Centre d'Etudes Prospectives et d'Informations Internationales (CEPII)'s database on geodesic distances.

[8] Data from the World Bank's *World Development Indicators.*

[9] Data for all the binary bilateral descriptive variables come from the CIA's *World Factbook.*

[10] Official exchange rate data from the International Monetary Fund's *International Financial Statistics.* Parallel market data from Reinhart and Rogoff (2002).

[11] Liberalization occurred in 1992 for Uganda, 1993 for Tanzania and Kenya, 1995 for Rwanda, and 1999 for Burundi.

[12] This was initially proposed by Anderson and Van Wincoop (2003).

[13] See data from the International Monetary Fund's *Annual Report on Exchange Arrangements and Exchange Restrictions*, various years.

[14] The anchors can be found in Shambaugh's (2004) exchange rate regime classification dataset.

[15] Data from the International Monetary Fund's *International Financial Statistics.*

[16] Results from the 1SLS and 2SLS regression are available upon request.

[17] The standard deviation of long-run real and nominal, short-run real and nominal volatility for the observations used in the estimation sample are as follows: 0.0195, 0.0633, 0.0543, and 0.0624. Although the coefficients for long-run real exchange rate volatility in the left panel of Table I are on average larger than the coefficients for the other volatility measures, its standard deviation is considerably smaller than the other measures, and hence the impact on trade is not drastically larger by comparison.

[18] Using only an average of the significant coefficients in the regressions, the reduction impact is shown to be: 29.58 percent (long-run real volatility), 25.5 percent (long-run nominal), 13.38 percent (short-run real), and 21.21 percent (short-run nominal).

[19] The standard deviation of the short-run parallel real exchange rate volatility measure of the observations used in the estimation is 0.4657. For the long-run parallel real exchange rate volatility (using only observations in the estimation sample), the standard deviation is 0.0119.

[20] However, this specification would also lead to redundancy in many of the time-invariant bilateral variables included in the other specifications, such as distance, common border, colony, language, religion.

References

Aitkens, Norman D. 1973. "The Effect of the EEC and EFTA on European Trade: a Temporal Cross Section Analysis." *American Economic Review* 63 (5): 881-892.

Anderson, James E. and Eric van Wincoop. 2003. "Gravity with Gravitas: A Solution to the Border Puzzle." *American Economic Review* 93 (1): 170-192.

Bergstrand, Jeffrey H. 1985. "The Gravity Equation in International Trade: Some Microeconimc Foundations and Empirical Evidence." *Review of Economics and Statistics* 67 (3): 474-481.

Bergstrangd, Jeffrey H. 1989. "The Generalized Gravity Equation, Monopolistic Competition, and Factor Proportions Theory in International Trade." *Review of Economics and Statistics* 71 (1): 143-153.

Booth, David, Diana Cammack, Thomas Kubua, Josaphat Kweka, and Nichodemus Rudaheranwa. 2007. "East African integration: How Can it Contribute to East African Development?" Overseas Development Institute Project Briefing.

Central Intelligence Agency. *The World Factbook.* https://www.cia.gov/library/publications/the-world-factbook (accessed January 3, 2011).

Centre d'Etudes Prospectives et d'Informations Internationales. Geodesic distances database. http://www.cepii.fr/anglaisgraph/bdd/distances.htm (accessed January 3, 2011).

Clark, Peter, Natalia Tamirisa, Shang-Jin Wei, Azim Sadikov, and Li Zeng. 2004. "Exchange Rate Volatility and Trade Flows – Some New Evidence." International Monetary Fund Occasional Paper No. 235.

Devereux, Michael B., and Philip R. Lane. 2003. "Understanding Bilateral Exchange Rate Volatility." *Journal of International Economics* 60 (1): 109-132.

Dell'Ariccia, Giovanni, Julian di Giovanni, Andre Faria, Ayhan Kose, Paolo Mauro, Jonathan Ostry, Martin Schindler, and Marco Terrones. 2008. "Reaping the Benefits of Financial Globalization." International Monetary Fund Occasional Paper No. 264.

di Giovanni, Julian, 2005. "What Drives Capital Flows? The Case of Cross-Border M&A Activity and Financial Deepening." *Journal of International Economics* 65 (1): 127–149.

Frankel, Jeffrey A. and Shang-Jin Wei. 1993. "Trade Blocs and Currency Blocks." NBER Working Paper No. 4335.

Helpman, Elhanan and Paul Krugman. 1985. *Market Structure and Foreign Trade: Increasing Returns, Imperfect Competitions and the International Economy.* Cambridge, MA: MIT Press.

International Monetary Fund. *Annual Report on Exchange Arrangements and Exchange Restrictions,* 1970-2009.

International Monetary Fund. *Direction of Trade Statistics*, 1970-2009.

International Monetary Fund. *International Financial Statistics,* 1970-2009.

Nguyen, Trung Kien. 2009. "Gravity Model by Panel Data Approach: An Empirical Application with Implications for the ASEAN Free Trade Area." *ASEAN Economic Bulletin* 26(3): 266-277.

Poyhonen, Pentti. 1963. "A Tentative Model for the Volume of Trade Between Countries." *Welwirtschaftliches Archiv* 90 (1): 93-99.

Prasad, Eswar, and Kenneth Rogoff, Shang-Jin Wei, and M. Ayhan Kose. 2003. "Effects of Financial Globalization on Developing Countries: Some Empirical Evidence." International Monetary Fund Occasional Paper No. 220.

Reinhart, Carmen M., and Kenneth R. Rogoff. 2002. "A Modern History of Exchange Rate Arrangements: A Reinterpretation." NBER Working Paper No. 8963.

Rutasitara, Longinus. 2004. "Exchange Rate Regimes and Inflation in Tanzania." African Economic Research Consortium Research Paper 138.

Shambaugh, Jay. 2004. "The Effects of Fixed Exchange Rates on Monetary Policy." *Quarterly Journal of Economics* 119 (1): 301-352.

Tenreyro, Silvana. 2003. "On the Trade Impact of Nominal Exchange Rate Volatility." Federal Reserve Bank of Boston Working Paper 03-2.

Tinbergen, Jan. 1962. *Shaping the World Economy: Suggestions for an International Economics Policy.* New York: The Twentieth Century Fund.

World Bank. *World Development Indicators*, 1970-2009.

8

Between Market Blip and Municipal Bloodbath: Understand Short- and Long-Term Crisis Avoidance in the Municipal Bond Market

Katie Cristol

This paper synthesizes analysis on current and future challenges in the municipal market, addresses structural problems of regulation and disclosure, and offers policy solutions advanced by experts in the field. It addresses features of the current market that challenge the overheated predictions about short-term crisis; namely, contained risk of default, price insensitive supply, and macro stability. But this review of the long-term threats to municipal credit-worthiness – pension and retiree health benefit liabilities, erosion of backstops, and bankruptcy risk of "private public debt" – suggests that these elements of the municipal bond landscape, coupled with structural failings in the marketplace, pose a threat in the mid- to long-range future. This paper argues that the municipal bond market bloodbath is neither imminent nor inevitable, but state, local, and federal governments must take responsibility for maintaining this critical source of funding for capital projects for future generations.

Katie Cristol is a 2011 graduate of the Master in Public Affairs program at Princeton University's Woodrow Wilson School of Public and International Affairs, where she focused on child and family issues and state fiscal policy. She can be reached at kcristol@gmail.com.

I. Introduction

Recent media coverage of the municipal debt market reads like reviews of a horror film: an unavoidable "bloodbath" (The Economist 2010) with municipal market mutual funds "hemorrhaging cash" (Seymour 2010), and "frightened money fleeing the market" (Seymour 2011), while "bond defaults stalk" citizens as state legislators "try to stem the bleeding" (Preston and Green 2010).

But these headlines overplay the immediate crisis in municipal debt while simultaneously underplaying the larger long-term threats to municipal bonds as a source of revenue for major infrastructure projects in states and localities. A better understanding of how the changing muni market affects state and local finance requires a more thorough investigation. This paper synthesizes reporting and expert analysis on current and future challenges in the municipal market and addresses structural problems of regulation and disclosure. It concludes by sharing policy solutions advanced by experts in the field.

First, a distinction: this paper largely addresses the primary market for municipal bonds, in which issuers sell bonds to investors – institutional or private – rather than the secondary market, in which buyers purchase after issuance, from other bondholders. The focus here is on the challenges facing the municipal bond market that directly relate to, and result from, municipal management, rather than the recent severe losses in secondary market value for municipal bond funds. Such losses have largely resulted from investor uncertainty about the expiration of federal programs such as income tax cuts, the 2009 American Recovery and Reinvestment Act, and Build America Bonds (Seymour 2010 and RBC Capital Markets 2010). Of course, declines in the value of the secondary bond market have consequences for municipal bond issuance and management, so this emphasis on the primary market reflects only the focus of this paper, not a dismissal of these challenges.

What is a Municipal Bond? The Simplified Model

Municipal bonds, at their most basic, "represent a promise by state or local governmental units (called the issuers) or other qualified issuers to repay to lenders (investors) an amount of money borrowed, called principal, along with interest according to a fixed schedule" (Wesalo Temel 2001, 1). The process of issuing a municipal bond generally begins with a state or local government identifying a capital need, such as infrastructure projects or undeveloped land to be purchased as a protected area. The government

then works with a financial advisor, who assists in preparing the terms of the bond and the prospectus, and with a municipal bond dealer, who underwrites, markets, and sells the new bond.

Variations on the Theme

As public finance has grown increasingly complex in recent decades, so has the potential for variation in the process described above. As Joseph Fichera, CEO of financial advisory firm Saber Partners stated, "the words 'municipal bond' mean a lot of things to a lot of different people," and a savvy observer of the municipal bond market must shed the assumption that everything in that market is the same (Fichera 2010). The variety in types of municipal bonds and manner in which they are issued is significant; there are a few distinctions that are critical to understanding the recent and long-term threats to the muni market.

General Obligation (GO) vs. Revenue Bonds – The fundamental distinction in municipal bonds is the source of repayment. General obligation bonds are backed by "the full faith and credit" and the taxing authority of a state or local government. In many states, the voting public must approve these bonds (National Counsel of State Legislatures 2010). By contrast, revenue bonds are backed by the profits (revenues) of the project or enterprise that the bond is designed to fund. According to former New Jersey Treasurer Clifford Goldman, investors view full faith and credit bonds as slightly more secure investments, although there are also very highly rated revenue bonds (Goldman 2010).

Municipality as Debtor vs. Municipality as Conduit – The traditional model of a municipal bond involves a government issuing its own debt for its projects. In recent years, however, conduit bonds have grown in market share. With conduit bonds, the government issues debt on behalf of another entity, such as a university or an industrial corporation (Walter 2009). Securities and Exchange Commissioner Elisse Walter explains, "conduit bonds have their name because the municipality acts as a 'conduit' through which investors lend money to a third party and the third party repays the debt. The credit of the third party supports the conduit bond, and the debt is generally not considered a liability of the municipality" (Walter 2009). Fichera estimates that conduit bonds or private activity bonds currently represent about 10-20 percent of the municipal bond market. Conduit bonds should not be confused with contract bonds, in which a state or local government works with an authority such as a development or transit authority to issue bonds and the authority serves as the debtor. Contract bonds do not require a public vote, which has evoked criticism

that such programs circumvent the public will.

Locality vs. State – The municipal descriptor can refer either to states or to localities such as cities, townships, or school boards. The structural differences between state and local government have some effects on the bonds that they issue. For example, Fichera notes, "local governments are creatures of the state governments, and rating agencies really like state governments because [states] have a lot of flexibility," relative to localities, because they can pass on budget troubles (Fichera 2010). Illustrating the agencies' preferences for state over local debt, a recent report from rating agency Standard & Poor's (S&P) suggests that "unlike states, local governments often have limited authority over their revenues, debt, and payment provisions, [and] their growth often doesn't keep pace with spending increases because restrictions on tax levies and assessed property value limit the degree to which governments may benefit from market value growth" (Standard & Poor's 2010).

II. The Current Landscape

As previously discussed, the coverage of the municipal bond market in recent months has been characterized by rhetoric of meltdown and market collapse. Yet a number of features of the market, including contained risk of default, price insensitive supply, and macro stability, suggest that short-term threats have been overstated. However, threats to the long-term health of the municipal debt market have been generally underemphasized. This is particularly true of pension and healthcare liabilities and the erosion of traditional backstops in this market.

Part I: Why Short-Term Worries are Overwrought

Contained Risk of Default

First, it is important to note that the default crisis is currently only anticipatory. As a recent *New York Times* article noted, "municipal bankruptcies or defaults have been extremely rare — no state has defaulted since the Great Depression, and only a handful of cities have declared bankruptcy or are considering doing so" (Cooper and Williams Walsh 2010). Indeed, many rating agencies and financial advisors argue that there will be few defaults coming. Analysts with the investment group RBC Capital Markets point out that "default risk in the municipal market has remained historically low" (RBC Capital Markets 2010). An end-of-year S&P report argued that despite vulnerabilities in the municipal sector and heated political rhetoric, the threats of default remain low; their internal analysis suggested that the

majority of state and municipal governments were very likely to survive the recession without defaulting on their debts (Standard and Poor's 2010).

According to former New Jersey Treasurer Goldman, a major reason for the rarity of defaults, as well as the relative confidence of some parties involved closely in the market, is the priority of bond debt for municipal revenues. The *New York Times* article similarly highlights that "states and cities typically make a priority of repaying their bond holders, even before paying for essential services" (Cooper and Williams Walsh 2010). This is because states and cities recognize the damage that a missed bond payment would do to their ability to raise funds in the municipal debt market at any point in the future.

Price Insensitive Supply

One issue that is largely overlooked in recent media coverage of volatility in the municipal market is that, regardless of investor behavior, state and local governments are largely unconcerned with market prices. Goldman emphasizes that "public entities are not sensitive to interest rates" in the same way as private companies or consumers, and that "if you need to do a project, you don't judge by interest rates, [rather,] you look at it as, I need the money now, interest rates could just as well go up as down, so let's do our business" (Goldman 2010).

In other words, issuance in the municipal market is unlikely to be affected by changes in interest rates. Even if investors demand higher interest rates to cover new uncertainties in a historically stable market, it is unlikely that states and localities will respond by changing the amount of debt that they would otherwise issue. Though the price insensitivity is not limitless, Goldman estimates plenty of room in the borrowing rate to absorb market shocks before governments stop issuing bonds, because interest rates are currently very low by historic standards (Goldman 2010).

Macro Stability

Contrary to notions of a second mortgage collapse – similar to the one advanced recently by *The Economist* – both of the previous points suggest that the risk of the municipal market bottoming out in the near term is low. As Goldman explains, "[there are] serious problems in states like New Jersey… and a serious problem in some of the cities, but I don't think those problems lead to a collapse in the municipal bond market for the simple reason that, in general, debt service costs are a small percentage of any government's budget – even the ones that over-borrow." Unlike the private financial sector, estimates suggest that overleveraged municipalities

are borrowing around 7-9 percent of their budget revenues, compared to an ideal 4 percent (Goldman 2010).

Regarding macro stability, it would be inaccurate to suggest that recent fluctuations in the prices of municipal bonds do not have consequences for state and local budgeting. However, this volatility affects the cost of financing projects and not the creditworthiness or borrowing potential of these governments. Losses or anticipated losses in the value of municipal bond funds in late 2010 and early 2011 have largely been driven by the expiration of the federal Build America Bonds (BABs) program, which was created by the 2009 American Recovery and Reinvestment Act. Under the program, municipalities could issue taxable bonds with a 35 percent rebate on their interest costs, such that a BAB had a lower borrowing rate than an ordinary, tax-exempt bond. The higher-interest BABs allowed municipalities to extend their reach into those areas of the investment market that demanded higher yields. Eager governments issued a total of $181 billion in BABs during the program's run from February 2009 to December 2010.

With the expiration of the program, *Bloomberg* has noted, "Investors have expressed concern that traditional tax-exempt debt issuance might surge next year. That may push up the interest-rates investors demand to hold tax-exempt bonds (Selway and McGrail 2010). This would increase borrowing costs for municipal governments. Since municipal governments are price-takers, the supply of municipal debt is relatively price-insensitive. Further, historically low interest rates leave room for absorbing shocks. As a result, the volatility surrounding the expiration of BABs may spell temporary unpleasantness for investors and governments, but is unlikely to lead to large-scale exit from the market for either party.

Part II: Underemphasized Long-Term Threats

The prior section illustrates why current volatility in the municipal market does not spell imminent catastrophe for state and local governments' ability to raise capital. However, the greater problem is the impact of the growing pension and health benefit liabilities, or underfunded retirement and health benefit systems for municipal government employees, on state and local budgets. Goldman, whose generally positive view on the state of the market was outlined above, qualifies that "the only thing I can envision that would trigger any jeopardy is the pension liability" (Goldman 2010).

Pension and Health Care Liabilities

Among investors and experts, concerns are growing that these pension

and health benefit liabilities can jeopardize debt payments. Arthur Levitt, former Securities and Exchange Commission (SEC) chairman, argues, "The balance sheets and income statements of governmental authorities are a mess. The growing crisis in public pensions, which are underfunded by as much as $3 trillion, is already beginning to shake up municipal bond investors, who recognize the risk of default even if offering documents papered over those liabilities" (Levitt 2010). Without comprehensive public accounting, there is no way of knowing the extent of these liabilities, but a 2010 study by the Pew Center on the States found that states had collective unfunded pension, healthcare, and other retiree benefit liabilities of $3.35 trillion at the end of fiscal year 2008. The study further concluded that only four states had fully funded pension systems, while many systems were underfunded by more than 30 percent (The Pew Center on the States 2010, 3). New Jersey, which was recently sued by the SEC for failing to disclose the extent of its liabilities, has one of the more extreme records on underfunding pensions, but many other states have deep troubles from years of insufficient or missed pension payments that were exacerbated by the economic crisis.

Disconcertingly, Goldman predicts that many state pension funds will run out of money during the life of bonds currently in the market, and notes that this possibility has been largely undisclosed. He asserts that, if a state's pension fund may go broke between a bond's issuance and maturity, there should be an obligation to disclose that to investors. Currently, rating agencies such as Standard and Poor's assure their investors that "pension reform movements [are] underway in certain states…[and] while we believe these liabilities represent true long-term pressure on government credit quality, they generally aren't immediately competing for most governments' capacity to fund their debt service or meet their other priority payment obligations" (Standard and Poor's 2010).

Yet, without large-scale pension reform, according to Goldman, states with growing liabilities will eventually reach the point where cutting *all* discretionary spending will still be insufficient to meet these deficits. At this point, the pension and health care liabilities will threaten bond repayment. Although general obligation bonds will still have to be paid – by raising taxes if necessary – the fate of revenue bonds is less assured under these circumstances. Many market analysts had assumed that the recent crisis would impose fiscal discipline on governments out of necessity (Fichera 2010). However, despite increasing attention to the issue, few lessons seem to have been learned about meeting pension obligations, and these growing liabilities on state and local balance sheets could have consequences for the payment of bond debts.

Erosion of Backstops

Fichera, a financial advisor with Saber Partners notes that most of the time, "you'll ultimately get your money" as a bondholder, even when municipalities struggle to pay back their debt, because of intervening entities like higher levels of government or insurers. This has been true historically, especially at the local level, to such an extent that there are concerns of moral hazard for cities and towns resulting from state insurance. Examples of this implied state insurance for localities include the recent Pennsylvania bailout of the town of Harrisburg, and promises such as those found in a recent S&P investor report about Detroit, which noted that "Michigan has repeatedly indicated to Standard & Poor's that it would take all steps necessary to prevent a [city financial] manager from filing for bankruptcy protection" (Standard & Poor's 2010). But poor fiscal practices, such as unmet pension and health care obligations, are undermining states' capacity to backstop localities. In fact, the U.S. Senate Budget Committee hearings in January 2011 raised the specter of states petitioning the federal government for their own bailouts. Senator Joe Manchin (D-WV), a former governor and Chair of the National Governor's Association, suggested that Federal Reserve Chairman Ben Bernanke consider "the possibility that states will default on their muni bonds" (Hume and Ackerman 2011).

This fraying safety net has been further weakened by the disappearance of the bond insurance that governments formerly bought to increase their credit. These insurers "used to cover roughly half of the market, [but] have retrenched or gone bust after making bad mortgage bets" (The Economist 2010). With traditional backstops missing, Fichera suggests that there could be concern about "cliff risk" – the potential for a worst-case scenario, or a bottoming out of the demand for any municipal debt, even backed by financially healthy governments. The capital markets "are brutal reactionaries" and a small but critical mass of defaults could collapse the market (Fichera 2010).

Bankruptcy and "Private Public Debt"

A final concern is the extent to which threats of default within the municipal bond market can actually be private and not governmental default. The conduit bonds described earlier have higher yields and are thus attractive to investors looking to increase returns on municipal bond funds, but conduit funds are backed by private entities and not the state or municipality. In short, the presence of such conduit bonds in the municipal bond market exposes the market not only to the risk of governments failing to meet payments, but also to potential defaults when private firms go bankrupt.

III. Structural Troubles for the Municipal Bond Market

Despite these storm clouds on the horizon, some state and local debt will continue to be well-managed and safe for investors. In the municipal market as well as private financial markets, credit risk, which is carried by the individual issuer or bond, is separate from market risk, or price fluctuations are caused by general shifts in municipal interest rates. But in municipal markets, credit and market risk are linked much more closely than they otherwise would be because failures of regulation and rating have left investors without the tools to discern good municipal debt risk from bad.

In other words, many of the concerns surrounding the mid- to long-range future of municipal bonds have been compounded by a lack of transparency. Not only are the aforementioned threats to state and locality fiscal health rarely reflected in the prospectuses or public information that municipal governments release, but the SEC also cannot demand this information nor compel its release. Further, after the dramatic failures of the large rating agencies to accurately assess vulnerabilities in private securities during the 2008-2009 market crash, investors and experts have increasingly begun to doubt the reliability of the agencies' ratings of municipal debt.

The next section is a more extensive review of the limitations on SEC regulation and the failures of the rating agencies. It is followed by a discussion of the consequences of the lack of accountability and the potential for changes in oversight and disclosure.

Limited SEC Regulatory Authority

By legislative design, the SEC has far less oversight of the municipal bond market than private markets. A number of recent SEC Commissioners, including Arthur Levitt, Elisse Walter, and current Commission Chair Mary Schapiro, have called for greater SEC authority over the municipal bond market. But the SEC's power in this area is constrained by congressional failure to overturn Section 15B(d)(1) of the 1975 Securities Exchange Act, commonly referred to as the Tower Amendment. The Amendment, originally designed to protect states from federal overreach, prohibits the SEC and the Municipal Securities Rulemaking Board (MSRB), which can issue, but not enforce, regulations for the municipal bond market, from requiring municipal bond issuers to file securities documentation before sale. According to Walter, "the MSRB is further limited in its ability to require any municipal issuer to furnish it or any purchaser or prospective purchaser with *any* documents" (Walter 2009).

This circumscribed authority has real consequences for the municipal bond market. In August 2010, Arthur Levitt evinced strong opinions on Congress' failure to give the SEC more authority to regulate municipal bonds "in spite of evidence that this market is rife with the hallmarks of abuse, [such as] poor disclosure, little regulatory oversight, made-to-order accounting rules and insider deals driven by banker and consultant fees" (Levitt 2010). Even apart from the abuses, experts question why state and local governments face far fewer requirements for disclosure than private companies (Goldman 2010). Currently, the market for municipal bonds operates under much less enforceable oversight than the private securities market.

Ineffectual Credit Rating Agencies

Secondly, credit rating agencies such as Standard & Poor's, Moody's, and Fitch have taken a severe credibility hit as a result of their widespread failures to detect or communicate the vulnerabilities of securities – particularly collateralized debt obligations – that soured during the recent financial crash. These failures also extend to the municipal bond market. According to Walter, "the record of credit rating agencies in recent years has led both retail and institutional investors to focus more closely on the disclosure documents of municipal issuers rather than merely relying on a bond rating" (Walter 2009).

Fichera explains that rating agencies "take things as a given," and cannot root out fraud or dangerous budgeting practices of municipal governments (Fichera 2010). He also explains that the agencies have traditionally had different standards and staff to rate municipalities, making it hard to develop consistent ratings throughout the market. This lack of common language is exacerbated by the fact that accounting standards for governments are neither as rigorous as they are in the corporate world, nor as timely. Municipalities may issue financial data later than private entities, so there is a time lag between evaluation, reporting, and the debt currently being traded in the market.

Even if the rating agencies improved their internal evaluation systems, or demanded more compliance with accounting standards, Fichera points out that there is still an inherent challenge to assessing the risk of municipal debt investments. "One of the biggest problems is how do you evaluate political risk" of taxpayer revolt or the election of fiscally irresponsible executives and legislators? Finally, Fichera believes that media coverage is unlikely to increase transparency regarding municipal bonds because coverage comes from political, rather than financial, reporters who are less likely to ask questions of financial specificity (Fichera 2010).

Consequences of Lack of Accountability

A lack of transparency and resulting asymmetries of information are dangerous not only for investors, but for issuers as well. Because there is so much time between the typical municipal bond's issuance and maturity (average maturity in the municipal bond market is around fifteen to twenty years), much of the bond's perceived value depends on projections about a government's financial future. Fichera highlights the phenomenon of "headline risk," or investor skittishness that results from political pundits' suppositions about bankruptcy and taxpayer revolts (Fichera 2010). Other experts have identified this skittishness as a driver of the losses in the municipal bond mutual fund market in late 2010 and early 2011. "Many analysts agree that the incessant drumbeat of prognosticators foretelling of widespread municipal insolvencies has taken a toll on demand," writes *The Bond Buyer*. The article quotes one investment manager who claims that "many clients of ours are nervous about it, and I'm sure we're not the only ones…there's plenty of scared money left in the market. I think there's still going to be flows out because of all the headline risk" (Seymour 2011). If more information were available about the creditworthiness of municipal governments, investors may be less likely to conflate the credit risk of a few headline-making states and localities with overall market risk.

Positive Trends in Regulation and Disclosure

There are some promising signals that the lack of regulation and disclosure might be challenged in months to come. For example, in August 2010, the SEC filed a fraud case against the state of New Jersey, claiming that the state failed to disclose to investors the liabilities of its two biggest pension plans when it sold $26 billion in bonds between 2001 and 2007. Fichera and Goldman both note that the SEC can get involved in municipal securities in circumstances of extreme fraud, and it would appear that the SEC is increasingly asserting this authority. According to Bloomberg, "the [New Jersey] case is the first SEC fraud charge against a state and follows the creation of a unit set up this year to focus on municipal securities and pension funds" (Seymour 2011).

Encouragingly, in light of Fichera's pessimism about the lack of media oversight for the market, the SEC case in New Jersey was largely seen as the result of a 2007 *New York Times* exposé. (Published April 4, 2007, the front-page exposé was authored by Mary Williams Walsh and entitled "NJ Pension Fund Endangered by Diverted Billions"). Furthermore, the Municipal Securities Review Board announced in late 2010 that it would form a commission to study pre-transaction pricing efficiency and

liquidity in the municipal market. MSRB Executive Director Lynnette Hotchkiss declared, "The study will help the board evaluate whether pricing mechanisms and liquidity in the market could be improved with higher levels of pre-trade price transparency" (Ackerman 2010). The MSRB announced that "specifically, [the study] will review transaction costs, price dispersion, and other market data to help the MSRB assess whether the market is operating as efficiently and fairly as possible" (Ackerman 2010). More aggressive regulatory action from the MSRB and the SEC, as well as increased media scrutiny, may be able to head off deleterious effects of pension liabilities and other fiscal mismanagement. However, congressional action and repeal of the Tower Amendment is still needed to bring true transparency to the municipal bond market.

Alternately, amid crisis-level talk on Capitol Hill about state bankruptcy, some more members of Congress are discussing legislation to mandate disclosure regarding pensions as a pre-condition to issuing municipal debt. According to the *The Bond Buyer,* congressmen at a February 9, 2011 hearing expressed that forcing states' hands on disclosure is a more practical and moderate step than legislation permitting states to file for bankruptcy protection. Representatives Devin Nunes (R-CA), Darrell Issa (R-CA), and Paul Ryan (R-WI) have introduced the Public Employee Pension Transparency Act, which "would impose reporting requirements on public-employee pensions. State and local governments that failed to comply with the requirements would forfeit their ability to issue tax-exempt bonds or receive federal subsidies on taxable municipal debt, such as Build America Bonds" (Quigley 2011). Senators Richard Burr (R-SC) and John Thune (R-SD) also plan to introduce an identical measure in the Senate.

IV. Averting a Crisis: Policy and Practice for Default Prevention

Beyond greater transparency and regulation as outlined above, additional policies should be adopted to strengthen the health of municipal bonds and their issuing governments. In addition to the priority of meeting retirement pension and health benefit obligations, some easily implemented practices surrounding debt issuance and management are outlined below.

Qualified Bonds

Aside from California's troubles, many of the news-making near-defaults in the municipal markets recently have come from locality issuers, not states. The interconnectedness of state and local budgeting, and the mechanism

of state aid to localities, provides an opportunity to support the solvency of locality-issued debt. During his time as State Treasurer in New Jersey, Goldman introduced a qualified bond system in which state aid to localities goes directly to the trustees of the municipal bonds. With the qualified bond system, the remainder of the aid goes to the city or town only after the debt service has been paid. Under the program, all municipal bonds for the major financially troubled cities in New Jersey have maintained high ratings despite other budgetary problems. In fact, an August 2010 announcement from the Fitch rating agency "re-affirmed its AA- rating on bonds issued by New Jersey's Municipal Qualified Bond Act, a mark of approval for the state's bond program" (Fitch 2010).

Goldman describes the qualified bond system as a "front end" way for the state to manage its obligations to municipalities, rather than stepping in when default is imminent. Though the program has remained under the radar, it continues to work well. Localities obtain better ratings and interest rates, as well as security, and because "they won't be permitted to borrow when their total debt isn't able to be covered by the state aid" (Goldman 2010). Consequently, the policy has prevented the worst cases of over-borrowing and mismanagement. Given New Jersey's positive experience, the qualified bond program could be a replicable policy for state governments looking to support their localities in issuing debt without creating the moral hazard of promising a backstop in the case of near-default.

Treating Conduit Bonds as Corporate Bonds

As previously mentioned, conduit bonds are arrangements where municipalities issue debt on behalf of private entities without backing them; as such, conduit bonds introduce private bankruptcy risks into the municipal market. Consequently, one good practice would be to treat the conduit bonds as the private debt that they are. Commissioner Walter has proposed:

> Commercial entities responsible for debt obligations under a conduit borrowing arrangement should be subject to the same level of disclosure obligations as a corporate issuer directly obtaining financing in the public securities markets. The fact that the bonds are tax-exempt does not change the fact that these are private obligations in which investors look to a private entity for repayment. (Walter 2009)

Until greater disclosure or oversight standards for the entire municipal market are introduced, closing this loophole for private debt treated like public debt would help better protect the municipal bond market from defaults related to private bankruptcy.

Accountability for Small Municipalities

For state-level debt, tighter standards may be necessary for accountability among seasoned state finance professionals. In small municipalities, the greater concern may be insufficient expertise, not intentionally misleading practices. In August 2010, a *Wall Street Journal* article about troubled debt in smaller towns quoted a California-based financial advisor that "small-town governments are usually made up of a few professional administrators and part-time politicians....They don't get into these [bonds] on their own. Municipal bankers will advise the city on the practicality of the bond, and you've got attorneys drawing up the statement and reviewing them, and a lot of these [politicians] don't have any experience with municipal bond finance" (Swarts 2010).

Levitt's proposals for regulation and fraud prosecution in the municipal markets provide one policy solution for these cases:

> Take on the bond advisers, not just the issuers. In every fraudulent bond deal, there are underwriters, legal counsel, auditors and a cast of others whose advice paved the way. Let them bear some of the responsibility when bonds are issued in bad faith. Charge them alongside the municipal authorities, and have them share in the fines. (Levitt 2010)

On December 20, 2010, the SEC announced a proposal for a new registration process for municipal advisors. The proposed rule would require municipal advisors "to submit more detailed information than is currently required and certify that they have met or will meet the qualifications and regulatory obligations required of them," forcing disclosure of any past disciplinary issues and making more information available to municipalities that may engage the services of these advisors (SEC 2010).

Though Levitt's proposal and the proposed SEC Permanent Rule are designed to reign in advisors in all cases – not just those of small localities – higher accountability standards for the advisors would likely impose greater fiduciary responsibility for the underwriters and attorneys working with less experienced local leadership.

V. Conclusion

Ultimately, these types of limited policy fixes must be paired with more fundamental reforms in how municipal debt is regulated and how municipal governments disclose financial and accounting practices. More importantly, better municipal financial management, particularly regarding retiree pension and health benefit obligations, is integral to sustaining a

healthy marketplace that can be relied upon for financing capital projects in the years to come. The "bloodbath" has neither arrived nor is inevitable, but state and local governments must take responsibility for removing the suspense and horror from their budget process.

References

Ackerman, Andrew. 2010. MSRB to Study Pre-Transaction Pricing. *The Bond Buyer*, December 8.

American Banking News. 2010. Press Releases: Fitch Approves New Jersey's Municipal Qualified Bond Act. August 10.

Cooper, Michael and Mary Williams Walsh. 2010. Mounting Debts by States Stoke Fears of Crisis. *New York Times*, December 4. www.nytimes.com/2010/12/05/us/politics/05states.html (accessed January 19, 2011).

The Economist. 2010. Municipal Bonds: the Mortgage Parallel. November 25. http://www.economist.com/node/17581914 (accessed January 19, 2011).

Fichera, Joseph (Senior Managing Director & CEO, Saber Partners, LLC), interviewed by phone, December 6, 2010.

Goldman, Clifford (Managing Partner, Goldman, Beale Associates; Former Treasurer, State of New Jersey), interviewed by phone, December 1, 2010.

Hume, Lynn and Andrew Ackerman. 2011. Senators Seek Ways to Assist Failed States. *The Bond Buyer*, January 10.

Levitt, Arthur. 2010. Drawing Blood Still to Come in Muni Market. *Bloomberg*, August 19.

McGee, Patrick. 2010. November Boosted by BABs. *The Bond Buyer*, December 1.

The Pew Center on the States. 2008. *The Trillion Dollar Gap: Underfunded State Retirement Systems and the Road to Reform*. Washington, DC: Pew Charitable Trust.

Preston, Darrell and Jeff Green. 2010. Bond Defaults Stalk Michigan's Wealthiest as Home Prices Crash. *Business Week*, June 23.

Quigley, Joan. 2011. Lawmakers Seek Pension Disclosure: Republicans Eye State Bankruptcy. *The Bond Buyer*, February 10.

Di Re, Remo and Erik Bresnahan. 2010. *A Focus on State and Local Government Municipal Debt*. Investor Report. New York, NY: RBC Capital Markets, LLC.

Securities and Exchange Commission. 2010. SEC Proposes Permanent Rule Requiring Municipal Advisors to Register With Agency. Press Release. Washington, DC: SEC. December 20.

Selway, William and Dunstan McNichol. 2010. SEC Muni-Bond Inquiry May Net Others After New Jersey. *Business Week*, August 19.

Selway, William and Brandon A. McGrail. 2010. Build America Bonds Program's End Poised to Batter Municipal-Debt Market. *Bloomberg*, December 10.

Seymour, Dan. 2010. Muni Mutual Funds Keep on Bleeding. *The Bond Buyer*, December 13.

—. 2011. Mutual Fund Cash Exodus Continues. *The Bond Buyer*, January 10.

Standard and Poors, 2010. *U.S. States And Municipalities Face Crises More Of Policy Than Debt*. New York, NY, November 8.

Swarts, Will. 2010. Tiny Bonds Wreaking Huge Local Havoc. *Wall Street Journal Smart Money*, August 25.

Temel, Judy Wesalo. 2001. *The Fundamentals of Municipal Bonds, 5th Edition; The Bond Market Association.* New York: John Wiley & Sons.

Walter, Commissioner Elisse B. 2009. Regulation of the Municipal Securities Market: Investors Are Not Second-Class Citizens. 10th Annual A. A. Sommer, Jr. Corporate, Securities and Financial Law Lecture. New York, New York. October 28.

Williams Walsh, Mary. 2007. NJ Pension Fund Endangered by Diverted Billions. *New York Times*, April 4.

http://www.nytimes.com/2007/04/04/nyregion/04pension.html (accessed January 19, 2011).

9

Entrepreneurship in Low-Income Countries: The Case of the Angolan Enterprise Program

Selina Carter

The Angolan Enterprise Program (AEP) is derived from C. K. Prahalad's Bottom-of-the-Pyramid (BOP) approach, which identifies the informal sector as a source of entrepreneurship. The formal private sector can profitably harness these entrepreneurs through microfinance and venture capitalist programs. The AEP presents a case in point on ways that the BOP approach can be successfully implemented. Angola represents a promising BOP testing ground because of the country's massive informal sector and recently liberalized economy. This paper makes critical observations of the AEP at its seven-year mark, concluding with recommendations for its future implementation and, more broadly, the applicability of the AEP as a model for other low-income countries.

Selina Carter is a third year graduate student in the Master in Public Administration, Economics, and International Relations Programs at the Maxwell School of Citizenship and Public Affairs at Syracuse University. She focuses on public-private partnerships, global markets, and international development policy. She can be reached at shcart01@syr.edu.

I. Introduction

Despite the billions of dollars in foreign aid that flow into developing countries each year,[1] these contributions will not have a lasting impact unless low-income countries develop a small-business sector.[2] According to United Nations (UN) sources and many development experts, the primary source of jobs in developing countries is small and medium-sized enterprise (UN 2004; Malloch 2004; Rivzi 2004; Versi 2006). Helping poor people run their own businesses is the key component to sustained and equitable economic growth. Most traditional foreign aid misses this point because it is donor-centric and fails to stimulate private-sector initiatives. Economist William Easterly notes that from 1965 to 1995, very few countries experienced economic growth subsequent to substantial aid flows (Easterly 2002, 37-9).

C. K. Prahalad's Bottom-of-the-Pyramid (BOP) approach is an emerging paradigm in development theory that attributes market potential to the roughly 4 billion people with annual incomes of less than $1,500 (Prahalad 2005, 4). Prahalad and others[3] assert that the poor represent a new consumer base for existing businesses, which can tap the low-income consumer market through high-volume and low-cost products. Microfinance is one component of this framework (*Ibid* 2005, 24). Given the opportunity, the poor represent an entrepreneurial revolution waiting to happen (*Ibid* 2005, 11). Since the majority of entrepreneurs in the developing world sustain themselves through the informal sector, harnessing this energy through private-sector action and public-sector stewardship is both pragmatic and profitable for all players. Some estimates of the informal assets in developing countries reach $9.4 trillion, which currently fall far short of their full productive use, as these resources can potentially be used as collateral in the banking system and thus spur entrepreneurial activity (UN Commission on the Private Sector and Development 2004, 9). For example, domestic banks in developing countries can potentially develop poor-friendly products such as microfinance programs and venture capitalist projects and by funding small business management schools.

This article presents the framework of the emerging BOP paradigm and analyzes its application in the case of the Angola Enterprise Program (AEP).[4] The AEP is a grassroots entrepreneurship program formed in 2004 through a partnership among the United Nations Development Program (UNDP), Government of Angola, and Chevron Corporation. Because of the country's massive informal sector and recently liberalized economy, Angola's case presents a vital learning opportunity for ways to implement

the BOP approach in practice. After making critical observations of the AEP at its seven-year mark, this article offers recommendations for its future implementation both locally and in other low-income countries (LICs) that house similar projects.

II. Implications of the BOP Paradigm for Public Sector Policy

In 2004, the UN Commission on the Private Sector and Development, of which Prahalad was a key member, published a major BOP public policy guide. Its report, *Unleashing Entrepreneurship: Making Business Work for the Poor*, advises LIC public authorities to welcome small-scale entrepreneurs by "eliminating artificial and policy-induced constraints to strong economic growth" and "engaging the private sector" through partnerships to facilitate access to financing options, promote skill development through business training institutions, and deliver basic services such as energy and water (UN Commission on the Private Sector and Development 2004, 13). Multinational corporations (MNCs) should be engaged in this process, particularly as partners to local companies with the goal of spurring innovation and technical skills for business owners (*Ibid* 2004, 3). In line with the BOP approach, the report states that successful approaches "rely on market mechanisms and private sector incentives" and hence lend themselves to the widespread, sustained success of entrepreneurs (*Ibid* 2004, ii). Moreover, few MNCs currently serve BOP markets because of institutional barriers such as lack of public infrastructure, property rights, skilled labor force, and formal capital markets (Webb et al 2010, 555). Alliances with NGOs or non-profit organizations such as the UNDP can facilitate these corporations' entry into BOP markets by linking their products and services with diverse stakeholder groups and removing barriers to entry (*Ibid*, 555).

The core of the BOP approach is the creation of "enabling environments" in LICs that allow entrepreneurs to thrive (UN Commission on the Private Sector and Development 2004, 2). Within many LIC governments, there is a need to reduce bureaucracy and fees for business registration, remove taxes on new businesses, and encourage private sector banks to create microfinance and "business incubator" institutions for entrepreneurs to become competitive. Governments should strengthen the rule of law, create transparent and effective regulatory oversight mechanisms, and stabilize the economy through prudent fiscal and monetary policies. They should also insert basic business skills into their national education systems by engaging private sector and NGO experts to create a business-

centric curriculum (*Ibid*, 3).

Most importantly, however, the BOP approach encourages governments to create partnerships with private-sector companies in order to create mutually advantageous projects that treat the poor as value-conscious consumers and creative entrepreneurs (Prahalad 2005, 1). Such partnerships have the potential to facilitate access to financing options, promote skill and knowledge development through microenterprise training institutions, and deliver basic services such as energy and water (UN Commission on the Private Sector and Development 2004, 3). Existing companies can be persuaded to view BOP markets as profitable, for example, by producing small unit packages in high volume or by financing start-up businesses. According to Prahalad, BOP borrowers do not represent a higher risk than top-of-the-pyramid borrowers (*Ibid*, 12), a perspective shared by Nobel Prize-recipient Muhammad Yunus, who created the Grameen Bank in Bangladesh (Nelson Mandela Foundation 2009). The growth of the microfinance sector also implies that a long-term supervisory role for governments is necessary: the November 2010 default crisis among Indian microfinance institutions is a reminder that, like mainstream financial institutions, governments must properly regulate BOP business practices in order to ensure prudence among lenders (Flintoff 2011).

The BOP approach also welcomes the efforts of MNCs to help reduce poverty with their superior resources and technology, particularly by tapping into the entrepreneurial spirit of the informal sector. Given their experience in private-sector activities, MNCs such as Chevron can be persuaded to lend expertise to domestic companies on ways to develop new poor-friendly products. Google and IDEO, a business design consultancy firm, are other examples of MNCs that can spur profit-driven projects to help poor people. NGO actors can also play an important intermediary role; a notable example is the Global Impact Investing Network (GIIN), a not-for-profit organization that informs profit-driven investors on methods for solving social and environmental problems (GIIN 2011). In order to elicit the expertise of these companies, development practitioners must leap beyond ideological boundaries that assume that all profit-seeking entities are untrustworthy or dangerous to poor peoples' interests. While existing Corporate Social Responsibility (CSR) programs are generally praiseworthy from a charitable standpoint, MNCs can be made most useful to development by disseminating their technical expertise and remodeling their business strategies to serve BOP markets.

Moreover, it is important to note that the BOP version of market-oriented development differs from the neoclassical or neoliberal economic theories

which have steered the Bretton Woods institutions since the 1970s.[5] First, the BOP approach seeks to convene public and private capabilities, and has its chief leadership within the UNDP, not the World Bank or the International Monetary Fund (IMF) (Prahalad 2005, xvi).[6] Second, the BOP approach promotes domestic-centric development, emphasizing "bottom-up" domestic entrepreneurs as opposed to "trickle-down" foreign direct investment (UN Commission on the Private Sector and Development 2004, 9). In addition, the BOP approach treats informal-sector actors as entrepreneurs whose energy should be harnessed, rather than surplus labor that will eventually join the formal sector as wage-earners (*Ibid* 2004, 8). Accordingly, this approach views economically marginalized groups such as women and the illiterate poor as valuable sources of creative energy that can spur private sector diversification. Finally, the BOP approach recognizes the benefits of foreign direct investment as a principle catalyst for technological innovation, but only if MNCs buy in to the BOP approach, target the poor as part of their overall corporate strategy, and help domestic industries do the same (*Ibid*, 9).[7]

III. Business Culture in Angola: A Testing Ground

Angola, a country of eighteen million people on Africa's southwest coast, represents an archetypal case of post-colonial economic transition. After independence from Portugal in 1975, Angola endured a thirty-year civil war costing 500,000 lives and displacing four million people. The fighting, loss of infrastructure, and central-planning tactics by the government perpetuated weak economic performance and human development. In the 1990s, inflation rose to 12,000 percent, and chronic fiscal and monetary imbalances led to low investment in social programs (Government of Angola and UNDP 2003, 3).

After the war's end in 2002, Angola's overall GDP growth rate soared to 21 percent in 2005 (World Development Indicators 2011). Paradoxically, the country has since gained low-middle income status with the World Bank, faring better than most of its neighbors (World Bank Data and Statistics 2009). However, most of this economic gain owes to Angola's massive oil and diamond industry;[8] oil revenue contributes to 85 percent of the country's annual GDP (CIA World Factbook 2009). With a Human Development Index score of 0.564, Angola is only slightly ahead of its Sub-Saharan African neighbors. The country has the highest infant mortality rate in the world, the highest overall death rate, and the lowest life expectancy, at just over thirty-eight years (*Ibid* 2009). Tersely stated by

the UN Millennium Development Goals Monitor, Angola is "off track" for eradicating extreme poverty and hunger (MDG Monitor 2009).

Unsurprisingly, new businesses face immense obstacles in Angola. The government previously outlawed mark-ups of more than 25 percent to final consumers, while profit taxes were effectively 47 percent. This has made the informal sector relatively more profitable than the formal sector, since formal business registration exposes businesses to official tax codes and other stringent regulations on start-ups (Government of Angola and UNDP 2003, 4). In 2004, about 75 percent of economically active Angolans worked in the informal sector, possibly the largest proportion of any other country in Sub-Saharan Africa (Barsky 2004). Around 82 percent of operators in the informal sector are self-employed, and most are women; 47 percent of women in Angola are involved in the informal sector, as opposed to only 27 percent of men (Government of Angola and UNDP 2003, 4). The government's earlier attempt to rescue the informal sector was to enforce complex business-registration procedures, which only created additional costs for entrepreneurs such as taxes, rents for market spaces, and extralegal charges paid to low-level public officials (*Ibid* 2003, 6).

Furthermore, many Angolan entrepreneurs are underserved by Angola's deficient vocational and educational system. Given that many entrepreneurs lack vocational skills and basic literacy – as of 2001 only 54 percent of women and 83 percent of men in Angola were literate – the provision of practical, tailored training to informal sector actors is crucial to raising their productive potential (CIA World Factbook 2009). The scarce vocational training centers that existed as of 2001 were all concentrated in Angola's capital city Luanda and focused on larger clients. This left most vocational training to informal apprenticeship, which has been of limited scope and availability (Government of Angola and UNDP 2003, 5).

Regarding access to credit, most private-sector banking services are concentrated in Luanda and tailor to larger clients. Micro-credit services are few, but the growing returns on this type of investment promise new opportunities for major banks (*Ibid* 2003, 5). In 2004, the National Bank of Angolan, the country's central bank, estimated that there were only 400,000 bank accounts countrywide, which accounted for roughly 6 percent of the population. With two working individuals per family and about two million families, the bank's estimates showed that there were $100 to $150 million in untapped capital within the banking sector. The Director of Research and Statistics of the bank stated, "if this could be drawn upon it would mean real potential for micro and small-scale credit" (Barsky 2004, 2). According to a report by Deloitte and Abanc, Angola's

bank association, there is significant potential for the retail-banking business in Angola. The percentage of the population using bank services could be about 20 percent, instead of its current 6 percent. In neighboring medium-income countries in sub-Saharan Africa, the average is 25 percent, and in South Africa's exceptionally vibrant economy it reaches 46 percent (Amaral 2008). This comparison shows that Angola has a large potential for private sector-driven access to financial services, which represents a major opportunity for domestic entrepreneurs.

Despite these abundant obstacles, the government's response to its post-war reconstruction challenge has been ambitious. Angola approved its national poverty reduction strategy[9] in 2004, and recently updated its targets for 2009-2013 (UNDP (b)). Its main priorities are "the creation of jobs, training of the labor force, the development of private initiatives, a significant increase of public investment, particularly in the social sector, and rehabilitation of infrastructure" (Government of Angola and UNDP 2005, 10). These goals align with the UNDP's strategy to promote small-scale entrepreneurs based on the BOP approach.

IV. The Angolan Enterprise Program: The BOP Approach in Action?

The AEP is an ongoing project initiated in 2004 by the UNDP, the Government of Angola, and Chevron Corporation, a United States-based MNC and Angola's largest oil producer with a fifty-year presence in Angola. The program's aim is to "promote the development of a diverse, robust, micro, small and medium enterprise sector in Angola" in order to generate employment and raise incomes (Government of Angola and UNDP 2003, 1). The AEP's core budget for the 2004-2006 period was $4 million, of which the UNDP provided $1 million and Chevron provided $3 million (UNDP Midterm Review 2006, 6). The project has since been extended to 2011 with additional support from the Spanish government, although the budget has not yet been clarified (interview with Ferrari dos Santos April 2011). Chevron's $3 million contribution to the AEP is a component of its Angola Partnership Initiative worth $25 million, which supports the UNDP as well as USAID projects (Chevron Corporate Responsibility Report 2008, 6). Although these investments are treated as "corporate responsibility" initiatives, many of Chevron's projects are financially self-sustaining and profitable. For example, Chevron's NovoBanco, a microfinance institution in which the company retains a 14 percent share, attracted $19 million in new deposits in 2009, and has expanded to eight branches since its founding in 2004 (Macauhub 2008;

Cabinda Gulf Company Limited 2009, 12). Given Chevron's recent capitalization of its business and banking expertise, its union with the UNDP and the Angolan government through the AEP represents an opportunity to critically evaluate best practices for the BOP approach.

The heart of the AEP is the equipping of the Angolan government with a coherent legal and institutional framework to encourage the micro- and small-business sector, as well as creating financially self-sustaining microfinance units for new businesses. The AEP's strategy consists of four main objectives: 1) create an "enabling environment" for entrepreneurs through public sector institutions; 2) strengthen vocational training programs for the small-business market; 3) expand the supply of micro and small business credit; and 4) introduce pilot models of business development service providers (Government of Angola and UNDP 2003, 6-7).

The main overseer of the AEP is the Steering Committee, which meets semi-annually. Representing Angolan institutions, this committee consists of appointed representatives from eight Angolan ministries,[10] the National Bank of Angola (BNA), Angola's Network of Micro-finance Institutions (RASM), and the Angolan Women's Entrepreneurs Federation (FMEA). International program advisors from Chevron and UNDP are also part of the Steering Committee (*Ibid* 2003, 14).

V. Outputs of the AEP: Public Sector- Versus Private Sector-Driven Schemes

According to the UNDP, the "on paper" outputs of the AEP by the end of 2010 include: twenty-five Business Development Services (BDS) centers in five provinces, market studies of the informal sector, vocational training needs assessments, counsel to the government on means of assisting new businesses, entrepreneurial training workshops, the creation of several microfinance institutions (MFIs), and a research unit within the Angolan Catholic University (UCAN) on small-business and informal-sector activities (UNDP(b)). The Business Incubator has trained 2,832 entrepreneurs, creating 117 new jobs. As of 2010, 3,200 students have taken the entrepreneurship course, and eighty secondary school teachers were trained in small-business education (*Ibid*).

Although these figures give some sense of the quantitative effectiveness of the AEP, qualitative results are arguably more relevant to understanding what challenges the program has faced. The Midterm Evaluation in 2006 was conducted in conjunction with outside auditors. The review consisted of a qualitative data collection through focus groups, interviews, and questionnaires with seventy-three stakeholders, which included AEP

administrators, representatives from the Angolan government, and outside consultants from NGOs such as ACCION and Development Works. The purpose of the review was to identify "key learning" about the process and direction of implementation qualitatively, rather than "assessing performance or impact" through quantitative observations alone (UNDP Midterm Review 2006, 6 and 13). Overall, the Midterm Review indicated that the AEP has helped small businesses by *initiating* the change process and raising awareness about the poverty-alleviating benefits of microfinance (*Ibid* 2006, 10). "Awareness" programs included media campaigns and outreach to local micro-finance institutions (MFIs), commercial banks, the central bank, NGOs, and the business community.

Given the AEP's recent implementation, it is premature to judge the project's long-term impact on poverty reduction. However, it is possible to observe patterns in the program's general direction even at its early stages. These patterns point to an underlying structure that could reveal its adherence or non-adherence to the BOP approach. In what ways is the AEP based on domestic private sector-driven opportunities for entrepreneurs? In what ways has it strayed from this strategy?

The project outputs can be categorized into three types: those which are public sector-centric, those which are private sector-centric, and those that essentially float between the two sectors. The first two output types, which target public and private actors, seek to improve the environment for entrepreneurs from different directions, both of which are necessary for a comprehensive, system-wide change in business culture in Angola. On the one hand, the AEP has targeted the government to reduce bureaucratic obstacles to entrepreneurs such as taxes on new businesses, which is an imperative reform considering the low economic position of most informal-sector entrepreneurs (*Ibid* 2006, 12). The private sector-driven outputs likewise represent crucial aspects of sustained financial and technical support. Examples of these outputs include the successful microfinance development unit through the National Bank of Angola and self-funding business incubators through commercial banks (*Ibid* 2006, 26). Business incubators are programs sponsored by governments, economic development institutions, universities, banks, or other entities that help entrepreneurs develop the experience, organizational structure, and contacts to launch successful start-up companies. Both public and private output types are intrinsic to the BOP approach (UN Commission on the Private Sector and Development 2004, 2).

The third output type, categorized by floating entities and programs that have no substantial linkages with the public and private realms, are

not thriving. An example is the initial stages of the Business Development Services (BDS) centers, which as of 2006 were languid. As of 2006, these centers survived from UNDP and donor funds, and represented a dead end for the AEP unless they can be underpinned with private-sector linkages and public-sector stewardship (UNDP Midterm Review 2006, 30).

The initial two "pilot model" BDS centers offered free educational programs for entrepreneurs. By 2010, their number grew to twenty-five such centers in five separate provinces (UNDP(b)). The original two centers are located in Luanda and act as business consultants, offering business-plan expertise, linkages with existing enterprises, access to international markets and foreign investment, and business incubator services for start-up companies (*Ibid* 2006, 20). The BDS centers were designed to be "managed in the same way as a private company" in that they could self-sustain their funding sources, recruit "motivated and highly professional staff," and assess results based on quantitative criteria (Government of Angola and UNDP 2003, 46).

In actuality, the initial BDS centers themselves were not managed like businesses. The staff admitted that they lack basic administrative skills such as business-plan preparation, accounting, monitoring and evaluation, and resource mobilization independent of the UNDP. Stakeholders complained that the centers "mostly waited for clients to come to them" and did not ambitiously advertise their services (UNDP Midterm Review 2006, 20). Stakeholders also indicated that training through the BDS Centers was "too long" and their "operation [was] not practical" (*Ibid* 2006, 15). In addition, the BDS centers did not provide clients with follow-up education after the initial training workshops. "Even the clients that received services from [the BDS Centers] were not happy about the quality of these services" (*Ibid* 2006, 20). Finally, the UNDP was often late in delivering salaries and administrative funding to the BDS Centers (*Ibid* 2006, 17).

Notably, these stakeholders indicate that the BDS component of the AEP was untenable. Given the quantity and nature of these flaws, they were symptoms that pointed to a larger question about the initial design of the AEP. Specifically, while the BDS centers were meant to operate like real businesses, they initially maintained a donor-dependent resource base, limiting their creative mobility. Instead of building a business-consulting model based on competitiveness and entrepreneurial spirit – the very values that the BDS centers should have imparted on clients – the AEP created donor-dependent, disengaged entities that were effectively severed from local private-sector linkages, relevant public-sector institutions, or grassroots realities. Rather than "increase the dynamism and competitiveness of the

domestic enterprise sector," the BDS centers essentially floated between sectors without a clear strategy for penetrating either the public or private spheres (Government of Angola and UNDP 2006, 46).

The experience of the BDS centers contrasts with the AEP's remarkably successful private-sector initiatives, the second output type. Specific examples include a microfinance-development unit in partnership with the National Bank of Angola and two microfinance institutions: Banco Sol (a commercial bank) and Development Works (an NGO) (*Ibid* 2003, 9 and 26). The AEP also built the first business incubator in Luanda, graduating six incubated companies in 2009. These particular companies created thirty job posts, and some of them have gained access to the credit of $40,000 through the support from the incubator. The AEP continues to work with major commercial banks in "downscaling" products to BOP markets (UNDP Angola (a)).

Although separate from the direct auspices of the AEP, Chevron's actions in the area of private sector-driven development models are likewise promising. Partnering with USAID and the International Finance Corporation (the private sector arm of the World Bank group), Chevron helped to establish NovoBanco in 2004, a microfinance institution that provides small-scale entrepreneurs and low-income households with access to finance. NovoBanco has expanded to three branches with approximately 30,000 clients and $27 million in net assets and plans to expand its operations by adding eleven new branches. In 2008 alone, it gave more than $10 million in loans to help Angolan entrepreneurs (Chevron Corporate Responsibility Report 2008, 30; IFC 2004; USAID 2009). In 2008, Angola's African Investment Bank (BAI)[11] acquired an 85.7 percent stake in NovoBanco, making it the lead manager. Chevron Sustainable Development Company retains the remaining 14.3 percent. These developments indicate that NovoBanco has been successful both financially and socially (Macauhub 2008).

These latter outputs of the private sector-driven classification serve as a reminder that programs grounded in profit-driven packages often represent more viable models than traditional, donor-funded entities, especially those that float between public and private realms with no concrete linkages with outside actors. The heart of the BOP approach stresses the need to move away from traditional donor-funded initiatives because they promote donor dependency and do not address the underlying problem: creating self-sustaining business opportunities for the poor. Instead, the private sector should be a central player in fostering grassroots entrepreneurs, while governments should ensure prudence and foresight among

profit-driven players, as well as making sure that products are designed with the interest of the poor in mind. The experience in Angola suggests that donor-funded entities are useful for short-term administrative and coordination purposes – such as the AEP coordination unit within the UNDP – but whose comparative advantage is not to autonomously harness creative entrepreneurship within a country's domestic private sector. The donor-dependent model forces otherwise knowledgeable actors – such as Chevron – to relegate their poor-friendly services to a traditional philanthropic, patron role, causing them to take a forced, uninspired, and somewhat paternalistic position based on "social responsibility," rather than market potential. That is, the BOP approach avoids charity as a model for serving poor people (Prahalad 2005, 3). Chevron's highly successful NovoBanco initiative highlights the market potential that MNCs can help create by offering informal-sector entrepreneurs products that suit their needs, rather than focusing on traditional charitable CSR programs. The AEP should not stifle this energy by reverting to traditional donor-centric traps.

Additionally, public-sector actors in Angola cannot be left out of this process. The Angolan Government needs to continuously adjust its policies in order to support private sector-led development initiatives, such as loosening the tax burden on new businesses, or providing tax breaks to commercial banks that offer microfinance programs. Unfortunately, the AEP has often left out the Angolan Government when administering the project. Stakeholders from the three partnering organizations stated that "the AEP is perceived currently to be in the driver's seat," the 2006 Midterm Report flatly states (UNDP Midterm Review 2006, 13). The report adds that the "AEP should support and facilitate; the Government should be in the driver's seat" (*Ibid* 2006, 13). Although the AEP National Steering Committee is composed of representatives from various government ministries as well as private sector and civil society members, it rarely convened and was "basically non-functioning" (*Ibid* 2006, 24). The government has strong networking power with existing private sector banks that it currently underutilizes. The National Bank of Angola has been the most active, but other ministries have done little to mitigate the bureaucratic or resource hurdles of entrepreneurs (*Ibid* 2006, 19). These facts indicate that the AEP needs to focus its energy on enhancing the public sector's role in fostering new enterprises, while targeting the private sector as an engine for entrepreneurial development. Moreover, AEP Program Officer Glayson Ferrari dos Santos also noted that "the AEP's main weakness was its lack of a network with outside actors, such as civil society, the public

sector, and businesses" (interview with Ferrari dos Santos, April 2011). "Floating" entities that are detached from public, or private realms, such as the initial BDS centers, should be incorporated into the existing native environment, rather than depend solely on exogenous support networks.

VI. Recommendations

While final evaluation of the AEP is still not complete, extrapolating from the underlying trend lines indicates the following policies may be feasible. In addition to increasing its staff capacity and improving bureaucratic efficiency (such as making timely payments to employees), the AEP management should reassert the core principles of the BOP approach – namely, that the small-scale entrepreneurs currently confined to the informal sector can flourish and expand if they have access to domestic private-sector credit, basic small-business training, and public-sector legality. These entrepreneurs can thrive if private-sector companies, governments, and non-profit organizations collaborate on solutions. The BOP consumers require new business models, training, and products that existing firms can help create. In turn, Angola's domestic private sector can gain from a massive increase in its client base as well as a more productive labor force. However, if small businesses are to thrive, then the Angolan government and elite private sector, in addition to foreign firms, must become principal actors in the AEP implementation process.

The BOP approach highlights the two-sided coin of harnessing informal sector entrepreneurial energy: public-sector stewardship and private-sector profit motivation. Stakeholders indicated in the 2006 Midterm Review that the AEP needs to streamline its priorities: its main effort should be helping to create an enabling environment within the public sector (goal one) and creating microfinance opportunities (goal three) (UNDP Midterm Review 2006, 14). This indicates that the AEP should focus on the enabling environment within the legal and bureaucratic levels of the government as well as on the ability to elicit microfinance opportunities from the existing banking system. Floating entities that fall into neither category should be revised to fit into an overall public-private linkage system. The BDS Centers reflect an aspect of the AEP that relies on exogenous support networks, rather than endogenous initiatives. The BDS Centers' current funding method, relying heavily on donors, represents the traditional donor-centric approach to development. As such, the primary recommendations for this program are:

The AEP should expand projects that treat informal sector entrepreneurs as a potentially profitable consumer base. Examples of such

projects include the successful Micro Finance Unit created within the National Bank of Angola, the two microfinance institutions established through Banco Sol and Development Works, and the "business incubator" model in Luanda. The AEP should continue to work with major commercial banks to downscale products to BOP markets, emphasizing the commercial profitability of microfinance products (UNDP Angola website (a)). Since 2002, around four to five new banks have started operating in Angola every year (Amaral 2008) on the basis that that retail banking, in addition to large-scale commercial banking, generates substantial profits while also diversifying a bank's balance sheet.

The AEP should continue to engage the Angolan government in the design and implementation of the program. One of the primary barriers to entrepreneurs is the legal system, which continues to neglect the needs of nascent businesses. The government should rationalize laws and the tax system so that entrepreneurs can actualize higher returns in the formal sector than in the informal sector. The AEP should galvanize cross-sector communication by reinitiating the Steering Committee meetings, and holding specific government ministers accountable for their absence or apathy in overseeing the project. If the government is not directly part of the project design, public officials may view UNDP counsel as unwarranted or useless, which would slow the process of legal and institutional change. The AEP also should target mid- and local-level government professionals as well as political appointees, since the public sector must be engaged across all relevant levels. Furthermore, AEP managers should incorporate lessons from India's recent microfinance default crisis and help the Angolan government to install robust regulatory frameworks for the country's growing microfinance sector. Like mainstream finance markets, loans should be made based on the credibility of business proposals and not for short-term consumption.

The AEP should link the BDS Centers with public- and private sector entities. The BDS centers offer valuable educational opportunities for entrepreneurs that should not be sabotaged by poor management. In addition to improving the quality of staff and training programs, these centers can be more efficiently run if they are connected with the Angolan civil-service system or through the university system via internships or knowledge-exchange partnerships. The BDS Centers should also establish more official linkages with the Angolan private sector. For example, the centers could help connect entrepreneurs with commercial-bank loan opportunities. An official mechanism for "graduating" business proposals and trained entrepreneurs from the BDS Centers is necessary to build the

credibility of new businesses in the eyes of potential lenders.

MNCs, such as Chevron, should be used as knowledge resources, not just financial resources. There is enormous potential for Chevron to transfer its business skills to local entrepreneurs, perhaps through skills workshops or training camps. Chevron managers and other private-sector actors can also share their expertise by advising the Ministry of Education on incorporating small-business skills into the core curriculum in public schools. In addition, the AEP should create linkages with Chevron's successful NovoBanco microfinance project and use this as a model for microfinance units within domestic banks. Considering that 75 percent of Chevron's professional employees in Angola are Angolan nationals, the AEP should not view this company as a traditional international donor but as an integral partner in private-sector development (Chevron, "Angola Fact Sheet" 2009). Furthermore, Chevron's contributions should be linked to the UNDP's efforts to improve governance processes within Angola's public sector. Because MNCs are dependent on the host-country government for contracts, they are poorly positioned to suggest institutional reform that would benefit local communities and improve the transparency of government revenues (Wiig et al 2010, 183). A longer-term issue that remains is how to insist that Angolan government authorities sign the Extractive Industries Transparency Initiative, which sets standards for how government revenues from extractive industries are publically reported (*Ibid* 2010, 183). Outside actors such as the UNDP are likely to be the best catalysts for this type of reform.

VII. Expanding the BOP Approach: Global Possibilities

Today, numerous businesses exist that target BOP markets. The Acumen Fund, established in 2001, is an example of a global venture fund that uses entrepreneurial approaches to solve the problems of global poverty. Likewise, the Omidyar Network is a philanthropic investment firm that invests in market-based projects to reduce poverty. Opportunity International and Kiva are more examples of microfinance organizations. IDEO, a business-design consultancy firm, helps governments and NGOs find creative solutions to help poor people. For Chevron, the AEP is currently a country-specific program, but executives have expressed unofficial plans to develop more global partnerships. Dennis Flemming, Chevron project director for the Angola Partnership Initiative in 2004, stated that although the program currently remains country-specific, he is optimistic about similar future alliances. He states, "UNDP works in just about every

country in which we operate so on a global scale this program will help us discover what partnerships are possible in the future" (Barsky 2004, 2). As part of its Private Sector Strategy, the UNDP currently promotes inclusive markets: competitive markets that extend opportunities to the poor, echoing the BOP approach. As of 2007, the UNDP private-sector portfolio consisted of approximately 530 private-sector development (PSD) programs and private-sector engagement (PSE) programs in more than 100 countries, with an annual value of at least $100 million (UNDP Private Sector Strategy Report 2007, 4). Cost-sharing private-sector partnerships, such as the agreement between UNDP-Angola and Chevron, are a recent, but increasingly popular, phenomenon in the global development community. The UNDP has three major global private-sector programs: the Growing Inclusive Markets (GIM) program, the Growing Sustainable Business (GSB) program, and the Public-Private Partnerships for Service Delivery (PPPSD) program (*Ibid* 2007, 4). All of these programs represent new initiatives within the development field, guided by Prahalad's emerging BOP paradigm. The most challenging aspect of this approach is defining the distinct roles of governments, domestic private-sector actors, MNCs, and international development organizations.

The BOP strategy is unique because it is based on what William Easterly would call a "searching" – as opposed to a rationally planned – set of initiatives (Easterly 2006, 3). That is, rather than attempt to define universal, overarching methods for removing poverty, the BOP approach prefers a pilot-project strategy that seeks bottom-up idea generation and implementation.[12] This represents an exciting new advancement in the development field that may help to reduce poverty at unprecedented levels. Based on the experience of the AEP, this article argues that the most successful aspects of BOP partnerships are those that enhance public-private sector linkages at the bottom of the pyramid, as opposed to traditional donor-supported entities that have neither sustained revenue channels nor built-in relationships with Angolan public-sector stewardship. Although many areas of development could require exclusive public-sector authority in certain cases, such as health and education in post-conflict environments, the long-term need for grassroots-driven private-sector growth stems from the BOP approach. The heart of this approach lies precisely in harnessing these profit-driven models of growth at the grassroots level, which should be a general guide for similar projects around the globe.

Notes

[1] For example, in 2008, Official Development Assistance (ODA) totaled $121.5 billion in current U.S. dollars. OECD Stat Extracts, accessed March 10, 2010.

[2] The working definition of "low-income countries" (LICs) for this paper comprises the World Bank's classification of low-income countries (those with less than $995 GNI per capita in 2009 dollars) and lower-middle income (those with $996-$3,945 GNI per capita in 2009 dollars) World Bank 2011.

[3] Prahalad also worked closely with Stuart Hart and Allan Hammond on these ideas.

[4] In Portuguese, the Angola Enterprise Program is *Programa Empresarial Angolano.*

[5] For a full characterization of the "Washington Consensus," see John Williamson, "What Washington Means by Policy Reform," 2002, Peterson Institute for International Economics.

[6] C.K. Prahalad's work inspired U.N. Secretary General Kofi Annan to create the U.N. Special Commission on the Private Sector and Development, of which Prahalad was a key member. This commission published the first major policy document outlining best practices for world leaders wishing to implement BOP strategies. Prahalad, xvi.

[7] The UN Commission on the Private Sector and Development Report states that the "focus on the domestic private sector does not diminish the importance of FDI. Beyond the financial resources that FDI brings, its infusion of a corporate culture can change the way business is done, bring managerial know-how and best practices, provide access to international markets, transfer technology and innovation, introduce competitive pressures in previously closed markets and be the principal driver for the growth of local business. In these situations, FDI can improve the overall investment climate." P. 9.

[8] Angola is second to Nigeria as an oil producer in Sub-Saharan Africa, and the fourth largest diamond producer. MDG Monitor, Angola Profile 2009.

[9] In Portuguese, Angola's Poverty Reduction Strategy is *Estratégia de Combate à Pobreza* (ECP).

[10] The seven ministries represented were the Ministry of Planning (MINPLAN), Ministry of Public Administration, Employment and Social Security (MAPESS), Ministry of Family and Promotion of Women (MINFAMU), Ministry of Commerce (MINCOM), Ministry of Finance (MINFIN), Ministry of Industry, and the Ministry of Fisheries, as well as the Chamber of Commerce and Industry. Government of Angola and UNDP, 14.

[11] In Portuguese, *Banco Africano de Investimentos*.

[12] As stated in the UN Commission on the Private Sector and Development Report, "We concluded at the outset that it would not be enough for this Commission to produce a traditional report voicing opinions and urging others to take action.

Instead, we believe that it is critical to develop a set of pilot actions and initiatives that would test the main observations and conclusions of our work—so that their relevance to the real world of development could be demonstrated on the ground." P. ii.

References

Amaral, Rodrigo. 2008. "Angola's retail banking network, while at an early stage of development, is growing fast along with an oil-driven economy." *The Banker,* January 1.

Barsky, Jennifer. 2004. "Reconstruction in Angola Means Business." *Choices.* 13 no2 Je.

Cabinda Gulf Company Limited: Corporate Responsibility Report 2009. Chevron Corporation. http://www.chevron.com/documents/pdf/AngolaCRenglish.pdf (accessed .March 26, 2011).

Chevron Corporate Responsibility Report 2008. http://www.chevron.com/globalissues/corporateresponsibility/2008/documents/Chevron_CR_Report_2008.pdf (accessed March 10, 2010).

Chevron Corporation. 2009. "Angola Fact Sheet." September. http://careers.chevron.com/global_operations/country_operations/angola/ (accessed 5 March 2010).

CIA World Factbook: Angola. 2009.

Easterly, William. 2002. *The Elusive Quest for Growth: Economists' Adventures and Misadventures in the Tropics.* MIT Press: Cambridge, Massachusetts.

Easterly, William. 2006. *The White Man's Burden: Why the West's Efforts to Aid the Rest have Done so Much Ill and so Little Good.* Penguin Press HC: London.

Ferrari dos Santos, Glayson. AEP Program Officer, UNDP-Angola. Personal interview. April 25, 2011.

Flintoff, Corey. 2010. "India's Poor Reel Under Microfinance Debt Burden." NPR, December 31. http://www.npr.org/2010/12/31/132497267/indias-poor-reel-under-microfinance-debt-burden (accessed February 27, 2011).

Government of Angola and UNDP. "Angola Millennium Goals Report Summary 2005." http://mirror.undp.org/angola/LinkRtf/MDGANG2005-eng.pdf (accessed February 26, 2011).

Government of Angola and United Nations Development Program. 2003. "Project Document: ANG/03/011/ - Angola Enterprise Programme – Support to the Development of the Microenterprise Sector in Angola." http://mirror.undp.org/angola/AEP%20Project.htm (accessed March 2010).

Global Impact Investing Network (GIIN) website. Updated in 2011. Available from: http://www.thegiin.org/cgi-bin/iowa/home/index.html (accessed on February 26, 2011).

Human Development Index Report (UNDP), Angola. 2009. http://hdrstats.undp.org/en/countries/country_fact_sheets/cty_fs_AGO.html (accessed March 2010).

International Finance Corporation (IFC). 2004. "IFC helps private sector boost incomes and job opportunities in Angola. http://www.ifc.org/ifcext/africa.nsf/Content/SelectedPR?OpenDocument&UNID= E8E6319F-48CF43E285256E350077AE54 (accessed March 26, 2011).

Macauhub. January 4, 2008. "Angola: Banco Africano de Investimentos acquires 85.7 pct of Novo Banco." http://www.macauhub.com.mo/en/2008/01/04/4321/ (accessed March 26, 2011).

Malloch, Mark. 2004. "Harnessing the Potential of the Domestic Private Sector for Development." *Choices,* 13 no2 Je.

Millennium Development Goal (MDG) Monitor, Angola Profile. 2009. http://www.mdgmonitor.org/country_progress.cfm?c=AGO&cd=24 (accessed February 2010).

Nelson Mandela Foundation. July 8, 2009. "Professor Muhammad Yunus shares his views on poverty and risk."http://www.nelsonmandela.org/index.php/news/article/professor_muhammad_yunus_shares_views_on_poverty_and_risk/ (accessed March 26, 2011).

OECD Stat Extracts. http://stats.oecd.org/Index.aspx?DatasetCode=TABLE1 (accessed March 10, 2010).

Prahalad, C.K. 2005. *The Fortune at the Bottom of the Pyramid*. NJ: Wharton School Publishing.

Rizvi, Haider. 2004. "Where Small Businesses Can't Grow, Businesses Can't Develop: Renewed emphasis on removing barriers to entrepreneurship and initiative." *Choices*, June.

UN Commission on the Private Sector and Development. 2004. "Unleashing Entrepreneurship: Making Business Work for the Poor." Report to the Secretary-General of the United Nations. New York: UNDP.

UNDP Angola website (a). "Angola Enterprise Program." http://mirror.undp.org/angola/AEP%20Project.htm (accessed 5 Mar 2010).

UNDP Angola website (b). "Poverty Reduction." Last updated January 23, 2011. http://mirror.undp.org/angola/Poverty%20Reduction.htm (accessed 5 March 2010).

UNDP Midterm Review. 2006. "Angola Enterprise Programme Mid-term Review/Evaluation." Luanda: UNDP, March. http://www.ao.undp.org/AEP%20Project.htm (accessed March 2010).

UNDP Private Sector Strategy Report. 2007. "Promoting Inclusive Market Development." September. http://www.undp.org/partners/business/resources/strategy_paper_ps_undp.pdf (accessed March 2010.)

USAID. January 2, 2009. "Global Development Alliance: Small Business Loans Help Angolan Economy." *USAID Frontlines*. http://www.usaid.gov/press/frontlines/fl_mar06/pillars.htm (accessed March 26, 2011).

Versi, Anver. 2006. "Hail the small-scale entrepreneur." *African Business,* October.

Webb, Justin W., Geoffrey M. Kistruck, R. Duane Ireland, and David H. Ketchen, Jr. May, 2010. "The Entrepreneurship Process in Base of the Pyramid Markets: The Case of Multinational Enterprise/Nongovernment Organization Alliances." *Entrepreneurship Theory and Practice,* Vol 34 No 3, pp. 555-581.

Wiig, Arne and Ivar Kolstad. 2010. "Multinational corporations and host country institutions: A case study of CSR activities in Angola." *International Business Review.* 19, pp. 178-190.

World Bank Data and Statistics: Angola. 2009.

World Bank. 2011. "How we Classify Countries." http://data.worldbank.org/about/country-classifications (accessed March 2011).

10

Institutions and the Transplant Effect

Van Trieu Le

Using a sample of forty-seven countries this paper demonstrates that the way in which institutions were historically transplanted is a predictor of contemporary institutional effectiveness. Countries that either developed their institutions internally or meaningfully adapted imported institutions to match local circumstances and context on average outperformed those that did not. The transplantation process influences the strength of both economic and political institutions, though there is stronger evidence found for the link between transplantation and economic institutions. In short, effective institutions are those that were suitably refined by leveraging local knowledge, history, participation and experimentation and those that match local demand. The policy implications are stark: if local support for institutional transplants is negligible, or worse, local resistance is sizable, then untailored insertions of transplants are likely to lead to unintended consequences or nonperformance—aid dollars should prudently be spent elsewhere, where they may yield greater returns.

Van Trieu Le is a graduate student in finance at Princeton University and holds degrees from the University of Cambridge and the London School of Economics and Political Science. He can be reached at vle@princeton.edu.

> "The majority of economists…paint a picture of an ideal economic system, and then, comparing it with what they observe (or think they observe), they prescribe what is necessary to reach this ideal state without much consideration for how this could be done. The analysis is carried out with great ingenuity but it floats in the air."
>
> *—Ronald Coase in The Firm, The Market, and the Law*

I. Introduction

The idea that sound institutional arrangements are necessary for growth and development comes as no surprise to a long line of economic and political thinkers, stretching at least as far back as Montesquieu (1748), including Adam Smith (1776), and uniting a younger generation of scholars from major research centers in North America and Europe. The core argument is relatively constant and simple enough: individuals must make relevant decisions and act within institutional settings and therefore institutions—properly defined—place constraints on behavior and shape human interaction. Good institutions, the argument goes, lead to good outcomes and poor institutions lead to poor outcomes. The institutions hypothesis has found support among a wide body of economists, political scientists, practitioners and lawyers. While the idea that institutions matter is as old as organized government itself, the recent spate of empirical studies purporting to settle the case once and for all in favor of the institutionists stands out as especially impressive. A group of researchers using quantitative data and instrumental strategies measured the effect of institutions on national income per capita and concluded that good institutions do not merely reflect growth but cause it. This body of work has engendered fresh optimism among the new institutionists. Even a critic of these studies observed that this academic cohort "has reached close to an intellectual consensus that the political institutions of limited government cause economic growth" (Glaeser et al. 2004, 272).

But if the march of the new institutionism is ascending, the history of institutional reform in developing countries has been lamentable. Following World War II, a group of U.S. policy advisors and scholars exported American institutions wholesale to countries in Latin America, Africa, and, to a lesser degree, Asia (Berkowitz et al. 1995, 163). By 1974, "the major protagonists of the movement announced its failure" (Berkowitz et al. 1995, 163). The second major wave of institutional transplantation of the last century followed the collapse of the Soviet Union. Once again, legal experts, scholars, economists, policy advisors and practitioners from the United States and Europe flooded "formerly socialist countries with

constitutions, codes, statutes and regulations" (Berkowitz et al. 1995, 164). Nearly twenty years later the results—once again—are pale.

The gap between the growing body of empirical research that highlights the importance of institutions and the poor batting average of institutional transplantation is striking. How do we resolve the apparent tension between the new institutionism and the history of institutional reform and development? This article takes the work of the new institutionists seriously and contends that transplant effects—mismatches between transplant institutions and the underlying characteristics of recipient countries (Berkowitz et al. 1995)—can account for the poor record of institutional reform. The basic argument of the article can be formulated as follows: institutional transplants are more likely to succeed and function effectively on average when (1) institutions are adapted to local conditions, incorporate local knowledge and satisfy local demand and (2) the wielders of political power are responsive to the evolving demand for institutions and can shape and reshape institutional arrangements to meet emerging challenges posed by local problems. Countries that either developed institutions internally or had a population already familiar with the principles of transplant institutions are more likely to build effective political and economic institutions.

Assuming the preceding argument holds in general, the policy implications are clear: reform and technical assistance must not rely too heavily on one-size-fits-all off-the-rack blueprints. In contrast, institutions must be suitably refined by leveraging local knowledge, history, participation and experimentation. More to the point, if local support for institutional transplants is negligible, or worse, local resistance is sizable, then untailored insertions of transplants are likely to lead to unintended consequences or nonperformance. In this case, aid dollars should be prudently spent elsewhere, where they may yield greater returns.

This article largely takes its cue from a pair of papers by Berkowitz, Pistor and Richard (BPR hereafter). In "Economic Development, Legality, and the Transplant Effect" and "The Transplant Effect," the trio argued that legality, or the effectiveness of a legal system, can be more soundly explained by "the way that the law was initially transplanted and received," rather than any reference to legal origins (Berkowitz et al. 1995, 2003). This essay also draws inspiration from Rodrik (1999) and Rodrik et al. (2004). In the first, Rodrik presents a qualitative case for institutional diversity and, citing empirical evidence, argues that democratic institutions are better positioned to deliver predictable growth rates, short-term stability, economic resilience and a more equitable distribution of wealth. In the second, Rodrik et al. demonstrate that institutions are the primary

determinant of growth over competing causal narratives that place trade or geography at the center. In both papers the lesson is clear: high-quality institutions cause high-quality growth.

This paper adds to the discussion by empirically investigating why certain countries are able to develop such institutions while others have failed to do so, to the consternation of well-intentioned American and Western European advisors. Unlike BPR, this paper is not concerned with a comparison of the transplant effect against legal origins to explain legality. The focus is not on legality per se. The central question of this article is not "Can transplant effects do a better job than legal origins of accounting for differences in legality?" but "Can transplant effects account for poor institutional performances?" This leads to the unique regressions contained below, where the most broadly used metrics of economic and political institutions are regressed against the transplant effect, in contrast to the use of a composite variable for legality in BPR. This allows the analysis to (1) measure the impact of the transplant effect on the standard proxies for institutions and (2) to isolate the transplant effect across economic and political institutions and compare the relative perniciousness of the transplantation process in each case. Furthermore, the paper contributes to the literature by accounting for a host of factors that may contribute to institutional failures. The paper demonstrates that the transplant effect continues to have an adverse effect on institutional quality controlling for a litany of the usual suspects that are held responsible for institutional failures: culture, religion, legal origins, national income per capita and inequality. To date, researchers have only controlled for OECD membership, which provides a good first approximation but more refined controls are needed to more rigorously test the strength of the hypothesis against competing explanations.

This article proceeds in five subsequent parts. Part II reviews institutional theory and the empirical results supporting it. Part III applies the idea of transplant effects to institution building exercises. Part IV discusses the data from a sample of forty-seven countries. Part V presents the results from OLS regressions and robustness checks. The results, this paper argues, demonstrate that transplant effects go a long way in predicting variations in institutional effectiveness. At the very least, they render the qualitative arguments plausible. Finally, part VI provides brief policy recommendations and overall conclusions.

II. Institutions

According to Douglass North (1990, 3), "Institutions are the rules of the game in a society or, more formally, are the humanly devised constraints that shape human interaction. In consequence they structure incentives in human exchange, whether political, social, or economic." This definition is deliberately broad enough to incorporate formal mechanisms (i.e. constitutions, laws, charters and treaties) and informal mechanisms (i.e. social norms, conventions and taboos). By this definition, an institution is both "law on the books," as it were, and the application and enforcement of law; it is property rights as well as social attitudes about property rights. Hall and Jones (1999) use the term "social infrastructure" to capture the comprehensive nature of institutions implied by North. Glaeser et al. (2004) emphasize North's formulation of institutions as constraints, placing a particularly heavy burden on constraints on executive power spelled out in formal constitutions. They also add reasonable permanence and durability to the definition. More recently, Acemoglu et al. center their discourse on economic institutions and political institutions, which can be further divided into two broad categories, de jure and de facto political institutions. Acemoglu et al. (2006, 395) write, "[W]e think of good economic institutions as those that provide security of property rights and relatively equal access to economic resources to a broad cross-section of society." De jure political power emanates from political institutions, and de facto political power is derived from the distribution of economic resources and the ability to solve collective action problems, independent of the powers conferred by formal political devices, such as constitutions. This paper follows the lead of Acemoglu et al. and focuses on economic and political institutions.

There are many channels through which good institutions may cause economic growth and development. Rodrik (1999) argues that institutions are necessary to support growth-enhancing market activity by providing property rights protection, appropriate regulation, macroeconomic stability, social insurance and conflict management. Besley and Ghatak (2009) show that improving property rights can improve economic outcomes at the microeconomic level. Hall and Jones (1999, 83) observe that growth theory can explain some variation in cross-country income differences but levels accounting finds a "large residual that varies considerably across countries." They maintain that differences in social infrastructure determine differences in physical and human capital accumulation, which in turn determine differences in national output. Using geographical and language

variables as instruments, they make a case that can perhaps withstand charges of endogeneity.

Acemoglu et al. (2001) use settler mortality as an instrument for institutions to show that the quality of institutions determines income per capita, controlling for geography and colonial history. In a follow-up paper, the authors, using historical data on urbanization patterns and population density, argue that the institutions hypothesis prevails over competing geographical and cultural explanations of cross-country income differences (Acemoglu et al. 2002). This is largely consistent with the economic historical analysis of Engermann and Sokoloff (1999). Dollar and Kraay (2003, 133) regress the log-level of per capita GDP on instrumented measures of trade and institutional quality and conclude that countries "with better institutions and countries that trade more grow faster." However the authors stop short of declaring the relative importance of institutions over trade, citing problems of multicollinearity. Rodrik et al. (2004) take this analysis a step further and with a similar regression conclude that institutions are the strongest determinant of cross-country income differences, over and above both trade and geography. Similarly, the work of Knack and Keefer (1995), Mauro (1995) and Easterly and Levine (2003), among many others, attests to the primacy of institutions in any meaningful diagnosis of cross-country income differences. In the face of such a wide and diverse body of research, one need not subscribe to the institutions hypothesis to realize that policy discussions that revolve around institution building exercises are not wholly misplaced and are in fact supported by both theory and empirics. For the sake of the argument developed hereafter, this paper takes the role that institutions play in determining income per capita as given. If this is so, finding the key drivers of institutional quality becomes not just an academic exercise but may provide a useful guide for development.

III. The Transplant Effect

The stature of the institutional literature in both academic and policy circles provides a sharp contrast from the actual history of institutional transplantation in transitional economies. Technical assistance and policy recommendations from American and Western European advisors oftentimes rest on the assumption that exporting Western style institutions and governance structures would lead to dramatic and welcome reform. This includes a more efficient allocation of resources by allowing market prices to equilibrate supply and demand, greater popular participation in political and economic markets, less abuse and corruption by political elites by

establishing credible checks on power, and greater general welfare defined by the conventional notion of consumer surplus. While some of these benefits have been realized, at least in part, most have not (Braguinsky and Yavlinsky 2000). The glaring shortcomings of such an approach invite a closer inspection of institution building processes. While the literature has often focused on the positive features of ideal institutions—well-defined property rights, effective contract enforcement, constraints on executive power, macroeconomic stability, the elimination of trade barriers, etc.—it has largely ignored the way in which institutions are transplanted.

Countries either developed their institutions internally, received their institutions from foreign powers voluntarily—in the case of Japan adopting German civil law during the Meiji era—or via colonization and conquest. Countries that developed their political and economic institutions internally enjoyed a comparative advantage. First, origin countries could benefit from substantial complementarities among formal institutions and between formal institutions and informal models of social organization, such as norms and conventions. Second, these institutions could incorporate local knowledge, participation and experimentation in a way that foreign institutions could not. Third, political, economic and legal intermediaries gain valuable knowledge and experience developing institutional technologies. In other words, designing institutions is a vital source of human capital that proves useful when it comes to implementing these very institutions in real time. Fourth, and perhaps most importantly, countries will build meaningful institutions to solve their collective action problems when the benefits of doing so outweigh the costs. It stands to reason then that countries that developed their institutions internally had much to gain from such exercises.

In contrast, countries that received their institutions from a foreign country, either voluntarily or through colonialism, had to come to terms with sometimes severe mismatches between foreign institutions and their domestic social, economic and legal contexts. "When a transplant country applies a rule that it received from an origin, it is effectively applying a rule to its own local circumstances that was developed in a foreign socioeconomic order" (Berkowitz et al. 1995, 177). Therefore countries that import their institutions should suffer from the so-called transplant effect—problems arising from mismatches between transplant institutions and the underlying characteristics of recipient countries—that will diminish an institution's effectiveness. However, countries that import their institutions may obviate the transplant effect by (1) having a population that is already familiar with the principles and mechanisms of the imported institution and/or

(2) appropriately modifying imported institutions to meet local challenges (Berkowitz et al. 1995).

Using the classification scheme of BPR (1995, 2003) countries are separated into two broad groups: origin countries, which developed their institutions internally, and transplant countries, which imported their institutions from another country. Transplant countries are further divided into two subgroups: receptive transplants, which either have a population already familiar with the mechanisms of the imported institutional order or have seriously adapted imported institutions to match local circumstances, and unreceptive transplants, which have neither a population familiar with imported institutions or taken steps to adapt such institutions to the local context. On average origin countries should have the most effective institutions followed by receptive transplants then by unreceptive transplants. A set of empirical regressions on a cross section sample of 47 countries supports this hypothesis.

IV. The Data

See Table 1 for a summary of the data and a description of the variables. Two measures of institutional quality are used as dependent variables: average risk against expropriation and constraints on executive power. The first variable is a measure of economic institutions. It is averaged over the period 1982 to 1995 from the scores of April and October from the monthly index. The second variable is a measure of political institutions. It is scaled from one to seven—one indicates unlimited authority while seven indicates strong and effective checks on executive power or subordination to other branches of government, such as the legislature or judiciary (Polity IV 2008).

To test the differences in institutional performances between origin countries, receptive transplants and unreceptive transplants, dummy variables are used for receptive and unreceptive transplants. Whether a country is categorized as an origin, a receptive transplant, or an unreceptive transplant depends on the circumstances surrounding its legal formation, as outlined in BPR (1995). Most countries formalized and consolidated their legal systems in the early-to-mid nineteenth century, but legal development in a few countries took place as late as the early twentieth century. Origin countries are those that developed their institutions internally. This includes Austria, Denmark, Finland, France, Germany, Norway, Sweden, Switzerland, the United Kingdom and, most controversially, the United States. BPR (2003, 5) defend this determination stating, "While English common law influenced the legal system in the United States during the

colonial period, legal development in the United States has sharply diverged from the English System after the colonial period." As a robustness check, BPR run additional regressions with the United States coded as a receptive transplant, which is adopted in this paper.

Transplant countries are those that imported their institutions from abroad. For a country to be coded as a receptive transplant, it must meet at least one of two criteria: (1) the country must have made serious adaptations to the foreign legal order for the purpose of local integration or (2) the country must have a population familiar with the principles of the foreign legal order. If a country does not meet any one of two criteria, it is coded as an unreceptive transplant. Unlike BPR this article is not solely interested in the process of the transplantation of legal systems but in the process of institutional transplantation in general. Are the BPR dummy variables then suitable for the purposes at hand? The author believes so. The scope of inquiry is restricted to political and economic institutions. Legal systems in large part codify political institutions through constitutions and economic institutions through anti-trust, commercial, security, property and contract law. The BPR data are reprinted in the appendix.

To control for various features of culture two variables are included in the regression: (1) ethnolinguistic fractionalization and (2) share of population belonging to the three largest religions—Roman Catholic, Protestantism, and Islam. Ethnolinguistic fractionalization is measured on a unit scale across five different component indices (La Porta et al. 1999).

In a few seminal articles, La Porta, Lopez-de-Silanes, Shleifer and Vishny (LLSV hereafter) assert that countries with French civil law tend to have weaker enforcement of shareholder and creditor rights, more procedural formalism and less judicial independence (La Porta et al. 1999). LLSV argue that as a result, French civil law countries also tend to have weaker enforcement of property rights, among other things. To control for this, a French legal origin dummy is included.

Because high levels of inequality can have an adverse effect on institutions, a variable for inequality is included. The Gini coefficient index was taken from the latest CIA World Factbook. The scores are normalized to the unit interval. A score of zero denotes perfect equality; one denotes perfect inequality. If there were more than one data point per country, the arithmetic average across all available points was used.

V. Methodology and Results

Data and Summary Statistics

A copy of the relevant tables appears in the Annex at the end of this paper.

Table 2 lists the summary statistics for the initial base sample of forty-nine countries. Hong Kong and Taiwan were excluded from the final sample as the complete set of data for both countries was unavailable. Hence the final sample used for OLS regressions consists of forty-seven countries.

From Table 2, one can see the general statistical features of origin, receptive transplant and unreceptive transplant countries. Origin countries generally have the strongest average protection against risk of expropriation followed by receptive transplants then unreceptive transplants. However, receptive transplant countries have the most effective constraints on their executives followed by origins then unreceptive transplants, even though the difference between origins and receptive transplants is statistically indistinguishable. Unreceptive transplant countries in the sample have the highest share of countries with French legal origins, followed by receptive transplants then origin countries. Unreceptive transplants also suffer from the most severe ethnolinguistic fractionalization, almost twice the rate of that of receptive transplants. Origin countries have a larger share of their population belonging to the world's three most populous religions, followed by receptive transplants then unreceptive transplants. Origin countries tend to be wealthier than both receptive and unreceptive transplants. In addition, origin countries, on average, suffer from less inequality, as measured by the Gini index, than transplants. Unreceptive transplant countries suffer from the most inequality with an average Gini score of nearly fifteen percent above that of origin countries.

The results of a series of differences-of-means tests are displayed in Table 3. There are not significant differences between origin countries and receptive transplants. Indeed, across six variables—average protection against expropriation, constraints on executive power, French legal origins, ethnolinguistic fractionalization, religion and the Gini index—origin and receptive transplant countries are statistically indistinguishable. However, the difference in means of log GNP per capita between origins and receptive transplants is significant at the 5 percent level. In contrast, the differences in means between receptive transplants and unreceptive transplants are statistically significant across all variables except one—the Gini index. Likewise, the differences between origins and unreceptive transplants are significant across the board, save for the religion variable.

Identification

The following reduced form equation is estimated using ordinary least squares

(B.1) $$y_i = a'x_i + b'z_i + c + u_i,\ i = 1, 2, \ldots, 47$$

where y_i denotes a measure for institutions, x_i is a 2-vector where the first coordinate is a binary indicator for receptive transplants and the second is an indicator for unreceptive transplants, z_i is an n-vector of control variables, u_i is the disturbance term and a, b and c are a 2-vector, n-vector and a scalar coefficient, respectively.

There are two potential limitations to the empirical work: omitted variables bias and endogeneity. It could be the case that an omitted variable is driving institutions and the transplant effect. To partially insulate the results herein from this weakness, a range of control variables are included in the regressions. The control variables include OECD membership, which serves as a proxy for economic development and relative distance from the technological frontier , and a dummy for French legal origins, which influences a number of substantive and procedural features of a legal system, such as shareholder and creditor protection, depth of financial markets, formalism and judicial independence (La Porta et al. 1999).

By construction, the transplant effect is a function of (1) meaningful adaptation of imported legal orders to local circumstances and (2) familiarity with those imported orders by the local population. As a result, the transplant effect is expected to capture various features of culture. The challenge then is to identify and control for these features of culture through which the transplant effect may influence institutional development. To accomplish this, proxies for ethnolinguistic fractionalization and religion are included in a second series of regressions.

The endogeneity problem is one that plagues most empirical studies, and this paper is no exception. Without question, the transplant effect, as defined in this paper, and institutional strength are endogenous variables. Since the transplant effect is quantified by a pair of binary variables from historical data ranging from the late eighteenth century to the early twentieth century, it is a lagged endogenous variable. Instrumental variables are currently unavailable—the very complexity of the transplantation process makes finding such an instrument extremely difficult if not impossible. Thus the results of the OLS regressions must be interpreted cautiously—no causal inferences can be made which are immune from the statistical problems discussed above.

Ordinary Least Squares Regressions

First this paper examines the role that the transplant effect plays in the performance of economic institutions, as measured by expropriation risk. In the first series of OLS regressions, the institutions variable—average protection against risk of expropriation—is regressed against the receptive

transplant and unreceptive transplant dummies controlling for French legal origins, OECD membership, ethnolinguistic fragmentation, religion and log GNP per capita. The main results are summarized in Table 4.

In every regression the unreceptive transplant coefficient is negative, in most regressions the result is statistically significant and in some cases this significance occurs at the one percent level. Also, the magnitude of the coefficient gives a rough indication of how much damage the transplant effect can inflict on national institutions. For example, the unreceptive transplant coefficient in column (2) is -2.5085. This is roughly the difference between Singapore and South Africa. While column (2) is the most dramatic example, the regressions as a whole show that the transplant effect is sizable. Furthermore, the average adjusted R-squared score of all regressions in Table 4 is an impressive .6959. Thus Table 4 provides strong evidence of the predictive power of the transplant effect on institutional performances: controlling for other factors, countries that suffer from the transplant effect have relatively weaker economic institutions.

In regressions (8), (9) and (10) log GNP per capita is included as a control variable. In all three regressions, log GNP per capita has the expected effect on institutions and this effect is significant at the 1 percent level. In contrast, the transplant effect is no longer statistically significant even though it continues to act as a negative drag on institutional performances. Again, these results are consistent with the framework that is taken for granted throughout the paper, namely that institutions and income move in the same direction, reinforcing the soundness of the starting assumption.

In another set of OLS regressions not printed here , dummies for Asian and African countries are added. The basic results do not change: in every regression transplant effects weaken economic institutions, half of the regressions are statistically significant, and out of those, two are significant at the 1 percent level. Both series of regressions show that the transplant effect is adversely related to the performance of economic institutions, as measured by protection against expropriation risk.

Next the empirical relation between the transplant effect and the performance of political institutions, as measured by constraints on executive power, is studied. Political institutions (executive constraints) is regressed against receptive transplants and unreceptive transplants controlling for French legal origins, OECD membership, ethnolinguistic fragmentation, religion and log GNP per capita. The results are contained in Table 5.

The unreceptive transplant coefficient is negative and sizable in every case. Hence the transplant effect is a reasonably good predictor of the quality of checks on political elites. However, in this series of regressions

there is significant variation in the residual terms. This may indicate that the transplant process relates to political institutions nonlinearly. This is not surprising given the complexity and variety of political organization across countries in the sample. To summarize the results of Table 5, there is evidence of the pernicious effects of the transplant process on political institutions, even though the results are not as pronounced as the relation between the transplant process and economic institutions, as seen in the slightly lower level of statistically significance and the wider residual variation across regressions.

A set of regressions not contained here shows that adding additional dummies for Asian and African countries does nothing to alter the previous results: unreceptive transplants continue to exert a negative influence on political institutions.

As a robustness check, all previous regressions are organized with the United States coded as a receptive transplant, rather than an origin country. The results are largely identical to the regressions reported in Tables 4 and 5, with no significant changes.

VI. Policy Rwcommendations and Conclusion

In four series of OLS cross section regressions, there is evidence that the transplant effect is negatively related to institutional performances. Mismatches between imported institutions and recipient countries correspond to weaker economic and political institutions, weaker protection against expropriation and weaker checks on executive power, even though the effects of transplantation on economic institutions are much more salient. Furthermore, there is no evidence that the transplant effect is a function of either ethnolinguistic fractionalization or religion.

On average, if a transplant country has (1) a population that is familiar with the imported institutional order or (2) meaningfully adapted the imported institutional order to its local context then it is more likely to build effective institutions. The policy implication is clear: development assistance must not rely too heavily on the wholesale exportation of American or Western European institutions for growth and reform. That is not to say that these economic and political institutions—free markets, property rights, contract enforcement, checks and balances on power, macroeconomic stabilization, trade barrier elimination—do not matter. On the contrary, they matter a great deal but in the right context.

How then should policy makers proceed? Three policy recommendations are briefly outlined. First, locals must take the lead. Local lawyers, politicians, economists and activists must forge their own constitutional

order and institutions, with the United States and Western Europe playing a supportive and facilitative role. Second, policy makers should set clear but realistic timelines for transitions. The process of institutional development takes time and patience. Institutions unfortunately cannot be ordered on demand. Third, local institutions should seek to leverage local knowledge deeply and broadly. This will ensure that local concerns are internalized and addressed in institution building. Harnessing the power of the Internet and social media to aggregate local information is especially apt given that many citizens of developing countries are young and possess a high level of technological literacy. Whether one uses technology or more traditional methods of information acquisition, all stakeholders must be given access to institutions to ensure they are robust and responsive.

The discussion thus far leads to a much deeper and unsettling question: if a country can improve the performance of its institutions by becoming more "receptive," then why doesn't it choose to do so over time? One answer is that a country will invest in legal, political and economic research and expertise to adapt institutions to local conditions if the costs of doing so are outweighed by the benefits. But this answer, while partially correct, largely ignores the internal political economic profile of a country. Who wins and loses in an improved institutional framework? Can the winners credibly compensate the losers so that they will not undermine the transplantation process altogether? What incentives do political, legal and economic intermediaries have for developing and maintaining functional institutions? Such questions lie at the heart of the new institutionism. To better understand the transplantation process—and by implication institution building—future research must explore the mapping between political economies and transplant effects. The scope of this article purposely refrains from addressing such a broad question. The modest objective of this article is to explore and establish the empirical link between institutions and the transplant effect, a reminder that institutional exportation exercises devoid of serious consideration of local people, contexts and needs are doomed to failure.

Annex

Table 1: Variable Descriptions

Country	Average protection against expropriation	Constraints on executive	French legal origins	Ethno-linguistic fractionali-zation	Religion	Log GNP per capita (1994)	Gini coefficient
Austria	9.69	7	0	0.0332	0.959	10.0652	0.285
Denmark	9.67	7	0	0.0275	0.960	10.1935	0.2685
Finland	9.67	7	0	0.1050	0.932	9.8679	0.2755
France	9.65	6	1	0.1455	0.818	10.0208	0.327
Germany	9.90	7	0	0.0438	0.814	10.0673	0.285
Norway	9.88	7	0	0.0699	0.982	10.1647	0.254
Sweden	9.40	7	0	0.0650	0.699	10.1161	0.24
Switzerland	9.98	7	0	0.3076	0.963	10.4846	0.334
United Kingdom	9.71	7	0	0.1063	0.306	9.8015	0.354
United States	9.98	7	0	0.2090	0.744	10.1161	0.429
Origins average	**9.753**	**6.900**	**0.100**	**0.1113**	**0.8177**	**10.0898**	**0.3052**
Australia	9.27	7	0	0.1128	0.533	9.7699	0.3285
Belgium	9.63	7	1	0.3638	0.915	9.9828	0.2835
Canada	9.67	7	0	0.3762	0.768	9.9020	0.3180
Ireland	9.67	7	0	0.0904	0.964	9.4727	0.3330
Israel	8.25	7	0	0.3271	0.092	9.5411	0.3735
Italy	9.35	7	1	0.0389	0.837	9.8955	0.2965
Japan	9.67	7	0	0.0099	0.015	10.3574	0.3150
Netherlands	9.98	7	1	0.0634	0.860	9.9499	0.3175
New Zealand	9.69	7	0	0.1476	0.566	9.4415	0.3620
Argentina	5.91	6	1	0.1789	0.993	8.8846	0.4570
Chile	7.50	7	1	0.0506	0.840	8.0619	0.5600
Receptive average	**8.9627**	**6.909**	**0.4545**	**0.1600**	**0.6712**	**9.5690**	**0.3586**
Brazil	7.62	6	1	0.0558	0.919	7.9828	0.5870
Columbia	6.95	6	1	0.0558	0.977	7.2442	0.5500
Ecuador	6.57	4	1	0.3254	0.983	7.0901	0.4920
Egypt	6.30	3	1	0.0231	0.822	6.4922	0.3440
Greece	7.12	7	1	0.0778	0.020	8.9079	0.3420
Hong Kong	8.29		0	0.2368	0.159	9.8015	0.5330
India	7.75	7	0	0.7422	0.140	5.7038	0.3730
Indonesia	7.16	7	1	0.6906	0.509	6.6067	0.3820
Jordan	6.07	3	1	0.0297	0.950	7.0817	0.3805
Kenya	5.98	7	0	0.8270	0.517	5.5984	0.4370
Malaysia	7.95	5	0	0.6104	0.536	8.0520	0.4760
Mexico	7.29	6	1	0.1741	0.959	8.1915	0.5065
Nigeria	5.33	5	0	0.8567	0.729	5.7038	0.4715
Pakistan	5.62	5	0	0.6216	0.981	6.0639	0.3580
Peru	5.54	7	1	0.4316	0.978	7.3065	0.4910
Philippines	5.22	6	1	0.7238	0.922	6.7452	0.4620
Portugal	8.90	7	1	0.0025	0.952	9.1193	0.3705
Singapore	9.30	3	0	0.3215	0.247	9.8960	0.4810
South Africa	6.88	7	0	0.8310	0.517	8.0000	0.6215
South Korea	8.31	6	0	0.0000	0.161	8.9438	0.3355
Spain	9.52	7	1	0.2745	0.970	9.5171	0.3225
Sri Lanka	6.05	5	0	0.3257	0.144	6.3970	0.4750
Taiwan	9.12	7	0	0.2551	0.050	9.2520	
Thailand	7.42	5	0	0.3569	0.045	7.6544	0.4250
Turkey	7.00	7	1	0.1636	0.993	7.9963	0.4180
Uruguay	6.58	7	1	0.0667	0.614	8.2506	0.4500
Venezuela	6.89	4	1	0.0525	0.958	7.9516	0.4885
Zimbabwe	5.61	2	0	0.5986	0.467	6.2538	0.5010
Unreceptive average	**7.0836**	**5.5926**	**0.5357**	**0.3475**	**0.6149**	**7.6359**	**0.4472**

Table 2: Summary Statistics

Country	Average protection against expropriation	Constraints on executive	French legal origins	Ethno-linguistic fractionali-zation	Religion	Log GNP per capita (1994)	Gini coefficient
Austria	9.69	7	0	0.0332	0.959	10.0652	0.285
Denmark	9.67	7	0	0.0275	0.960	10.1935	0.2685
Finland	9.67	7	0	0.1050	0.932	9.8679	0.2755
France	9.65	6	1	0.1455	0.818	10.0208	0.327
Germany	9.90	7	0	0.0438	0.814	10.0673	0.285
Norway	9.88	7	0	0.0699	0.982	10.1647	0.254
Sweden	9.40	7	0	0.0650	0.699	10.1161	0.24
Switzerland	9.98	7	0	0.3076	0.963	10.4846	0.334
United Kingdom	9.71	7	0	0.1063	0.306	9.8015	0.354
United States	9.98	7	0	0.2090	0.744	10.1161	0.429
Origins average	**9.753**	**6.900**	**0.100**	**0.1113**	**0.8177**	**10.0898**	**0.3052**
Australia	9.27	7	0	0.1128	0.533	9.7699	0.3285
Belgium	9.63	7	1	0.3638	0.915	9.9828	0.2835
Canada	9.67	7	0	0.3762	0.768	9.9020	0.3180
Ireland	9.67	7	0	0.0904	0.964	9.4727	0.3330
Israel	8.25	7	0	0.3271	0.092	9.5411	0.3735
Italy	9.35	7	1	0.0389	0.837	9.8955	0.2965
Japan	9.67	7	0	0.0099	0.015	10.3574	0.3150
Netherlands	9.98	7	1	0.0634	0.860	9.9499	0.3175
New Zealand	9.69	7	0	0.1476	0.566	9.4415	0.3620
Argentina	5.91	6	1	0.1789	0.993	8.8846	0.4570
Chile	7.50	7	1	0.0506	0.840	8.0619	0.5600
Receptive average	**8.9627**	**6.909**	**0.4545**	**0.1600**	**0.6712**	**9.5690**	**0.3586**
Brazil	7.62	6	1	0.0558	0.919	7.9828	0.5870
Columbia	6.95	6	1	0.0558	0.977	7.2442	0.5500
Ecuador	6.57	4	1	0.3254	0.983	7.0901	0.4920
Egypt	6.30	3	1	0.0231	0.822	6.4922	0.3440
Greece	7.12	7	1	0.0778	0.020	8.9079	0.3420
Hong Kong	8.29		0	0.2368	0.159	9.8015	0.5330
India	7.75	7	0	0.7422	0.140	5.7038	0.3730
Indonesia	7.16	7	1	0.6906	0.509	6.6067	0.3820
Jordan	6.07	3	1	0.0297	0.950	7.0817	0.3805
Kenya	5.98	7	0	0.8270	0.517	5.5984	0.4370
Malaysia	7.95	5	0	0.6104	0.536	8.0520	0.4760
Mexico	7.29	6	1	0.1741	0.959	8.1915	0.5065
Nigeria	5.33	5	0	0.8567	0.729	5.7038	0.4715
Pakistan	5.62	5	0	0.6216	0.981	6.0639	0.3580
Peru	5.54	7	1	0.4316	0.978	7.3065	0.4910
Philippines	5.22	6	1	0.7238	0.922	6.7452	0.4620
Portugal	8.90	7	1	0.0025	0.952	9.1193	0.3705
Singapore	9.30	3	0	0.3215	0.247	9.8960	0.4810
South Africa	6.88	7	0	0.8310	0.517	8.0000	0.6215
South Korea	8.31	6	0	0.0000	0.161	8.9438	0.3355
Spain	9.52	7	1	0.2745	0.970	9.5171	0.3225
Sri Lanka	6.05	5	0	0.3257	0.144	6.3970	0.4750
Taiwan	9.12	7	0	0.2551	0.050	9.2520	
Thailand	7.42	5	0	0.3569	0.045	7.6544	0.4250
Turkey	7.00	7	1	0.1636	0.993	7.9963	0.4180
Uruguay	6.58	7	1	0.0667	0.614	8.2506	0.4500
Venezuela	6.89	4	1	0.0525	0.958	7.9516	0.4885
Zimbabwe	5.61	2	0	0.5986	0.467	6.2538	0.5010
Unreceptive average	**7.0836**	**5.5926**	**0.5357**	**0.3475**	**0.6149**	**7.6359**	**0.4472**

Table 3: Test of Means (T-Statistics)

	Average protection against expropriation	Constraints on executive	French legal origins	Ethno-linguistic fractionali-zation	Religion	Log GNP per capita	Gini coefficient
origin v. receptive	1.9700	–0.0674	–1.8571	–0.9653	1.1752	2.5147**	–1.7172
receptive v. unreceptive	4.2768*	2.7728*	–0.4457***	–2.0355**	0.4456**	4.7491*	–3.0836
origin v. unreceptive	6.7909*	2.6237**	–2.5305**	–2.4995**	1.6572	5.9777*	–5.1500*

NOTE: Robust standard errors in parentheses.
* Significant at the 1 percent level.
** Significant at the 5 percent level.
*** Significant at the 10 percent level.

Table 4: OLS Regressions: Protection Against Expropriation

INDEPENDENT VARIABLE	DEPENDENT VARIABLE IS AVERAGE PROTECTION AGAINST RISK OF EXPROPRIATION									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Receptive transplant		−0.8596***	−0.5834	−0.4486			−0.1517			−0.0354
		(0.4568)	(0.5054)	(0.3929)			(0.4282)			(0.3352)
Unreceptive transplant	−2.0291*	−2.5085*	−2.0898*	−1.2326*	−0.9197**	−0.6481***	−0.7776	−0.4830	−0.2876	−0.0312
	(0.3395)	(0.4172)	(0.5342)	(0.4509)	(0.3593)	(0.3746)	(0.5263)	(0.3230)	(0.3321)	(0.4347)
French legal origins			−0.5443			−0.6347***	−0.5791		−0.51128***	−0.4768
			(0.4378)			(0.3228)	(0.3619)		(0.2778)	(0.2833)
OECD				1.6854*	1.7863*	1.7268*	1.6978*			0.9727*
				(0.3774)	(0.3682)	(0.3580)	(0.3708)			(0.3206)
Ethnolinguistic fragmentation	−1.5918	−1.5255**	−1.9168**	−0.9255	−0.9222	−1.3760**	−1.3374**	0.3988	−0.0173	−0.0274
	(0.6670)	(0.6499)	(0.7185)	(0.5598)	(0.5617)	(0.5912)	(0.6071)	(0.5437)	(0.5757)	(0.5352)
Religion	−0.8683	−0.9893**	−0.5472	−1.1230*	−1.0713*	−0.5943	−0.6535	−0.2846	0.0741	−0.1845
	(0.4694)	(0.4612)	(0.5801)	(0.3867)	(0.3854)	(0.4453)	(0.4800)	(0.3367)	(0.3815)	(0.3852)
Log GNP / capita								0.8675*	0.8376*	0.6631*
								(0.1238)	(0.1216)	(0.1256)
Intercept	10.2001*	10.7310*	10.4678*	9.0881*	8.7278*	8.6689*	8.7958*	0.9796	1.2099	2.2913***
	(0.4359)	(0.5094)	(0.5487)	(0.5628)	(0.4677)	(0.4541)	(0.5822)	(1.3500)	(1.3207)	(1.3135)
Adjusted R^2	0.5531	0.5770	0.5822	0.7043	0.7022	0.7205	0.7146	0.7841	0.7952	0.8260

NOTE: Robust standard errors in parentheses.
* Significant at the 1 percent level.
** Significant at the 5 percent level.
*** Significant at the 10 percent level.

Table 5: OLS Regressions: Constraints on Executive

INDEPENDENT VARIABLE	DEPENDENT VARIABLE IS CONSTRAINTS ON EXECUTIVE									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Receptive transplant		−0.0464	−0.3784	0.2298				0.1519		−0.0520
		(0.5378)	(0.5945)	(0.5302)				(0.5333)		(0.5841)
Unreceptive transplant	−1.4245*	−1.4503*	−1.9619*	−0.6046	−0.7644	−1.0284***	−0.7169	−0.6106	−0.9657***	−0.8177
	(0.3871)	(0.4925)	(0.6335)	(0.6081)	(0.4790)	(0.5115)	(0.5328)	(0.6553)	(0.5583)	(0.7709)
French legal origins			0.6580			0.6073			0.6111	0.6602
			(0.5179)			(0.4419)			(0.4487)	(0.4955)
OECD				1.1243**	1.0721**	1.1336**				0.8371
				(0.5101)	(0.4909)	(0.4880)				(0.5713)
Ethnolinguistic fragmentation	0.4018	0.4052	0.8882	0.8157	0.8126	1.2595	1.2425	1.2574	1.7338***	1.7195***
	(0.7586)	(0.7684)	(0.8524)	(0.7593)	(0.7520)	(0.8123)	(0.8632)	(0.8741)	(0.9277)	(0.9326)
Religion	−0.2395	−0.2463	−0.7641	−0.3186	−0.3470	−0.7863	−0.0437	−0.0156	−0.4642	−0.6778
	(0.5439)	(0.5558)	(0.6861)	(0.5334)	(0.5243)	(0.6095)	(0.5392)	(0.5540)	(0.6168)	(0.6712)
Log GNP / capita							0.3834	0.3952	0.4143**	0.2430
							(0.2043)	(0.2107)	(0.2034)	(0.2326)
Intercept	7.0273*	7.0561*	7.3597*	5.9452*	6.1318*	6.1699	3.0040	2.7847	2.7814	3.9075
	(0.5028)	(0.6088)	(0.6500)	(0.7708)	(0.6332)	(0.6274)	(2.1989)	(2.3523)	(2.1835)	(2.3811)
Adjusted R^2	0.1983	0.1798	0.1914	0.2473	0.2616	0.2765	0.2418	0.2252	0.2566	0.2610

NOTE: Robust standard errors in parentheses.
* Significant at the 1 percent level.
** Significant at the 5 percent level.
*** Significant at the 10 percent level.

Notes

[i] See, e.g., Acemoglu et al. (2001, 2002), Acemoglu (2006), Rodrik (1999), Rodrik at al. (2004), Easterly and Levine (2003) and Dollar and Kraay (2003).

[ii] For a contrary interpretation see Glaeser et al. (2004).

[iii] See Rodrik (1999).

[iv] For a survey of the legal origins literature by the founders and pioneers of the hypothesis, see La Porta et al. (2008).

[v] For more on the coding of countries see Berkowitz et al. (1995, 2003).

[vi] The component indices include: (1) the probability that two randomly selected people will not belong to the same ethnolinguistic group; (2) the probability of two randomly selected individuals speak different languages; (3) the probability of two randomly selected individuals do not speak the same language; (4) share of the population not speaking the official language; and (5) share of the population not speaking the most widely used language.

[vii] Differences-in-means t-tests are performed between OECD countries and non-OECD countries over the variables average protection against risk of expropriation and log GNP per capita. Both tests are significant at the 1 percent level.

[viii] To test whether the transplant effect is a function of ethnolinguistic fractionalization or religion, this paper regresses unreceptive transplants against ethnolinguistic fractionalization and religion, while controlling for OECD membership. Ethnolinguistic fractionalization moves in the same direction as the transplant effect while religion moves in the opposite direction. The results however are not statistically significant. The same regression is run replacing OECD membership with log GNP per capita as a control. In the second regression, both ethnolinguistic fractionalization and religion move in opposite directions but, once again, fail to be statistically significant. Thus this paper finds no evidence that the transplant effect is in fact driven by ethnolinguistic fractionalization or the religious makeup of a country's population. That is, the effectiveness of institutions are not in fact determined by ethnolinguistic fractionalization or religion acting through the channel of transplantation. Thus it is appropriate to focus on transplantation processes rather than other possible determinants.

[ix] These results are available upon request.

[x] These results are available upon request.

* The author would like to thank Philip Armour, Samir Choksy, Billie Chow and Dimple Pajwani for helpful comments and suggestions. The author is also grateful for the advising of Professor Tim Besley. Needless to say, the responsibility for all errors resides with the author.

References

Acemoglu, Daron, Simon Johnson and James A. Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review* 91(5): 1369–1401.

———. 2002. "Reversal of Fortune: Geography and Development in the Making of the Modern World Income Distribution." *Quarterly Journal of Economics* 117(4): 1231–1294.

———. 2005. "Institutions as a Fundamental Cause of Economic Growth." In *Handbook of Economic Growth*, edited by Philippe Aghion and Steven N. Durlauf, 385–472. Amsterdam: Elsevier B.V.

———. 2001. "A Theory of Political Transitions." *American Economic Review* 91(4): 938–963.

———. 2000. "Political losers as a barrier to economic development." *American Economic Review* 90(2): 126–130.

Berkowitz, Daniel, Katherina Pistor and Jean-Francois Richard. 1995. "The Transplant Effect." *American Journal of Comparative Law* 51: 163–204.

———. 2003. "Economic Development, Legality, and the Transplant Effect." *European Economic Review* 47: 165–95.

Besley, Timothy and Maitreesh Ghatak. 2009. "Property Rights and Economic Development." In *Handbook of Development Economics, Volume 5*, edited by Dani Rodrik and Mark Rosenzweig, 4525–4596. Amsterdam: Elsevier B.V.

Braguinsky, Serguey and Grigory Yavlinsky. 2000. *Incentives and Institutions: The Transition to a Market Economy in Russia*. Princeton, New Jersey: Princeton University Press.

Buchanan, James M., and Gordon Tullock. 1962. *The Calculus of Consent: Logical Foundations of Constitutional Democracy*. Ann Arbor: University of Michigan Press.

Dollar, David, and Aart Kraay. 2003. "Institutions, Trade and Growth." *Journal of Monetary Economics* 50(1): 133–162.

Easterly, William, and Ross Levine. 2003. "Tropics, Germs, and Crops: How Endowments Influence Economic Development." *Journal of Monetary Economics* 50(1): 3–39.

Economy, Elizabeth C. 2011. *Roots of Protest and the Party Response*. Prepared statement before the U.S.-China Economic and Security Review Commission, U.S. Senate/U.S. House of Representatives, First Session, 112th Congress.

Engerman, Stanley L., and Kenneth L. Sokoloff. 1991. "Factor Endowments, Institutions, and Differential Paths of Growth among New World Economies." in *How Latin America Fell Behind*, edited by Stephen Haber, 260–306. Stanford, CA: Stanford University Press.

———. 1991b. "Institutions, Factor Endowments, and Paths of Development in the New World." *The Journal of Economic Perspectives* 14(3): 217–232.

Glaeser, Edward, Rafael La Porta, Florencio Lopez-de-Silanes and Andrei Shleifer. 2004. "Do Institutions Cause Growth." *Journal of Economic Growth* 9(3): 271–303.

Haber, Stephen, ed. 2002. *Crony Capitalism and Economic Growth in Latin America: Theory and Evidence*. Stanford, CA: Hoover Institution Press, Stanford University.

Hall, Robert E., and Charles I. Jones. 1999. "Why Do Some Countries Produce

So Much More Output per Worker Than Others?" *Quarterly Journal of Economics* 114(1): 83–116.

International Country Risk Guide. 1996. Political Risk Services, East Syracuse, NY.

Knack, S., and P. Keefer. 1995. "Institutions and Economic Performance: Cross Country Tests Using Alternative Measures." *Economics and Politics* 7(3): 207–227.

Kurtlantzick, Joshua. 2011. Lessons for the Mideast from Asia's Revolutions. http://www.cfr.org/middle-east/lessons-mideast-asias-revolutions/p24246 (accessed February 25, 2011)

La Porta, Rafael, Florencio Lopez-de-Silanes and Andrei Shleifer. 2008. "The Economic Consequences of Legal Origins." *Journal of Economic Literature* 46(2): 285–332.

La Porta, Rafael, Florencio Lopez-de-Silanes, A. Shleifer and Robert Vishny. 1997. "Legal Determinants of External Finance." *Journal of Finance* 52(3): 1131–1150.

———. 1998. "Law and Finance." *Journal of Political Economy* 106(6): 1113–1155.

———. 1999. "The Quality of Government." J*ournal of Law, Economics, and Organization* 15(1): 222–279.

Mauro, Paolo. 1995. "Corruption and Growth." *Quarterly Journal of Economics* 110: 681–712.

North, Douglass C. 1981. *Structure and Change in Economic History*. New York: Norton and Co.

———. 1990. *Institutions, Institutional Change, and Economic Performance*. Cambridge: Cambridge University Press.

North, Douglass C., and Robert P. Thomas. 1973. *The Rise of the Western World: A New Economic History*. Cambridge: Cambridge University Press.

Rodrik, Dani. 1999. "Institutions for High-Quality Growth: What They Are and How To Acquire Them." Draft prepared for the IMF conference on Second Generation Reforms, Washington, DC, 8–9 November 1999.

Rodrik, Dani, Arvind Subramanian and Francesco Trebbi. 2004. "Institutions Rule: The Primacy Institutions over Geography and Integration in Economic Development." *Journal of Economic Growth* 9: 131–65.

Trubek, David M., and Marc Galanter. 1974. "Scholars in Self-Estrangement: Some Reflections on the Crisis in Law and Development Studies in the United States." *Wis. L. Rev*. 1062.

The World Factbook. 2010. Central Intelligence Agency, Washington D.C.