Spring 2012

Baker Center Journal of Applied Public Policy - Vol. IV, No. I

Thomas Graham

Robert F. Durant

Robert M. Stein

John M. Scheb II

Colin Glennon

See next page for additional authors

Authors

Thomas Graham, Robert F. Durant, Robert M. Stein, John M. Scheb II, Colin Glennon, Hemant Kumar Sharma, Dorian Stiefel, Kristin Wagers, Mark Harmon, Cristina Reiser, and Benjamin Signer

BAKER CENTER JOURNAL OF APPLIED PUBLIC POLICY

Published by Howard H. Baker Jr. Center for Public Policy at the University of Tennessee, Knoxville

COMMENTARY

ARTICLES

STUDENT ARTICLES

BAKER CENTER JOURNAL OF APPLIED PUBLIC POLICY

VOL. IV NO. 1 | SPRING 2012

The Baker Center Journal of Applied Public Policy is published by the Howard H. Baker Jr. Center for Public Policy at the University of Tennessee, Knoxville.

The vision of the *Baker Center Journal of Applied Public Policy* is to provide public officials, policymakers, political activists, scholars, and citizens with forward-looking commentary and research on matters of public policy. The journal will focus on pragmatic, rather than theoretical, analysis of issues that are regional, national, or international in scope.

Thank you for your interest, and please direct any questions to Dr. Nissa Dahlin-Brown, Associate Director, Howard H. Baker Jr. Center for Public Policy, (865) 974-0931. www.bakercenter.utk.edu

Baker Center Journal of Applied Public Policy
Published by THE UNIVERSITY OF TENNESSEE
Howard H. Baker Jr. Center for Public Policy
VOL. IV NO. I Spring 2012

THE UNIVERSITY of TENNESSEE KNOXVILLE

University Administration

Joe DiPietro, DVM
President

Jimmy G. Cheek, PhD
Chancellor

Susan D. Martin, PhD
Provost

Howard H. Baker Jr. Center for Public Policy

Matthew M. Murray, PhD
Director

Carl A. Pierce, JD
Director Emeritus

Nissa Dahlin-Brown, EdD
Associate Director

Patti Rebholz
Business Manager

Michelle Castro
Development Director

Commentary

The Nuclear Non-Proliferation Treaty Under Threat: *Iran And North Korea*[1]

Ambassador Thomas Graham Jr.[2]

The world changed in 1945 with the advent of the atomic bomb. For the first time, humankind possessed a weapon with which it could destroy itself. However, disarmament efforts gradually gained momentum and over time a web of international treaties and agreements was constructed which limited weapon development and inhibited the spread of nuclear weapons as well as chemical and biological weapons. These efforts changed the course of history and made the world a safer place.

Nuclear weapons are truly unlike any other form of weaponry. The atomic bomb used against Hiroshima in 1945 was 14 kilotons or 14 thousand tons of TNT explosive equivalent. In a few years, the United States and the Soviet Union were testing nuclear weapons in the megaton range or million tons of TNT explosive equivalent. Soon, a vast nuclear arms race was underway. The Soviet Union built 55,000 nuclear weapons, the United States some 72,000 and at one time had 32,500 in its arsenal, the Soviet Union possessed around 45,000. This effort eventually bankrupted the Soviet Union and cost the United States in excess of $5.5 trillion in 2004 dollars.[3]

President John F. Kennedy truly believed that there was a serious risk that nuclear weapons were destined to sweep all over the world. In March of 1963 in response to a reporter's question at a news conference, he said, "[P]ersonally, I am haunted by the feeling that by 1970 . . . there may be 10 nuclear powers instead of 4, and by 1975, 15 or 20. . . . I would regard that as the greatest possible danger and hazard."[4] He spent much of his presidency pursuing the cause of nonproliferation. President Kennedy had been told by the outgoing Secretary of State, Christian Herter, in December of 1960 that nuclear weapons would spread to additional countries and that the most likely next

1 This article is drawn in significant part from a forthcoming book by the author.

2 Ambassador Graham is a former Special Representative of the President for Arms Control, Non-Proliferation, and Disarmament (1994-1997), in which capacity and others, he participated in the major arms-control/non-proliferation negotiations in which the United States engaged between 1970 and 1997. He served as General Counsel to the United States Arms Control and Disarmament Agency (ACDA) (1977-1981, 1983-1993), as Acting Director of ACDA (1993), and as Acting Deputy Director of ACDA (1993-1994).

3 Robert S. Norris, and William M. Arkin, "NRDC Nuclear Notebook: Global Nuclear Stockpiles, 1945-2000," *Bulletin of the Atomic Scientists* 56 (March-April 2000): 79.

4 Robert Dallek, *An Unfinished Life: John F. Kennedy, 1917-1963* (Boston: Little, Brown & Co.2003), p. 615.

nuclear weapon states were India and Israel. He took this very seriously.[5]

If such anticipated proliferation had in fact happened, there could be far more than two dozen nuclear weapon states in the world today, with nuclear weapons integrated into their national arsenals. Mohamed ElBaradei, the former Director General of the International Atomic Energy Agency (IAEA), and more recently a leader of the opposition during the revolution in Egypt expressed this concern in 2004 when in a speech in Washington D.C., he said, "The danger is so imminent . . . not only with regard to countries acquiring nuclear weapons, but also terrorists getting their hands on some of these nuclear materials – uranium or plutonium."[6] Director General ElBaradei said in another speech around the same time that more than 40 countries then had the capability to build nuclear weapons. Under such circumstances with that many nuclear weapon states in existence, potentially every significant conflict could have brought with it the risk of going nuclear. It might have also become extremely difficult to keep nuclear weapons out of the hands of terrorist organizations.

In 1965, the United Nations (UN) General Assembly took up the subject. A resolution was passed, which over the next few years, proved to be the blueprint of the Nuclear Non-proliferation Treaty (NPT). Among other things, this resolution called for "balanced obligations" between nuclear weapon and non-nuclear weapon states in the treaty to be negotiated. The NPT was signed in 1968 and entered into force in 1970 and came to be recognized as the principal reason, along with the parallel extended deterrence policies of the United States and the Soviet Union, that President Kennedy's darkest fears have thus far not been realized.

The success of the NPT was no accident. The treaty was based on a carefully crafted central bargain, which incorporated the "balanced obligations" concept. In exchange for a commitment from the non-nuclear weapon states (today more than 180 nations, most of the world) not to acquire nuclear weapons and to submit to international safeguards to verify compliance with this commitment, the NPT nuclear weapon states (the United States, the United Kingdom, France, Russia, and China) pledged unfettered access to peaceful nuclear technologies and undertook to engage in nuclear disarmament negotiations aimed at the ultimate elimination of their nuclear arsenals. It is this basic bargain that for the last four decades has formed the central underpinnings of the international non-proliferation regime.

However, one of the principal problems with this arrangement has been that the NPT nuclear weapon states have never fully delivered on the disarmament part of this bargain. The essence of the disarmament commitment in 1968 and thereafter was that, pending the eventual elimination of nuclear weapon arsenals, the nuclear weapon states would: agree to a treaty prohibiting all nuclear weapon tests, that is, a comprehensive nuclear test ban; negotiate an agreement prohibiting the further production of nuclear bomb explosive material; undertake obligations to drastically reduce their nuclear arsenals; and give legally binding commitments that they would never use nuclear weapons against NPT non-nuclear weapon states.

5 Jeffrey T. Richelson, *Spying on the Bomb: American Nuclear Intelligence from Nazi Germany to Iran and North Korea* (New York: W.W. Norton, 2006), p. 254. Reported in an editorial, "A Warning About WMD," in the *Washington Times*, June 25, 2004.

6 Reported in an editorial, "A Warning About WMD," in the *Washington Times*, June 25, 2004.

However, few of these disarmament elements of the NPT basic bargain have been actually accomplished forty years later.

The NPT is essentially an international strategic political undertaking, which should be observed. It is not a gift from the non-nuclear weapon states. Few deny that today the NPT is in trouble. The question is how long it can remain viable as an unbalanced treaty with an important part of its basic bargain unrealized and a significant part unraveling as North Korea and Iran pursue the bomb. Recognizing this vulnerability of the NPT, and with the end of the Cold War accompanied by the potential spread of nuclear weapon technology to failed and failing states and international terrorist organizations, serious efforts have begun to attempt to move toward the complete elimination of nuclear weapons, as called for in the NPT.

Since the mid-twentieth century almost all American presidents have placed arms control and non-proliferation policy high on their agendas. President Dwight D. Eisenhower considered his failure to have achieved a nuclear test ban to be the greatest disappointment of his presidency. The NPT was signed on President Lyndon Johnson's watch. President Richard Nixon oversaw the negotiation of the SALT I Agreements and the beginning of the SALT II Treaty process. The SALT II process was continued under President Gerald Ford and concluded under President Jimmy Carter. President Carter also attempted to negotiate a comprehensive nuclear test ban, which was finally concluded under President Bill Clinton's leadership. President Ronald Reagan advocated the abolition of all nuclear weapons and completed the medium range nuclear missile treaty. The most successful arms-control President was President George H.W. Bush. His Administration concluded four major arms-control treaties during his four years as President: the START I Treaty, the START II Treaty, the Conventional Armed Forces in Europe Treaty, and the Chemical Weapons Convention. No other President has successfully completed more than one. Thus, nuclear arms control, non-proliferation, and disarmament negotiations have been at the center of U.S. foreign policy for much of the last fifty years.

No other President has spoken out more eloquently and in such a comprehensive manner, than President Barack Obama, who in Prague in April of 2009 declared his strong support for a replacement START Treaty to be followed by deeper cuts in nuclear weapons leading to a multilateral nuclear weapon reduction negotiation involving all of the nuclear weapon states. He reiterated his support for U.S. ratification and entry into force of the Comprehensive Nuclear Test Ban Treaty (CTBT), and he confirmed his own support for a process that would lead to a nuclear weapon free world. He underscored his commitment to the strengthening of the NPT, along with measures to enact stronger safeguards of fissile material around the world. In addition, he urged the prompt negotiation of a treaty prohibiting the further production of fissile material. The following September, with President Obama in the chair, the United Nations Security Council endorsed the goal of the elimination of nuclear weapons.

In addition, the long-awaited Nuclear Posture Review (NPR) for this Administration was released in the spring of 2010. Among many other things, the NPR brings U.S. national policy into line with the U.S. 1995 NPT political commit-

ment, effectively never to use nuclear weapons against non-nuclear weapon NPT parties in good standing. It also reduces the role of nuclear weapons in U.S. security policy, another NPT commitment, made in 2000. The NPT is the central international agreement underlying international peace and security in today's world. The principal quid for the quo of most nations of the world to never acquire nuclear weapons under this Treaty was the test ban. It is the only arms-control agreement explicitly mentioned in the NPT, and it is the most significant commitment made by the nuclear weapon states to bring the necessary political balance to the NPT, the idea of "balanced obligations" mentioned above. The 1995 Statement of Principles, which was the political price for making the NPT a permanent Treaty, referred to an indefinite extension of the NPT in 1995 and explicitly called for the negotiation of a CTBT in one year, that is, by the end of 1996.

This 1996 deadline from the NPT was met, and the CTBT was signed in September 1996 with the United States as the first signatory, but the U.S. Senate rejected the CTBT in 1999. Over 150 states have now ratified the CTBT, including 36 of of the 44 states whose ratification is required by the Treaty for entry into force. Most of the rest of this required group are waiting upon ratification by the United States, China, Israel, and Indonesia more or less explicitly.

While President Obama could have immediately pushed for ratification of the CTBT in 2009—at the time when history tells us that a new President's political strength is at its zenith— his Administration, perhaps correctly, chose not to do so. Passing the CTBT in the U.S. Senate has never been primarily a matter of the merits of the Treaty; it largely has been about politics and views about nuclear weapons policy. Most Senators are unacquainted with the Treaty details and important specific issues such as verification and nuclear stockpile maintenance. Rather, they follow the lead of a few Senators who are knowledgeable and who have specific nuclear issue agendas which they wish to pursue. This was true in 1999, was true in February, 2009, and is true today. And, without U.S. ratification, the CTBT will never come into force.

A further complication has been the complexity of the process of achieving START Treaty ratification. The negotiations were completed in late March, 2010, after a year of vigorous effort. A deployed total warhead level of 1,550 operational strategic warheads and a limit of 800 strategic nuclear-weapon delivery vehicles were agreed. These are significant contributions. Their achievement have opened the door to negotiations toward further, much steeper U.S.-Russian reductions in strategic nuclear weapons, perhaps to the level of 1,000 total nuclear weapons each. This level is regarded as a necessary step to permit consideration of multilateral nuclear weapon reduction negotiations involving all nuclear weapon states which, over a long period of time, if successful, would put the world community on the road toward the eventual elimination of nuclear weapons. Of course, as long as the NPT holds.

However, the NPT remains in crisis, in part because of the essential failure of the disarmament agenda, and in part because of the continuing high political value of nuclear weapons, whereby the possession of nuclear weapons is seen as the distinguishing feature separating great powers from other states. With the NPT basic bargain remaining unrealized after forty years, states eventually could leave

the Treaty; North Korea already has de jure, Iran has de facto. Without the NPT, disarmament is nothing. The NPT is threatened from many sides today but primarily at this time by the nuclear programs in Iran and North Korea. The nuclear programs in these two states threaten to break open the NPT regime in the Middle East and Northeast Asia and thereby unleash the wide-scale nuclear proliferation that President Kennedy so greatly and rightly feared.

When considering the Iran issue, one should remember that Iran has been pursuing nuclear weapons for many decades. With the assistance of Germany, France, and South Africa, Iran began making plans for a large nuclear infrastructure under the former Shah in the 1970s. A contract was signed with Siemens of Germany to construct two 1,000 megawatt nuclear power reactors at Bushehr, near the border with Iraq, in the late 1970s. Ultimately, Iran hoped to build twenty power reactors or perhaps even more, making it a major player in the world nuclear industry.

In 2004, long-held suspicions that the quest for nuclear power production was linked to a nuclear weapons program were largely confirmed by one of the Shah's former foreign ministers, Ardeshir Zahedi, who said, "the assumption within the policymaking elite was that Iran should be in a position to develop and test a nuclear device within 18 months" of making the decision to construct nuclear weapons.[7]

Why would Iran want nuclear weapons? The motivations of the Islamic Republic today are, of course, different than those under the western-aligned Shah, but the radically different regimes share one common and compelling reason for pursuing the bomb: international prestige. Since early in the Cold War, the possession of nuclear weapons has distinguished great powers from other states. The political value of nuclear weapons has not declined since the end of the Cold War, despite urgings that nuclear weapons should play a lesser role in the security policies of states. The United Nations Permanent Five (the United States, the United Kingdom, France, Russia, and China) by no accident are the nuclear weapon states authorized by the NPT. In 1958, British Prime Minister Harold MacMillan said that the British program "puts us where we ought to be . . . in the position of a great power." President Charles de Gaulle of France said in 1961 that "a great state," which does not possess nuclear weapons when others do, "does not command its own destiny." Prime Minister Vajpayee of India indicated in 1998 after the Indian tests that India now was a truly important country since "We have a big bomb now."[8] When the Permanent Five met in Paris to fashion a response to the Indian and Pakistani tests, there were reports that Germany and Japan sought to come as well but were told no, at some level, since they were not nuclear weapon states.

Iran is a proud country. The Persian cultural heritage is one of the richest in world civilization. The Persian Empire was once the world's most powerful in history. If former provinces and client states of that empire now have nuclear weapons, why should modern Persia not possess them as well? The view of many Iranians likely is that Iran deserves to be a great power. Dr. Akbar Etemad, the director of

7 Ray Takeyh, *Hidden Iran: Paradox and Power in the Islamic Republic* (New York: Henry Holt & Co., 2007), p. 136.

8 Thomas Graham, Jr., "Avoiding the Tipping Point," review of *The Nuclear Tipping Point: Why States Reconsider Their Nuclear Choices*, ed. by Kurt M. Campbell et al., *Arms Control Today* 34 (Nov. 1, 2004): 45.

Iran's nuclear program at the time of the monarchy, asserted that the Shah's program was designed to provide the nuclear option should any of Iran's regional competitors move in that direction.[9] But there is more to it than that. In the 1970s, none of Iran's potential competitors were moving toward nuclear weapons. However, if there was one thing that the Shah was interested in pursuing, it was international prestige for Iran and derivatively for himself.

Many Iranians believe that, as a great civilization with a long history, Iran has a right to acquire nuclear weapons.[10] The nuclear program and Iran's national identity have become linked in the minds of Iran's rulers. For instance, Ali Hussein-Tash, Deputy Secretary of the Supreme National Security Council, said in 2006, "A nation that does not engage in risks and difficult challenges, and a nation which does not stand up for itself, can never be a proud nation."[11] A reformist activist, Sayyed Mostafa Tajzadeh, noted in 2003, "It's basically a matter of equilibrium; if I [do not] have a nuclear bomb, I don't have security."[12]

For better or for worse, Iran suddenly was able to see itself as the region's major power with the demise of Saddam Hussein. The acquisition of nuclear weapons by Iran could enhance its role as a regional power and increase Iranian influence in the Middle East. Of course, the Islamic Republic has other motivations that the Shah's regime did not. Iran under the Shah was allied with and protected by the United States; this is no longer true. The Islamic Republic was deeply scarred by the damage inflicted on it during the 1980-1988 Iran-Iraq war, in which more than a million people died; that experience left a strong "never again" sentiment among both Iran's rulers and its population. Nuclear weapons for the Islamic Republic are seen as insurance that something like the war with Iraq will never happen to Iran again. It is also seen as a measure of fair treatment. Iranians note that, under the Shah, the Western powers supported the nuclear power program; they also note the West's silence about Israel's nuclear-weapons program. Why should Iran be any different?

After the Islamic Revolution in 1979, Iran initially did not pursue nuclear weapons. Ayatollah Khomeini, as well as others in the regime, was of the view that the widespread, indiscriminate nature of nuclear weapons was contrary to Islam's rules on warfare. Toward the end of the devastating war with Iraq, Ayatollah Khomeini apparently changed his mind.

The nuclear program began again in earnest under President Rafsanjani, who took office in 1989, and has been continued by his successors, both the reformist Mohammad Khatami and the conservative, Mahmoud Ahmadinejad. The Germans were unwilling to return and rebuild the two Bushehr reactors on which Germany had begun construction work prior to the Iranian Revolution and which were badly damaged during the war with Iran. Iran then turned to Russia, which signed a $800 million contract with Iran to complete the construction of one of the two reactors. Thus began a close Russian-Iranian commercial nuclear relationship

9 Ibid.

10 Ibid., p. 154.

11 Ibid., p. 149.

12 Ibid., p. 155.

that has continued to the present day. As a practical matter, the damage was such that Russia had to begin construction anew at the Bushehr site. Despite constant pressure from the United States, Russia persevered in this effort and announced completion of construction of the reactor in 2004. Nevertheless, Russia delayed delivering fuel for the reactor for some years and argued that its reliable supply obviated the need for Iran to engage in domestic uranium enrichment. The first fuel load was finally delivered in late 2009.

The United States has been concerned about the Iranian nuclear program for many years. During the 1980s, the Reagan Administration persuaded European governments to adopt strict controls on exports to Iran of items that might be used for a nuclear weapons program. In 1992, Robert Gates, then director of the Central Intelligence Agency (CIA), testified before Congress that Iran was seeking nuclear weapons and might acquire them by the year 2000. The United States has tried repeatedly over the years to limit the involvement of Russia and other nations in the Iranian program, with only partial success. Many agreements were made, the most notable being the December 1995 accord negotiated by Vice President Al Gore and Prime Minister Viktor Chernomyrdin. Russia agreed to limit its cooperation with Iran to only one reactor at Bushehr. In 1992, the U.S. was able to block the sale of nuclear items to Iran. Argentina was persuaded not to ship fuel fabrication equipment to Iran, which was already packed and ready to go. Later, China agreed not to sell to Iran a reactor it was seeking. In 1995, China had plans to sell two 300-megawatt pressurized water reactors to Iran. U.S. Secretary of State Warren Christopher spoke with Chinese Foreign Minister Qian Qichen and China eventually backed away from the sale. "A regime with this kind of record simply cannot be permitted to get its hands on nuclear weapons," Christopher told reporters.[13]

Perhaps the most revealing incident of all took place the previous year, when U.S. and other Western intelligence learned that Iranian officials had visited the Ublinsky Metallurgical Works in the former Soviet Republic of Kazakhstan. The Iranian visitors had expressed interest in highly enriched uranium (HEU), which was stored at the facility apparently in a "highly insecure way." In order to ensure the HEU would not fall into Iranian hands, the United States ultimately purchased all 1,320 pounds of the HEU stored at the plant and removed it in three transport aircrafts to Oak Ridge National Laboratory in an operation known as "Sapphire."[14] There could be no other purpose for Iran to attempt to acquire such material except for the construction of nuclear weapons.

Nevertheless, the Iranian program appeared to slow somewhat during the latter half of the 1990s. It appeared that diplomacy worked, complemented by faltering steps within Iran; however, American officials were unaware of the ongoing cooperation between Iran and the A.Q. Khan network. The U.S. intelligence community had become aware of Khan's efforts to trade uranium enrichment technology to North Korea, but the CIA had missed the Iranian arrangement with Khan. Over many years, Iran had been acquiring uranium enrichment technology from Khan for deployment at its facilities being developed at Natanz. Most significantly,

13 Richelson, *Spying on the Bomb*,, pp. 508-09.

14 Ibid., p. 507.

this involved the transfer of plans and models of centrifuge systems. In addition, it later became known that Khan had acquired from China the blueprints for a nuclear weapon tested by China in 1966 at around the 20-kiloton level. The plans for this weapon were later found to have been transferred to Libya. It is reasonably likely that they were transferred to Iran as well.

All this changed in 2000, when the CIA was able to penetrate the Khan network. One of the key figures in the Khan ring agreed to cooperate with the CIA. He was the network's senior representative at the transshipment post of Dubai in the United Arab Emirates. In this way, the CIA learned that centrifuges and other equipment were being sent by Khan to Iran and Libya. The CIA not only learned about the illegal transfers, but it also was given access to the equipment as it passed through Dubai. This access permitted the CIA secretly to cripple this equipment, including the technology going to Iran. The sabotage was effective and verified by the IAEA as inspectors who travelled to Iran. Inspectors noticed vacuum pumps that had been cleverly damaged so they would not work and a power supply that had been shipped to Iran was found to be defective.

Then at a briefing at the National Press Club in Washington in 2002, the National Council of Resistance of Iran, the political arm of the People's Mujahedin, known as the MEK, revealed the existence of a heavy-water reactor program at Arak and enrichment facilities at Natanz. Although this organization is often referred to in American media as a "dissident organization," in reality, the MEK had been a terrorist organization operating in Iran and was listed as such by the United States for many years. The organization was expelled from Iran in the 1980s to Iraq, where it became a wholly owned subsidiary of Saddam Hussein. Thus the source was suspect, but the facts apparently were real.

The Natanz facility, construction of which began in 2000, included a pilot enrichment plant, which could house some 1,000 centrifuges and a large underground facility eventually intended for perhaps fifty thousand centrifuges. At Natanz, Iranians would eventually be able to enrich on an industrial scale and produce enough highly enriched uranium for 20 weapons a year or more. The heavy-water reactor at Arak, once operational, might produce plutonium sufficient in amount for one to two weapons per year. It is approximately the same size as the North Korean reactor at Yongbyon, which has fueled that country's nuclear weapons program. Overall, the Atomic Energy Organization established by the Shah did not provide administrative direction for the nuclear program in Iran. Rather, the program was, and is, led by the elite Revolutionary Guards, removing any doubt that it is intended as a military program the ultimate objective of which is nuclear weapons.

Iran's nuclear program includes a number of facilities that have been developed over the years. In addition to the facilities at Arak and Natanz, there is a Nuclear Technology Center located at Isphahan that includes mini reactors, subcritical assemblies, and a fuel fabrication laboratory. It may be Iran's equivalent of Los Alamos. Also located at Isphahan is the Uranium Conversion Facility, which converts uranium "yellowcake" to uranium hexafluoride gas for separation in the centrifuges at Natanz. In Tehran is the Tehran Nuclear Research Center with an operating research reactor, a radio isotope production facility, various laboratories, and a waste handling facility.

The University of Tehran houses a five-megawatt research reactor, provided by the United States to the Shah's regime, which makes medical isotopes under IAEA safeguards. Argentina supplied fuel (enriched to nearly twenty percent as opposed to two to four percent for commercial power reactors) for this reactor, but it is running low. There are also various front companies that buy technology and material abroad for the nuclear program, such as Kala Electric in Tehran.[15] In 2010, the world learned of the secret enrichment facility under construction at a Revolutionary Guard base near the holy city of Qom. The underground facility may have been intended to house potentially around 3,000 centrifuges, enough for a weapons program, but not for a power program. Since the discovery of the site, the Iranian Government may have lost interest in this facility.

In the days and weeks after September 11, 2001, Iran took steps to distance itself from al Qaeda and the attacks in New York and Washington. The mayor of Tehran sent his condolences. When the U.S. was preparing its assault on Afghanistan after the Taliban government refused to give up Osama bin Laden, Iran indicated that it would return American warplane pilots downed in the forthcoming fighting. Iran also pressed its allies in the Northern Alliance to work with the Americans and permitted the United States to ship food trucks across its territory to Afghanistan. At the January 2002 conference of donor nations in Tokyo, Iran pledged $350 million for rebuilding Afghanistan. After the Taliban government was overturned, Iran played a helpful role in the creation of the Karzai government in the Bonn negotiations in December 2001.

In addition, Iranian leaders confirmed their willingness to cooperate with the United States to further Iran's economic interests. President Muhammad Khatami said: "The government cannot come up with the money needed to create a million jobs a year. We need private foreign investments." Supreme Leader Ali Khamenei also asserted, "We and the U.S. have many differences. But this does not mean that we cannot adopt a regular policy in view of our national interests."[16]

Nevertheless, the second Bush Administration firmly shut the door on this historic opportunity in the President's 2002 State of the Union speech in which he denounced the Islamic Republic as part of an "Axis of Evil" along with Iraq and North Korea. To President George W. Bush, Iran was a repressive, aggressive, and ideological state that must be treated with hostility until its regime changed. Bush announced that Iran was a "major sponsor of terrorism," condemning the "unelected few" who were suppressing Iranians. Bush promised that the United States "would not permit the world's most dangerous regimes to threaten us with the world's most dangerous weapons."[17] In the context of the attack of September 11th and the subsequent U.S. retaliatory invasion of Afghanistan, this could only be seen as a direct threat. Since this occurred, there has never been any real chance of heading off the Iranian nuclear program with one remarkable exception.

The Iranian response was predictable. President Khatami had said a few

15 Ibid., pp. 503-05.

16 Tayeyh, *Hidden Iran*, p. 121.

17 Ibid., pp. 128-29.

months prior that there were "no obstacles preventing economic cooperation with the U.S." However, after having heard Bush's harsh rhetoric, Supreme Leader Khamenei now responded that the "drunkard shouts of American officials reveals the truth that the enemy is the enemy." Even President Khatami rejected Bush's statement as "war mongering and insulting toward the Iranian nation." The Iranian public was alienated by the State of the Union speech, and Iranians resented being defamed by an American politician.[18] It was a watershed in U.S.-Iranian relations. The Iranian establishment had reached far in the direction of the United States to achieve improvement in relations, but the Bush Administration's worldview prevented any success.

Perhaps the most unfortunate episode of all took place over a year after the "Axis of Evil" speech, following the invasion of Iraq by the United States. Remarkably, the United States was given a third and last opportunity to reach some sort of an accommodation with Iran. This time it appeared to present the chance to at least negotiate the framework of a "grand bargain."

In a proposed dialogue of "mutual respect," the Iranian proposal set forth a sweeping proposal. Tehran offered to end support for Hamas and Islamic Jihad and to pressure those militant groups to cease attacks on Israel. The Iranians would disarm Hezbollah and ensure it became a purely political party in Lebanon. They offered to open up their nuclear program completely to inspection to alleviate any fears of weaponization. Iran would sign the NPT and offered extensive American involvement in their nuclear program. With respect to terrorism, Iran offered full cooperation against all terrorist organizations—especially al Qaeda. On Iraq, Iran would work actively with the United States to establish political stabilization and a democratic, secular government. Finally, Iran would accept the 2002 Saudi peace plan for Palestine, recognizing and making peace with Israel in return for the latter's withdrawal from the occupied territories and a two-state solution.

In return, Iran wanted members of the MEK, which had been based in Iraq under Saddam's protection for some years, to be turned over to them. Tehran was willing to exchange al Qaeda militants in their custody, potentially making themselves a target, in exchange for members of the MEK. Iran wanted removal from the "Axis of Evil" list, an end to all U.S. sanctions, and recognition of war reparation claims from the Iran-Iraq war. The Iranians also requested that full access to nuclear, biological, and chemical technology within the limits of relevant treaties be respected and asked that there be greater recognition of Iran's legitimate security interests in the region. This was a remarkable, unprecedented proposal. If real, it could have led to a completely different situation in the Middle East. Subsequently, some argued that it was a ploy, others that it was real and presented an historic negotiating opportunity. Tragically, we shall never know.

At the State Department, Secretary Colin Powell wanted to accept this proposal as the basis for negotiations. However, Vice President Cheney and Secretary of Defense Rumsfeld reportedly blocked it in the White House, heading off any internal U.S. government consideration. Their plan was to overthrow the Iranian regime after finishing with Iraq, so negotiations seemed unnecessary. As Under-

18 Ibid., pp. 121, 129.

secretary of State John Bolton had put it, go to Baghdad and "turn right." The United States never responded to the proposal but sent a message to Iran through the Swiss government making it clear such a deal would not be seriously considered.[19] After January of 2002, Iran's approach to the United States in general and with respect to the nuclear program in particular was entirely tactical.

On August 5, 2005, the EU-Three (United Kingdom, France, and Germany), which had been conducting negotiations with Iran since late 2003, proposed significant economic and political cooperation if Tehran gave up its uranium enrichment program. Iran flatly rejected this proposal, and the EU-Three cancelled further negotiations. A day or two after that, the Isphahan conversion facility resumed operations after a moratorium of some months. In September, the IAEA Board passed a resolution warning Iran of a possible referral to the Security Council unless measures were adopted to increase transparency, reestablish suspension of uranium enrichment activities, and reconsider the heavy-water reactor at Arak; Iran rejected the resolution.

In early January 2006, Iran informed the IAEA of its intent to resume research and development of peaceful nuclear technology and a few days later removed the IAEA seals at enrichment sites. By then there had been considerable construction at Natanz, and operations could begin. The next week, the UN Permanent Five met in Jordan to consider the Iranian nuclear crisis. On February 4, a resolution was adopted by the IAEA Board calling on the Director General to refer Iran to the UN Security Council. The vote was 27 to three, with five abstentions and with all the UN Permanent Five voting in the affirmative.[20] On March 8, the IAEA formally submitted a report on Iran's nuclear program to the Security Council. With this submission Iran could no longer claim to be a NPT party in good standing. Under the treaty, Iran no longer had an "inalienable" right to peaceful nuclear technology and the Security Council would be able to approve sanctions on Iran.

Thus, the desultory negotiation phase came to an end. The EU-Three did sporadically continue to meet with Iran, and the U.S. Undersecretary of State, William Burns, did join a negotiating session in 2008. The emphasis shifted to sanctions rather than negotiations, neither of which has been successful on slowing the Iranian nuclear program. Subsequently, the Security Council adopted three more sanctions resolutions, but as Washington seeks to engage Iran on its nuclear program, it must recognize that hostile rhetoric and threats appear to have only a limited effect on Iran's behavior.

When President Obama took office, he inherited a situation in Iran that appeared unamenable to a solution. Over the previous six years, negotiations had made no progress and sanctions had failed to have their intended effect. President Obama announced that he would try for a year to press for real negotiations with Iran. If this did not succeed, then other measures would be considered. President Obama sent messages to Tehran, but the establishment in Tehran did not appear to

19 Trita Parsi, *Treacherous Alliance: The Secret Dealings of Israel, Iran, and the United States* (New Haven: Yale University Press, 2007), pp. 243-57.

20 Shahram Chubin, *Iran's Nuclear Ambitions* (Washington, D.C.: Carnegie Endowment for International Peace, 2006), p. xx.

have any interest in negotiations. Then came the disputed Iranian presidential election of June, 2009, and the subsequent near rebellion by the opposition, followed by a brutal crackdown by the government.

While the Obama Administration initially had concerns about negotiating with the Iranian government under such circumstances, Washington decided to go ahead. In the fall of 2009, the IAEA developed an agreement whereby Iran would transfer most of its stocks of low enriched uranium (LEU) to Russia to be modified to up to about 20 percent enriched. The process was suitable for the production of medical isotopes by Iran's TRIGA research reactor, which was then running low on fuel. After enrichment modification, the uranium would be transferred to France for fabrication into research reactor fuel and then returned to Iran. This agreement, at the time, would have significantly reduced Iran's bomb-making ability for about a year while fulfilling an Iranian requirement. The Iranian representative agreed to it in Vienna, but it was quickly disavowed by Tehran.

Shortly after the negotiation at the IAEA, at the G-20 meeting in Pittsburgh, President Obama, President Nicolas Sarkozy of France, and Prime Minister Gordon Brown of Great Britain announced the discovery of a secret uranium enrichment site at a Revolutionary Guard base near the holy city of Qom. CIA monitors actually had discovered it some time before. The allied leaders demanded that construction be immediately halted and took the position that the failure to notify the IAEA of the commencement of construction several years before was a violation of Iran's NPT Safeguards Agreement. The leaders also insisted that the site be made immediately available for IAEA inspections. Upon their arrival a few weeks later, IAEA inspectors estimated that the facility could house up to 3000 centrifuges, which was impractical for commercial enrichment but not for a bomb program.

On November 27, 2009, the IAEA board adopted a resolution declaring the secret site to be an NPT violation, demanded a work stoppage, and once again called for the referral of this new discovery to the UN Security Council. In retaliation, Iran announced plans to build ten more sites, five immediately, a highly impractical idea.

In May 2010, the presidents of Iran, Brazil, and Turkey signed a new Tripartite Agreement in Tehran. It was similar to the previously proposed IAEA Agreement. Iran would deposit 1200 kilograms—somewhat over half of its then current supply—of LEU, low enriched uranium (two to four percent), in Turkey for "safekeeping" under IAEA safeguards. In return, Iran would receive within one year 120 kilos of medical-research reactor fuel enriched to 19.75 percent for use in its research reactor supplied by the "Vienna Group" (the United States, France, Russia, and the IAEA). Nothing else would be affected. Iran would continue to enrich, both to two to four percent as well as to twenty percent.

However, while the fall 2009 agreement would have removed more than three-quarters of Iran's LEU, this new agreement would remove only around half, leaving Iran with enough enriched uranium to make one nuclear weapon if further enriched to the 90-percent level. Western nations had little interest in this proposal, regarding it largely as a ploy to avoid sanctions.

By this time, President Obama's one-year negotiating time frame had long

since expired, stalemate still reigned, and the Iranians continued to move ahead, albeit with some doubts expressed internationally about the rapidity of progress on the program. The focus was now on achieving a fourth UN Security Council Resolution in response to the secret site. A new draft resolution approved by all of the Permanent Five, was released in New York the day after the announcement of the agreement in Tehran. Secretary of State Hillary Clinton asserted that the draft resolution was the appropriate response to the announcement in Tehran. The Security Council adopted the resolution by a vote of twelve to two, with only Brazil and Turkey voting no.

While Iran continues to assert that its program is peaceful, aimed only at nuclear power, the entire history of its program appears to tell a different story. Beginning with the Shah in the 1970s, there was an apparent interest in the prestige and power associated with nuclear weapons. When Ayatollah Khomeini disavowed an interest in nuclear weapons, the Bushehr reactor program was put on the shelf; when Iranian policy changed, the reactor program was revived. A.Q. Khan was not known as a promoter of nuclear power; he was selling nuclear weapon capability. Iran had a nearly twenty-year clandestine relationship with him and acquired from him centrifuge enrichment technology and possibly the design of a Chinese nuclear weapon.

There are other indicia, such as the many links of the nuclear program to the military, the domination of the program by the quasi-autonomous Revolutionary Guard, as well as constantly changing explanations and the destruction of evidence and buildings before inspection. The assertion of an "inalienable right" to peaceful nuclear technology is not relevant in this case. The NPT does grant such a right but only to treaty parties in compliance with their non-proliferation obligations. Iran has twice been found by the IAEA to be in violation of NPT-related obligations. The Iran case has been referred to and is in the hands of the United Nations Security Council. So until a nuclear explosive test removes all doubt, or until a deal can be reached to satisfy all sides, one must prudently assume that the Iranian objective is, in fact, nuclear weapons or at least the achievement of a nuclear weapons technical capability from which status weapons could be quickly produced if desired.

What does this mean in terms of policy for the international community? The time for truly effective negotiations with Iran likely ended early in 2002 with the "Axis of Evil" speech, with the one exception noted. Sanctions have not changed Iranian behavior. The new sanctions—the fourth round—adopted by the Security Council on June 9, 2010, also seem unlikely to change Iranian behavior, but possibly they may have some useful effect. As Steven E. Miller, Director of the International Security Program at Harvard University's Belfer Center for Science and International Affairs, said, "I think that by default we end up with sanctions because we don't know what else to do."[21] For his part, President Ahmadinejad the day before had declared. "If the U.S. and its allies think they could hold the stick of

21 Neil MacFarquhar, "U.N. Approves New Sanctions to Deter Iran," *New York Time*s, June 10, 2010, pp. A1, A4.

sanctions and then sit and negotiate with us, they are seriously mistaken."[22] Russia and China probably will be opposed to further sanctions.

The military option does not seem practical either. Secretary Gates has said that it would only delay the Iranian program by "a few years." To accomplish anything more than such a temporary delay would appear to require something truly massive, conceivably a month or more of sustained bombing, followed by an invasion by Special Forces and perhaps mainline military units. This will not happen, and, in any case, Iranian retaliation would be severe. They would probably begin with attempts to devastate Gulf State allies of the West with missiles and air attacks. All of this being the case, perhaps the policy of the international community should be to try to continue negotiations in the hope that something will come of them one day.

In any case, Iranian acquisition of nuclear weapons, or even a nuclear weapons capability, will create many international complications. Since the primary cause of all this is the unremitting hostility of successive Iranian Islamic Republic governments toward everything Western, maybe someday all this will cease, Iran will become more of a country than a cause, and these sorts of calculations will no longer need to be made. For now, after the opportunities lost in the 2001-2003 timeframe, Iran remains a problem with no easy solution.

The roots of the North Korean nuclear problem stretch deep into history as well. In 1964, North Korean dictator Kim Il Sung, who was installed by the Soviets after World War II and later triggered the Korean War when he launched an attack on the South, journeyed to China seeking nuclear weapons technology. Nuclear cooperation agreements had been signed with the Soviet Union in 1956 and 1959, as well as with China in 1959. The United Nations forces, led by the United States, humbled Kim's forces and an uneasy truce had succeeded the Korean War in the region of the 38th parallel, near the original border between North and South. These cease-fire arrangements, referred to as the Demilitarized Zone or DMZ, still are in place some fifty-seven years after the cessation of hostilities.

When China tested its first nuclear weapon in 1964, it became a brand new member of a nuclear weapons club that at the time included four other countries (the United States, the United Kingdom, the Soviet Union, and France). China was not interested in becoming a nuclear-weapons proliferator and politely declined Kim's request. Muammar Qadhafi also tried to obtain nuclear technology from China in 1970, and Kim Il Sung made his request again in 1974, at a time when South Korea was exploring a nuclear option. China's response to both requests was negative once again. In the late 1970s, Kim gave the order to North Korean government officials to begin seeking nuclear weapons on their own. The Soviet Union had sold a small research reactor, capable of little beyond laboratory work, to North Korea in the 1960s. The reactor was built at Yongbyon, north of Pyongyang; the Soviets also established a research center there. In the early 1980s, North Korea, or the Democratic People's Republic of Korea (the DPRK), began building a larger research reactor at Yongbyon in the twenty to thirty megawatts (thermal) range - five

22 David E. Sanger, "U.S. Presses Its Case Against Iran Ahead of Sanctions Vote," *New York Times*, June 8, 2010, *available at* http://www.nytimes.com/2010/06/08/world/middleeast/08nuke.html.

megawatts (electric). That reactor, which became operational in 1986, produced heat and electricity and also plutonium. By the end of the decade, North Korea was mining uranium and building two larger reactors, fifty megawatts (electric) and two hundred megawatts (electric).

With so much nuclear activity underway, North Korea came under heavy pressure from the Soviet Union to join the NPT. It did so in 1986, the year its new five-megawatt reactor at Yongbyon became operational. Initially, North Korea refused to negotiate a Safeguards Agreement with the IAEA as required by the NPT. Finally, facing significant international pressure, North Korea signed such an agreement with the IAEA in early 1992.

When IAEA inspectors, led by IAEA Director General Hans Blix, began their work in the DPRK, they soon discovered an undeclared reprocessing facility. North Korea referred to the facility as a "radioisotope laboratory," supposedly capable of only research. Inspectors also became suspicious of two undeclared waste storage sites. They asked to inspect them and were denied permission to do so. It was later surmised that North Korea shut down its five-megawatt reactor in 1989 for three months, long enough to withdraw fuel rods sufficient to reprocess enough plutonium for one to two weapons. Inspectors analyzed nuclear waste samples, and the results of their analysis indicated that more than the ninety grams of plutonium North Korea had admitted to producing had actually been reprocessed.

In February 1993, the board of the IAEA in Vienna requested a "special inspection" of the two waste storage sites which it was believed, based on satellite imagery supplied by U.S. intelligence,[23] could indicate that there had been undeclared plutonium production.

The IAEA board had initially been divided over whether to make such a demand, but the satellite imagery persuaded those who were uncertain. The board gave Pyongyang 30 days to cede to its request—but did not need to wait nearly that long for the response. North Korea rejected the inspection request the very next day and, about two weeks later, on March 12, gave the three-month notice required by the withdrawal provision of the NPT. North Korea halted IAEA inspections of any kind. This was a significant shock to the international community as no state had ever before exercised the withdrawal provision of the NPT or that of any other international arms control or non-proliferation agreement. Near the end of March, the IAEA board approved a resolution stating that it could no longer certify that illegal nuclear material diversion from peaceful activities to a nuclear weapons program had not taken place in North Korea. The board forwarded the DPRK case to the UN Security Council for consideration of sanctions.

Even though it was by no means certain that China, one of the Security Council's permanent members, would support imposing international sanctions on an already destitute North Korea, messages arrived from North Korea indicating a willingness to negotiate. By the early spring of 1993, the United States and Kim Il Sung's increasingly isolated regime agreed to negotiate. In June, on the eighty-ninth day of the ninety-day notice period Pyongyang had given, North Korea "suspended"

23 Mitchell Reiss, *Bridled Ambition: Why Countries Constrain Their Nuclear Capabilities* (Washington, D.C.: Woodrow Wilson Center Press, 1995), pp. 249-50.

its withdrawal from the NPT.[24] (Whether it was truly a "suspension" or a cancellation of plans to withdraw has been debated for years.)

When North Korea finally did withdraw from the NPT in January 2003, officials in Pyongyang claimed they could withdraw with just one day's notice since they had used up eighty-nine of the ninety days in 1993. Some United Nations and other international organization officials, as well as officials of some states, argue that the 90-day clock begins all over again under the NPT and that, as a result, North Korea's withdrawal was invalid. They insist that North Korea remains legally bound by NPT provisions. At some NPT meetings, an empty chair is maintained for the DPRK.

When Kim Il Sung, at the age of 82, suddenly died of a heart attack in July 1994, his son and successor, Kim Jong-il, did not change his policies. In October of 1994, the United States and the DPRK reached an "Agreed Framework,"[25] meaning in effect that the agreement was not a treaty requiring approval by the U.S. Senate. Once the Agreed Framework was in place, the fuel rods that had been withdrawn from the five-megawatt reactor stayed in the spent fuel pond and the reactor ceased operations, all under continuous monitoring by returning IAEA inspectors. Under the Agreed Framework, the nuclear program that North Korea had begun in the 1980s was frozen. The five-megawatt reactor could make enough plutonium for one nuclear weapon a year. The Agreement was in force for eight years, meaning that it prevented the production of material for eight nuclear weapons that North Korea might otherwise have acquired.

North Korea nonetheless continued to expand its ballistic missile program and in 1998 launched a Taepodong I ballistic missile over Japan into the Pacific. Pyongyang claimed this to be an attempt to place a satellite in orbit, but this assertion was widely regarded as implausible. The third stage of the missile failed, but the test outraged Japan, which had long feared an unpredictable, nuclear-capable neighbor to its west. After threats of UN sanctions, North Korea agreed to a moratorium on further such launches. The message was clear: the DPRK could develop a ballistic missile delivery system for nuclear weapons threatening its regional neighbors.

However, another threat, even more menacing than the DPRK ballistic missile program, soon emerged. In the early 1990s, A.Q. Khan, the scientist who had pioneered Pakistan's nuclear-weapons program, brokered an arrangement to transfer uranium enrichment technology to North Korea in exchange for designs and parts for North Korea's medium-range ballistic missile, the Nodong. This missile, capable of delivering nuclear weapons, became known as the Ghauri in Pakistan. Khan made his first trip to North Korea in 1993 with then Prime Minister Benezir Bhutto, who was seeking friendly relations with reclusive North Korea. On the side, Khan made his deal. He made perhaps up to a dozen more trips in Pakistani air force planes—thereby implying government or at least armed forces support—to North Korea in the 1990s, delivering plans, parts, as well as completed centrifuge machines.

North Korea is a dangerous state. It has a long track record of being will-

24 Ibid., p. 253.

25 Ibid., pp. 275-76.

ing to sell anything to anyone for its own benefit and a history of state terrorism against South Korea. A nuclear-armed North Korea, with ballistic missiles capable of reaching targets throughout Northeast Asia, posed a double danger. First, the closed, inscrutable government of dictator Kim Jong-il, who may be preparing to pass control to his youngest son, could sell nuclear weapons to Iran or to terrorist organizations. He could also transfer bomb production technology as it did to Syria in the 2005-2007 time-frame. Second, a nuclear-armed North Korea, because of its nuclear capable medium range missile, the Nodong, is a grave threat to Japanese and South Korean cities.

Pyongyang's track record also includes a certain realpolitik and willingness to negotiate. Above all, the North Korean regime, grappling with overwhelming poverty, occasional famine, and few allies, is interested in survival and in economic benefits. Military action against the DPRK is not an attractive option due to the threat to Seoul from the huge North Korean artillery and rocket forces arrayed along the DMZ border less than twenty miles away. Diplomacy is the only practical option.

In October 2000, Kim Jong-il sent his number two official, Vice Marshall Jo Myong Rok of the Korean People's Army, to Washington. Marshall Jo met with President Clinton in the Oval Office along with other high-ranking national security officials, including Secretary of State Madeline Albright and National Security Advisor Sandy Berger. The marshal wore his dress uniform covered with medals; he was cordially greeted by President Clinton.

Marshall Jo sat next to President Clinton in the Oval Office meeting, holding in his hand a brown folder. The President asked Jo if it was a letter for him, whereupon Marshall Jo stood and formally handed over the folder to Clinton. Clinton promptly opened the folder and read the letter, pronouncing it a "good letter." In the missive, Kim Jong-il stated that he was prepared to stop the production, export, and use of long-range ballistic missiles. Then Marshall Jo unveiled another surprise. He said that on behalf of Kim Jong-il, he would like to invite President Clinton to come to Pyongyang to sign an agreement on missiles. "If you come to Pyongyang, . . . Kim Jong-il will guarantee that he will satisfy all your security concerns," he said to President Clinton. Indeed he went further and said to Clinton, "I need to secure your agreement to come to Pyongyang. I really need to take back a positive answer."[26] Clinton was noncommittal but generally spoke positively with Jo. He wanted the DPRK to understand that they could work with him. All in the room sensed that a deal might be possible that might take the two countries all the way to a position that would enable them to normalize relations.

Since the time of the Agreed Framework negotiation, North Korea's objective had been, for its own purposes, to end its adversarial relationship with the United States. For the time being, the Agreed Framework had accomplished that goal, but it was not long before the agreement began to fray. After the Clinton-Jo meeting, a communiqué was released which stated that neither government would have "hostile intent" toward the other and that both were committed to "build a new

26 Mike Chinoy, *Meltdown: The Inside Story of the North Korean Nuclear Crisis* (New York: St. Martin's Press, 2008), p. 25.

relationship free from past enmity."[27] Clinton delayed making his decision to go to North Korea and sent Secretary Albright to Pyongyang some weeks later to convey President Clinton's views and also to prepare for a possible visit by the President. Charles Kartman, Jack Pritchard, Kartman's deputy, Wendy Sherman, State Department Coordinator on policy toward the DPRK, and, Assistant Secretary of State for East Asian Affairs Stanley Roth accompanied the Secretary on this trip.[28]

In discussions over two days with Albright, Kim Jong-il displayed a mastery of his brief, indicating that a missile deal was possible and making it clear that he wanted improved relations with the United States. He expressed the hope that the United States would no longer view North Korea as an adversary. When the American election day arrived on November 7, 2000, the DPRK government made clear that it wanted relations with the United States to continue to improve and hoped to do business with whoever became the next President. In a commentary on the election in Pyongyang's official newspaper, it was asserted "the improved relations between the two countries are in line with the desire and interests of the two peoples."[29]

The long stalemate over who would be the next President greatly complicated decision-making on North Korea within the Clinton Administration. Clinton very much wanted to respond to the DPRK's conciliatory gestures and visit Pyongyang. Even though many government experts believed the North Koreans would accede to an agreement if Clinton took the risk, the long internal debate finally concluded that too many loose ends remained and that one more visit by a high-level emissary was necessary before a summit trip could be agreed. Wendy Sherman, with an accompanying delegation, could be sent instead, with a proposed date for President Clinton's trip in her pocket. She would put the visit on the table as soon as the remaining principal issues, especially verification, were resolved.[30]

Many experts in government were convinced, however, that if Clinton simply took the risk and went, the North Koreans would see to it that agreement was in fact reached. Ultimately the White House did not want to take a risk until the election impasse had been resolved. The Clinton Administration did not want to go forward even with the two-phased approach of sending Sherman first. Clinton officials believed that the new President needed to be briefed on this, but he could not be briefed until they knew who had won the election. Finally, George W. Bush was declared to be President. By this time, it was too late for President Clinton to go to North Korea.

As Bush entered office, the DPRK attempted to signal that it hoped to resume a dialogue with the new Administration. On February 8, 2001, at the Atlantic Council in Washington, Li Gun, the Deputy North Korean UN Representative, said, "We hope the Bush administration maintains the U.S. engagement policy to-

27 Ibid., p. 26.

28 Ibid., pp. 26-27.

29 Ibid., p. 35.

30 Ibid., pp. 35-36.

ward North Korea."[31] The North Koreans were reassured by the fact that there remained the communiqué signed by President Clinton and Marshall Jo pledging "no hostile intent" between the two countries.

However, the personnel of the new Administration, particularly on the National Security Council staff, made no secret of their distrust of North Korea in general and of Kim Jong-il in particular. They also made clear their desire to scuttle the Agreed Framework as fast as possible. State Department experts explained to skeptical Bush Administration staffers that the DPRK regarded the Clinton-Jo Communiqué as the foundation stone of their relationship with the United States in much the same way that China regarded the Shanghai Communiqué of 1972 signed by Richard Nixon. This argument was not accepted by senior NSC staff officers. Secretary Powell and Deputy Secretary Richard Armitage, on the other hand held quite different views. While Powell thought that it would have been a mistake for Clinton to have gone to North Korea, he very much wanted to continue the Clinton engagement policy. Powell brought James Kelly to the post of Assistant Secretary for East Asia. Although Kelly was somewhat skeptical on the Clinton policy toward North Korea, he was experienced and highly respected. He was a supporter of the Agreed Framework.[32]

With the new Administration now in place in Washington, South Korean President Kim Dae-jung was worried about a possible change in policy in Washington toward engagement with the DPRK. Thus, South Korea decided to press for an early meeting between Bush and Kim. Many career U.S. officials and outside experts were counseling South Korea to wait, saying that the new Administration was not yet ready for a summit meeting. There was an ongoing policy review on North Korea, and it was believed that many top officials had not shifted intellectually from being part of a presidential campaign to governing. In other words, they had not moved from rhetoric to practical decisions on policy. However, Kim was determined to come to Washington. He had been a long-time dissident who resisted military dictatorships in South Korea. He had survived several attempts on his life. Upon coming to office, he had created a Sunshine Policy of engagement with the North, for which he had received the Nobel Peace Prize. He was convinced that he could persuade President Bush to continue the dialogue with Kim Jong-il.[33]

The meeting was set for March 7, and in preparation for the visit, the South Korean Foreign Minister met in Washington with Secretary Powell. Powell reassured him that the administration was supportive of the Sunshine Policy and that it intended to continue with the Agreed Framework.

On the day President Kim arrived in Washington, National Security Advisor Condoleezza Rice gave an interview to *The New York Times*. Speaking as an anonymous "senior administration official," she said the DPRK as well as Kim Jong-il were a "problem." At the same time, Powell met with reporters and indicated that the Bush Administration planned to pick up where the Clinton administration left

31 Ibid., p. 43.

32 Ibid., pp. 44-48.

33 Ibid., pp. 49-50.

off in dealing with North Korea.[34]

The next morning, Secretary Powell hosted President Kim and his National Security Advisor at Blair House for breakfast. Powell was accompanied by several senior State Department officials. The meeting, according to one of the participants, was "wonderful." Powell and Kim were comfortable with one another, and Powell made clear that the two governments shared the same view as how to deal with the DPRK. The policy of engagement would continue. Kim used the occasion to try out the presentation that he planned to make to the President. The Americans who were present believed the arrangement was brilliant. [35]

However, the situation was to change quickly. The following morning, the *Washington Post* ran a story about Powell's meeting with the press the day before the talks. The headline read "Bush to Pick Up Clinton Talks on N. Korean Missiles." The headlines appeared to cause a firestorm at the White House. Powell received a call from Rice at Blair House as he was preparing to head over to the White House with President Kim.. Rice demanded that Powell set affairs straight with the press right away before Bush and Kim did their press conference after their meeting. So Powell had to leave the meeting between Kim and Bush, and, in a humiliating way, say to a confused group of reporters waiting outside the Oval Office that no agreement really had yet been reached with regard to the engagement policy toward the DPRK, andthat the situation would need additional study.[36]

The meeting between President Bush and President Kim has accurately been described as "a disaster." Kim Dae-jung was seventy-five years old. He had suffered for years struggling for human rights against South Korean dictators of the past. He had been jailed, tortured, and had survived a death sentence on sedition charges and an assassination attempt.[37] He was the popularly elected President of South Korea and had been awarded the Nobel Prize the previous year. He was used to and deserved respect. George W. Bush had no clear idea who Kim Dae-jung was or what he represented.[38] He only knew he did not like being lectured on the importance of negotiating with Kim Jong-il, a man whom he had never met but somehow intensely disliked; to be urged to follow the policies of former President Clinton; and to be friendly with a man who had publically questioned the value of ballistic missile defense, one of his Administration's top two objectives. After Kim made the same speech to Bush that he had given at Blair House, Bush challenged Kim and his ideas and made it clear he would not follow the Clinton policies toward North Korea and would not resume the missile talks in the foreseeable future. Basically, it was a posture of: we detest Kim Jong-il, and we do not agree with your

34 Ibid., p. 53.

35 Ibid., p. 53-54.

36 Ibid., p. 55

37 Ibid.; Choe Sang-hun, "Kim Dea-jung, Ex-President of S. Korea, Dies at 83," *New York Times*, Aug. 19, 2009, p. A25.

38 See Chinoy, *Meltdown*, pp. 50, 56.

Sunshine Policy.[39] To the Bush Administration, North Korea represented one of the most compelling justifications for its pursuit of a ballistic missile defense policy.[40] So the Administration did not want to forego North Korea's hostility. This policy view thus was solidly based on U.S. domestic policies as opposed to national security analysis. After this unfortunate meeting, between President Bush and his South Korean counterpart the internal Administration policy review with respect to North Korea continued.

When the policy review finally ended in early June, President Bush announced that the U.S. was prepared to resume "serious negotiations" with North Korea, but there were conditions, such as improved verification of the Agreed Framework, a ban on missile exports, controls on the DPRK missile forces, and reductions in North Korea's conventional military capability. In addition, there was no renewal of the "no hostile intent" communiqué from the Clinton Administration. The Bush Administration wanted concessions on all these issues from the DPRK before it would do anything. This could not possibly look any other way but intensely hostile to North Korea.

A few months later, shortly after the terrorist attacks on the United States of September 11, 2001, North Korea in an official statement described the attacks as "a very regretful and tragic incident" and stated that the DPRK was "opposed to all forms of terrorism."[41] Throughout the fall of 2001, the DPRK tried its best to distinguish itself from the international terrorists to avoid being grouped into the "bad guy" camp. The DPRK signed on to UN counter-terrorism conventions that it had not theretofore ratified. However, it was no use; President Bush denounced Kim Jong-il on October 16th as he was leaving for an Asia Pacific Economic Forum meeting in Shanghai, saying that Kim should stop spreading weapons of mass destruction around the world. In November, President Bush made a similar denunciation while speaking to reporters in the Rose Garden, asserting that to have a relationship with the U.S., North Korea would have to stop proliferating weapons of mass destruction, implying that the DPRK would be "held accountable."[42]

Then the Nuclear Posture Review, which is the Pentagon Report on nuclear policy, was completed in December 2001. Portions of it became public, in particular the section asserting that the United States needed to maintain nuclear weapons for the possible use of such weapons against Iraq, Iran, Syria, Libya, and North Korea at that time, even though perceived as adversaries, all NPT non-nuclear-weapon states. This campaign of increasing the negative pressure came to its high point in the President's State of the Union speech in January of 2002. The relevant section denounced the "Axis of Evil" threatening the United States—in its near final draft mentioning only Iraq. The White House, in order not to be seen as singling out Iraq, instructed the speech writer to add Iran (which had just been helping the

39 Ibid., p. 55.

40 Ibid., p. 59.

41 Ibid., p. 65.

42 Ibid., pp. 67-68.

U.S., in its fall effort to drive the Taliban out of Afghanistan and establish a new government) and North Korea (so as to have the President appear less anti-Islamic in his State of the Union speech). The public in South Korea saw this language as possibly justifying a war on the Korean Peninsula. The U.S. buildup against Iraq was well under way. Naturally the DPRK reacted angrily but at least for the short term acted with caution.[43]

In spite of all of this, the chief advisor on North Korean affairs to President Kim Dae-jung went to Pyongyang for several days of talks. After his return, he announced that Kim Jong-il had said that he was prepared to accept a visit by an American government envoy. After many months, the DPRK expressed a willingness to accept the call of the Bush Administration for negotiations. Secretary Powell pressed for authority to respond with what became known within the United States government as the "Bold Approach," in other words comprehensive talks. This touched off another internal debate, but this time, the State Department prevailed, and Assistant Secretary for East Asian and Pacific Affairs James Kelly was authorized to lead a delegation to Pyongyang on July 10, 2002.[44]

A new surprise was in store during their visit. The Central Intelligence Agency approached the State Department with the message that they had new information. For some time, CIA officials said, the DPRK had been conducting a secret uranium enrichment program to develop nuclear weapons,[45] which was highly contrary to the spirit and the rationale of the Agreed Framework. The information caused a full review of available intelligence to be ordered. The review disclosed an equipment procurement effort that hitherto had not been noticed. The operation began on a small scale in the 1990s but had become a major program by 2002. The new intelligence may have come from an agent inside North Korea or it may have come from an increasing understanding of Pakistani nuclear scientist A.Q. Khan's illegal, clandestine, nuclear proliferation network.

By the end of the summer, the CIA had pieced together quite a full picture of the DPRK enrichment program. The operation had remained a small research effort until 2002 when it became a much more serious program. Principals were briefed by the CIA on the program in September, and it was concluded that business as usual was no longer possible with the DPRK. It was feared that North Korea might in fact be close to a uranium bomb. Kelly was authorized to go to Pyongyang with the sole mission of accusing the DPRK of this program, not to negotiate.[46]

On the morning of October 3, 2002, Assistant Secretary Kelly and his small delegation landed at Pyongyang airport. His mission, after further machinations in Washington, was to accuse the DPRK of violating the Agreed Framework Agreement, at least in principle, and nothing more. The North Koreans were expecting a negotiation. As was common practice, the North Koreans would host a lavish

43 Ibid., pp. 68-72.

44 Ibid., p. 80.

45 Ibid., p. 82.

46 Ibid., pp. 114-15.

dinner for the visiting American delegation. Normally, the U.S. would reciprocate the second night, but Kelly was under further orders to rub it in and refuse to reciprocate. The North Koreans were notified of this in advance, confirmed upon Kelly's arrival. Due to this behavior, the DPRK housed Kelly and his delegation in a downtown hotel where they had to check in like regular tourists, rather than one of the guest homes maintained by the Foreign Ministry.[47]

The first meeting was scheduled that afternoon with Kim Gye-gwan, one of several vice-ministers and a long time interlocutor with Americans. This was intended as a short discussion. A larger meeting was scheduled for the next morning with Kim, followed by a meeting with the First Vice Minister, Kang Sok-ju, a key figure in the government. At the afternoon meeting Kim invited Kelly to begin. Kim said he wanted to learn about current U.S. policy toward North Korea and to understand the true intention of the United States with respect to dialogue with the DPRK. Kelly decided that instead of an introductory discussion he would read his prepared script. The basic message was that the President had planned to have serious discussions with the DPRK about transforming the U.S.–North Korean relationship. However, the U.S. had irrefutable evidence that North Korea had embarked on a secret program to develop nuclear weapons through uranium enrichment, so such dialogue was no longer possible. Kelly's brief also covered subjects such as U.S. concerns about the Agreed Framework, terrorism, missiles, conventional forces, and human rights.[48]

After Kelly's presentation, Kim asked for a break and left the room, probably to brief the First Vice Minister. When he returned, his response was relatively brief, rejecting the charges about a covert HEU program out of hand and stating that those who were opposed to a better relationship between the U.S. and the DPRK fabricated such charges. He then presented standard talking points on the relationship. The dinner that evening was uneventful although the North Koreans remained offended that the U.S. would not be reciprocating.[49]

The next morning, Kelly presented virtually the same script that he had the previous afternoon. Kim responded that it was clear that the United States now intended to pursue a policy of "strangulation" of North Korea and intended to change North Korea's system by force. Therefore, only a hard-line response would be appropriate, involving conventional forces and missiles. The HEU charges did not merit a response he said.[50]

Not long after this discussion, the U.S. delegation had a nearly one hour meeting with Vice Minister Kang. This perhaps was the most fateful meeting in the long and tortured history of U.S.-North Korea discussions. Kang said that he had been thoroughly briefed by Vice Minister Kim and had conducted discussions with

47 Ibid., pp. 115-16; Charles L. Pritchard, *Failed Diplomacy: The Tragic Story of How North Korea Got the Bomb* (Washington, D.C.: Brookings Institution Press, 2007), p. 35.

48 Pritchard, *Failed Diplomacy*, p. 35.

49 Ibid., p. 36.

50 Ibid., p. 37.

senior government officials through the night. He said that the United States had completely destroyed the Agreed Framework as a result of designating North Korea as part of the "axis of evil," establishing a preemptive nuclear strike policy, and singling the DPRK out for nuclear attack. He noted that Article III of the Agreed Framework provided that the U.S. would not use or threaten to use nuclear weapons against the DPRK. The new U.S. policy was a direct violation of the arrangement. The DPRK had no choice but to modernize its armed forces. The U.S. claims that the DPRK has a uranium enrichment program for the fabrication of nuclear weapons. The DPRK was prepared to develop even more advanced weapons and put itself on an equal footing with the United States to discuss denuclearization.[51]

"We are part of the axis of evil and you are gentlemen," Kang quipped. "This is our relationship. We cannot discuss matters like gentlemen. If we disarm ourselves because of U.S. pressure, then we will become like Yugoslavia or Afghanistan's Taliban, to be beaten to death." All eight of the Americans concluded that, even though Kang did not explicitly say so, they believed his statements amounted to an admission by North Korea that it had a HEU nuclear- weapons program. Kang asserted after discussion that the DPRK had to counter the "physical declaration of war" by the United States. The American delegation's conclusion was subsequently and consistently denied by the DPRK.[52]

In response, the United States lobbied its partners in the Korean Energy Development Organization (KEDO) to terminate heavy fuel oil (HFO) shipments to North Korea. Also, on November 14, 2002, KEDO announced the suspension of HFO shipments in a statement that condemned North Korea for building nuclear weapons. North Korea's response came late the next month when it announced the expulsion of IAEA monitors and the removable of IAEA monitoring devices from the nuclear complex at Yongbyon. On January 10, 2003, North Korea declared that the ninety-day NPT withdrawal period, which had begun ten years earlier, was now completed. The DPRK was no longer a party to the Nuclear Non-Proliferation Treaty.[53]

In 1994, the Clinton Administration had been prepared to go to war if North Korea began reprocessing the fuel rods that had been pulled from the Yongbyon reactor for plutonium. In 2003 and thereafter, the Bush Administration simply stood by, hoping for a regime change and moving ahead as intended with its missile defense policy as North Korea reprocessed spent fuel twice in 2003 and 2005 from the Yongbyon reactor, thus enabling the DPRK to have obtained enough plutonium for an estimated up to eight additional nuclear weapons.[54] Adding this estimate to the plutonium gained from the three-month reactor shutdown in 1989 and the intelligence estimate for the DPRK, they could have generated enough plutonium for eight to ten nuclear weapons.

51 Ibid., pp. 37-38.

52 Ibid., pp. 38-39.

53 Ibid., pp. 40-44.

54 Ibid., p. 44.

In the spring of 2005, the Bush Administration's view was that North Korea could be planning an imminent nuclear weapon test. The previous year, in an attempt to reverse the deteriorating situation, the Six-Party Talks process involving North and South Korea, the United States, Japan, China, and Russia had begun with the stated goal of ending the DPRK nuclear program. In September 2005, with the United States downplaying the HEU issue, North Korea agreed in principle to terminate its nuclear program pursuant to the negotiations. However, a long period of stalemate followed with no further forward progress in the negotiations. On October 9, 2006, North Korea performed a nuclear test. The test was largely a failure in that an intended four kiloton test resulted in considerably less than a one kiloton explosion.[55] The nuclear explosion in the DPRK was detected, down to the kiloton, all over the world by seismic stations established by the Comprehensive Test Ban Treaty Office. The test signaled a new, more dangerous era in which one of he world's most unstable regimes could evolve into a nuclear weapon capable state, putting its regional neighbors in harm's way.

A short time afterward, North Korea signed an agreement to close down its five-megawatt reactor at Yongbyon and begin the termination of its nuclear program. In July 2007, the reactor was closed down in the presence of IAEA inspectors.[56] All of these developments came in spite of the fact that President Bush somewhat earlier had once again had attacked Kim Jong-il personally in a press conference on April 28, 2005, volunteering that "Kim Jong-Il is a dangerous person. He's a man who starves his people. He [has] got huge concentration camps. And . . . there is concern about his capacity to deliver a nuclear weapon."[57] However, after one attempted North Korean test, the negotiations seemed to be finding the right track. Secretary of State Rice had sufficient clout with the White House to press forward with a diplomatic effort at State in the face of the continued White House preference for regime change, which had become a virtual non-policy.

In 2007, progress was made in disabling the Yongbyon reactor and in understanding the size and scope of the DPRK program. There were many U.S. and other Western technical experts in North Korea and the DPRK who agreed to make a comprehensive declaration of the DPRK nuclear program by the end of 2007. The deadline was not met, but a declaration was finally made in June 2008. The declaration contained significant information on the plutonium program but nothing on the uranium enrichment program. Neither did the declaration include anything regarding an alleged DPRK role in building a research reactor in Syria capable of producing plutonium, which was destroyed by Israeli warplanes in September 2007.

As agreed, President Bush gave notice of his intent to remove the DPRK from the State Department's list of state sponsors of terrorism. When at the end of the forty-five-day notice period, the DPRK had not agreed to a verification arrangement for its declaration, which the United States insisted should occur. The Presi-

55 See ibid., pp. 150-52, 194-98.

56 Chinoy, *Meltdown*, pp. 325-26, 339.

57 Pritchard, *Failed Diplomacy*, pp. 15-16.

dent let the notice period expire without taking any action.

In early September, however, Washington announced that it would not take North Korea off the state sponsors of terrorism list until it agreed to an extensive written verification arrangement. The U.S. demanded full access to "any site, facility or location" deemed relevant to the DPRK nuclear program. David Albright, the U.S. nuclear expert, referred to this demand as a "verification wish list" comparable to what Iraq had agreed to "only after it was bombed." North Korea made a counter proposal in which it agreed to some provisions but objected to two key elements—visits to undeclared facilities and the taking of samples.[58] However, the United States did admit that it did not have any agreement in writing from the DPRK to establish a verification arrangement and had communicated its wish for a verification system only verbally. The United States added this unilateral demand after North Korea already complied with the previous commitment. For their part, the DPRK declared it would restore and restart the Yongbyon reactor, thereby enabling them to make more plutonium if this dispute continued. Once again, it appeared as though the process with North Korea might be unraveling.

However, the situation began to improve in October. North Korea had barred inspectors from Yongbyon and had prepared to resume plutonium production. The U.S. then suddenly decided that the DPRK had shown progress and announced that a verbal agreement had been reached on verification. The agreement would permit visits to undeclared sites and for inspectors to take away environmental samples from facilities, which could determine how much plutonium had been produced. No text of such an agreement has ever been produced, but this did take the dispute out of the news until after the American elections.

On October 11, 2008, President Bush removed the DPRK from the State Department's state-sponsors-of-terrorism list. Inspectors were readmitted, and it appeared that the disarmament process was back on track. However, in November of that year, the DPRK refused to allow inspectors to remove samples from the Yongbyon nuclear complex, saying in a formal statement that it had never agreed to sampling. In December, the DPRK explicitly refused to agree to a written verification plan, thereby ending the nuclear disarmament process on the Korean Peninsula for the Bush Administration. Thus, in 2007 and 2008 some limited progress was made in the Six-Party Talks. In the end, however, the Bush Administration left an entirely unconstrained situation for its successor.

In the spring of 2009, North Korea conducted the launch of a communication satellite that experts concluded was a subterfuge to test the Taepodong-II missile, which in theory would be capable of reaching North America. The launch was judged to be a failure, but there was speculation that it was the intent of the DPRK to develop nuclear weapons that would be designed to be deployed on medium- and long-range ballistic missiles. Later in 2009, North Korea conducted a second nuclear test that was more successful and about six kilotons. The DPRK announced that it was now a nuclear weapon state. The nightmare scenario of a nuclear weapon capable North Korea possessing medium range ballistic missiles threatening the Northeast Asia region took a significant step toward reality.

58 Glenn Kessler, "Far-Reaching U.S. Plan Impaired N. Korea Deal," *Washington Post*, Sept. 26, 2008, p. A20.

In November 2010, Stanford Professor and former Director of the Los Alamos National Laboratory Siegfried Hecker and two Stanford colleagues were visiting North Korea. While there, they were unexpectedly given a tour of a formerly secret uranium enrichment facility with perhaps up to 2,000 centrifuges. Hecker said that he was "stunned" that North Korea had built such a plant so quickly. Some American officials speculated that the purpose of showing off the plant was to advertise this capability to other countries who might want to buy what it could produce. American officials regarded this plant as "significantly more advanced" than what Iran possessed and were of the view that this plant could not have been built so quickly unless there was a network of secret nuclear sites elsewhere.[59]

When the Bush Administration came into office early in 2001, it inherited a North Korean situation at a promising point. The long-running crisis had been contained, and the process was on a road toward resolution. The DPRK had in its possession perhaps enough plutonium for one on two nuclear weapons, but the plutonium program was capped and contained by the Agreed Framework. U.S. officials were beginning to learn the details of North Korea's illicit bargaining with A.Q. Khan; but at this stage, the DPRK had actually done nothing with the uranium enrichment nuclear-weapons program beyond receipt of the centrifuge technology from Khan and related research. There was near agreement on termination of the missile program, both the domestic and import parts, and progress had been made toward some sort of broad settlement with North Korea. However uneasy, it might have at least removed the DPRK from the ranks of rogue nations.

Today there is no Six-Party Talks process for the nuclear disarmament of North Korea, and there is an active ballistic missile development program. North Korea has conducted two nuclear weapon tests and has declared that it is a nuclear weapon state. These are developments deeply contrary to U.S. and world security interests. These developments could drive Japan and South Korea toward nuclear-weapons programs and thus serve as the death knell for the NPT. The DPRK could try again to help Syria create a nuclear-weapons program, continue to assist Iran on its program, and help the dictatorial regime in Myanmar establish a nuclear-weapons program. Lastly, the transfer of nuclear weapon material to terrorist organizations by North Korea cannot be ruled out of the question. All of this would bring President Kennedy's darkest fears closer to realization.

The nuclear-weapons programs in Iran and North Korea during the Bush Administration were allowed to develop into grave threats to the continued viability of the NPT as well as to United States national security and the peace and security of the world community. Nevertheless, in spite of all these very real dangers along with other threats, this does still remain a time of promise because of the commitment of President Obama and his Administration to nuclear-arms control and non-proliferation. Widespread support for progress on nuclear-arms control and non-proliferation in the international community as well as the growing support for Global Zero also make this endeavor possible. Of course, it is also a time of great difficulty because of the many overwhelmingly serious crises that were left

59 David E Sanger and William J. Broad, "U.S. Concludes N. Korea Has More Missile Sites," *New York Times*, Dec. 15, 2010, p. A13.

at the end of 2008. These crises include the world economic downturn, climate change, Afghanistan, Iraq, and Palestine, in addition to the matters discussed herein, particularly the cases of Iran and North Korea.

The polarized political situation in the United States is a serious obstacle to progress, and much time has passed. Great damage to the disarmament process has been sustained and the proliferation crises seem to grow steadily more dangerous. However, while the hour is growing late, it is not too late. Success remains possible; that safer and more secure world that all of us want can still be built. We must not give up. In a real sense, we must all truly keep the faith and, to paraphrase a famous British statesman, never, never give in.[60]

60 See Winston S. Churchill, "Speech at Harrow School," Oct. 29, 1941, *reprinted in* Robert Rhodes James, ed., *The Complete Speeches of Winston S. Churchill*, vol. 6 (London: Chelsea House, 1974), p. 6499.

The Twilight of the Neoadministrative State: *Crises, American Political Development, and the "New Interventionism"*

Robert F. Durant, American University

Introduction

Until the worldwide financial meltdown of 2007-2008, "collaboration," "partnerships," and "networks" were the catchwords and phrases of our time. In fact, public administration scholars routinely talked less about "government" and more about "governance." The latter incorporates the idea that public, private, and nonprofit organizations working together, rather than government agencies acting unilaterally, are the dominant mode for delivering goods, services, and opportunities to citizens. Theirs was a *non-interventionist* view of the role of the state, with many proponents viewing government as the source of, rather than the solution to, societies' ills.

By the end of the first decade of the twenty-first century, however, a full-blown assault was underway on the power of this neoadministrative state to deal adequately—technically, politically, and philosophically—with the world's formidable challenges. Perceptions of crises—immediate, looming, or contrived—stimulated such thinking, along with perceptions of government failure. The former included global warming, the war on terror, spiraling national and foreign debt, revelations of the contributions of crony capitalism and government deregulation to financial crises worldwide (e.g., in the United States, Greece, Iceland, and Portugal), and underfunded entitlement programs in major industrialized nations exacerbated by an aging population (e.g., in France). For the United States, the latter were ensconced in perceptions of misfeasance, malfeasance, and nonfeasance related to the Iraqi war, Hurricane Katrina, and—later—the BP Gulf oil spill.

It was the worldwide financial crisis of 2007-2008 and its aftermath, however, that brought about what many see as the most immediate and significant reinvigoration of the state—i.e., the federal government in the United States since the Great Depression of the 1930s. The unprecedented monetary size of the financial rescue efforts in the United States; the partial nationalization of nine major American banks and portions of the automobile industry; and the enactment or expansion of classic command-and-control regulatory authority in new federal laws regulating the mortgage, derivatives, and credit card industries proved enough for opponents of the Bush and Obama Administrations and prominent journalists to claim that the United States was on its way to state socialism (Meacham and Thomas 2009).

Relying heavily on my previous publications on this topic, this article argues that those who see the rise of a new Leviathan in the United States under the Bush and (especially) the Obama Administrations falsely conflate the "ends" or

"how much?" question regarding state intervention in markets with the "means" or "how implemented?" question. When one looks today and historically at the "how implemented" question in the face of domestic and global crises, one persistently finds interventionist state policies creating implementation structures which ensure that organized market interests preserve power, access, and influence in the administration of programs. The paper attributes this to the three causal mechanisms in American political development: path dependency, the constitutive effects of public policies, and layering (Pierson 2004). Thus, those expecting, hoping for, or lamenting a period of state dominance over markets as a result of the "new interventionism" in the United States under President Obama fail to grasp the realpolitik of administrative reform historically in this nation.

The article begins by offering a synopsis of a theoretical framework for understanding the dynamics of American political development in response to crises—real, perceived, or imagined—presented in greater detail in my previous publications (Durant 2009 and 2011). I then show the empirical foundations of this framework by giving an overview of the reactions of federal policymakers historically to these crises (for a more in-depth treatment, see Durant 2009). Reviewed next is how similar trends of state activism, tempered by implementation structures reserving corporate access, influence, and power, are already occurring in the nation's response to the 2007-2008 financial crisis.

Crises, Path Dependency, and American Exceptionalism

Much talk has arisen of late—some positive and some critical—about American exceptionalism. These values include faith in markets, individualism, and minimal state preferences (Lipset 1996). As used in this article, the term should not be interpreted as implying the moral superiority ofthe United States, but rather the uniqueness or *distinctiveness* of the American Madisonian system of diffused (rather than concentrated) responsibility occasioned by separation of powers, checks and balances, and federalism (Kingdon 1999). Thus, no matter how aggressive or sizeable the initiative taken legislatively to advance state over market interests in the face of crises (domestic, international, or intermestic), the compromises necessary to pass legislation in such a hyperfragmented system ensure implementation structures that allow market interests access and influence during rulemaking and enforcement.

Indeed, the trajectory or arch of reform in response to crises exhibits scant variation acrossAmerican history. First come calls by reformers that existing administrative structures are ill-suited for dealing with emerging trends, contemporary problems, or crises. In the case of crises, calls for state intervention are typically grounded in three sets of rhetorical issue-framing: the need for new structures are needed to produce administrative synergies that were previously nonexistent; the need to respond quickly and effectively to perceived imminent danger (i.e., if we do not act, the consequences will be profound); and the need to be on the "right side of history," given fundamental changes in society. These arguments are quickly countered by opponents and agnostics as likely to have perverse effects, and they would be futile given the complexity of the issue, or jeopardize other

things we value (e.g., state-led healthcare reform jeopardizing freedom of healthcare choices by individuals).

For their part, affected market interests (large and small business alike) raise rhetorical claims such as these, but only up to the point at which they realize that something is going to happen. Then they try to shape implementation structures that ensure them additional access and influence as agencies exercise their discretion in carrying out the law. Alternatively, they may lead the call for reform before others do when events and the inability of existing structures to advance their interests occur. In either case, the legislative process requires a consolidation of perspectives that results in implementation structures that are "halfway, halting, and patchworked" (Skowronek 1982) and that do not replace the status quo. Rather, they get layered within existing and crowded policy spaces and their accompanying implementation structures, thus ensuring difficulty of implementation and conflict among disparate agencies (among any new ones created and legacy agencies that remain in existence).

These structures also produce biases within the system, privileging some policies and interest groups and diminishing or marginalizing others over subsequent decades. In other words, it is not just that politics creates policy, but policy creates politics (Schattschneider 1975). More formally, policies have "constitutive" effects. They do so in several ways: (1) they generate new interest groups that participate in subsequent legislative reauthorization and appropriation decisions in Congress; (2) they either legitimize or marginalize different groups; (3) they either enhance or diminish citizens' sense of political worth and efficacy; and (4) they create implementation structures for carrying out policy that ensure access to some and deny it to others. These, in turn, exhibit "path dependency" and "amplifying" effects. That is, early policy decisions and implementation structures constrain future policy options, participants, and opportunities.

In the case of administrative governance reforms in reaction to national and global crises, these "constitutive" effects create and amplify the ability of corporate interests to ensure that their capacity to influence the discretion that agencies exercise is institutionalized in governance (and agency) structures. Cross-sectoral implementation structures that are developed also create resource dependence, monitoring, and accountability challenges that are less tractable when programs are housed within single public agencies (Frederickson and Stazyk 2010)—thus setting the stage for additional leverage for private actors over policy implementation. This "amplifying effect" also sets the stage for disappointments, if not failures, in reforms, which diminish citizen faith in government and provoke either the next crisis or future calls for administrative reform—at which point, the cycle begins anew. Failure to recognize this realpolitik lies at the heart of mistaking increasing levels of state activism in response to crises for a major aggrandizement of state power and a commensurate loss of power for private enterprise. To see how and why this is the case, we turn next to the empirical foundations of these claims.

Discerning the Forest from the Trees

To paraphrase Kierkegaard's famous observation, "We live our lives forward,

but understand them backwards."[1] Given the anti-state, pro-market, and competitive cultural foundations of the United States, creating a "government out of sight" was the default option of American leaders in the wake of crises in the eighteenth and nineteenth centuries (e.g., the Whiskey Rebellion, the War of 1812, and the depressions of 1819, 1837, and 1873) (Balogh 2009; Howe 1979; Wilentz 2005). During the nation's first century, they did so by, among other things, relying on tariffs collected at American ports rather than on internal taxes; by depending on the decentralized nature of local government with its watchful eye on the judiciary; by subsidizing and giving tax-free deals for settling and developing western lands and infrastructure; and by not recognizing a distinction "between state and civil society or, for that matter, public and private roles for citizens" (Balogh 2009, 24). In this sense, energetic government by the state came packaged through the talents of all sectors (public, private, and nonprofit) rather than through reforms dominated by the national government. Wrought also in the nineteenth century was a linkage between markets and citizenship that helped further put the United States on a networked, cross-sectoral governance path. Indeed, during the first three decades of the nineteenth century, the notions of democracy and capitalism themselves became linked in Americans' minds to create an enduring "myth of national identity" (Howe 1979). Even the Civil War did not leave as its legacy a "permanent autonomous federal bureaucracy" (Skocpol 1992, 68) as a result of Southern reactions to Recontruction.

However, a critical juncture of claims that existing administrative structures were not up to the task of contemporary problems combined with enduring cultural predilections to launch the Progressive reform movement of the late nineteenth and early twentieth centuries. Progressives saw existing governance structures as unable to cope with, for example, health and safety problems in urban areas where immigrants congregated. These structures were also incapable of ending the corruption of political machines dependent on immigrants for electoral support. Likewise mismatched were state regulatory authorities with a burgeoning industrial system that crossed state lines, along with international economic, industrial, and foreign policy threats that reformers said only professionalization and executive-centered government could allay. Madisonian checks-and-balances were too cumbersome, slow, and amateurish to cope with these challenges and would put the United States at a commercial and military disadvantage.

The 1880s witnessed the most dynamic period of economic growth in the United States since the Civil War, but it was largely concentrated in rural,

1 Kierkegaard appears to have expressed this concept over time in a variety of forms. *See, e.g.*, Alexander Dru, ed. and trans., *The Journals of Søren Kierkegaard* (London: Oxford University Press, 1951), 127 ("It is perfectly true, as philosophers say, that life must be understood backwards. But they forget the other proposition, that is must be lived forwards." [1843, Entry 465]) ; Howard V. Hong and Edna H. Hong, eds. and trans., *Søren Kierkegaard's Journals and Papers* (Bloomington, Ind.: Indiana University Press, 1967), I:449 ("Life can be interpreted only after it has been experienced . . . ." (1838, Entry 1025], 450 ("Philosophy is perfectly right in saying that life must be understood backwards. But then one forgets the other clause—that it must be lived forwards." [1843, Entry 1030]), 469 ("The sad thing about us human beings is really that in almost everything in our lives hindsight is best; that is, after we have done something, often badly, then we know how we should have done it." (1849, Entry 1074]; *see also* William James, *Essays on Radical Empiricism* (London: Longmans Green, 1922), 238 (" . . . a saying of Kierkegaard's to the effect that we live forwards, but we understand backwards.") (quoted in Hong & Hong, eds., *Journals and Papers*, I:537 n.553).

sole proprietary business ventures. That era observed the beginnings of trade associations of businesses, but nowhere near the proliferation that came during the Progressive Era (e.g., the Indianapolis Monetary Commission, the National Association of Manufacturers, the U.S. Chamber of Commerce, and the National Foreign Trade Council). Indeed, as Lears (2009) argues, commerce and courage were linked in many minds to the manly virtues that an "overciviliz[ed]" Gilded Age and Victorian society had been jeopardizing since the end of the Civil War.

However, with the Panic of 1893, the nation went through a financial depression unparalleled since the 1819 and 1837 depressions and not experienced again until the Great Depression of the mid-twentieth century. Lasting four years, the devastation of the 1893 depression was immense and is typically seen as the wellspring of the Populist movement that preceded the regulatory agenda-setting efforts of the Progressive reform movement. Also contributing to these movements was the rampant labor unrest spawned by a reconstruction of American capitalism that followed in the wake of the Panic of 1893. The result was an unprecedented corporate reorganization of American industry, especially in the form of a noticeable increase in industrial mergers. While approximately 20 industrial corporate consolidations had resulted from merger activity since the 1880s and 1890s, nearly 300 industrial corporate consolidations had occurred by 1904 (Sklar 1988, 45-46).

Underlying these trends was a growing recognition by business—large and small alike—that classic economic liberalism was now obsolete. Its focus on rugged individualism, localism, and competition that was so central to the nation's economic development in the early to mid-1800s, an era of largely rural and proprietary capitalism (small artisan and business competition), was no longer working to the nation's advantage. Competition, corporate capitalists averred, was "socially inefficient," wasteful of resources, and unnecessary with the revolution of mechanization in America. Cooperation, coordination, and the application of scientific principles to industry management was their preferred solution. Meanwhile, small business wanted protection from large producers. Thus, for different reasons, both wanted a "government of laws, not of persons." They saw this as conducive to a favorable investment environment, techno-economic progress, interclass cooperation, and social stability—in short, "clean government" and "social efficiency" (Sklar 1988, 30-31).

As this "revolution of cooperation" or "cooperative commonwealth" took root (Weinstein 1968), labor unrest occurred on a scale seldom seen before or since. During the 1880s and 1890s, the "efficiency" focus of corporations sparked a working-class "national strike day" in 1886 to demand an eight-hour workday. Wage cuts also contributed to repeated and sometimes violent strikes involving hundreds of thousands of unskilled industrial workers, with unrest spiking in 1877, 1886, 1892, and 1893 (Trachtenberg 2007, 88-89). Concomitantly, traditional definitions of responsibility that were associated with individuals were redefined by corporate interests in the National Civic Federation (see more below) as "socially responsible" behavior—defined partially as not fomenting economic class divisions.

In the midst of this unrest, both the large and small business communities had common interests in advancing some form of regulatory agenda. Small

business was disenchanted with laissez-faire competition and wanted it curtailed by regulation. For larger businesses, politics was too unpredictable, inefficient (because of patronage), and responsive to untutored public opinion to afford a stable, predictable, and safe business environment. At the same time, many middle-class social Progressive reformers, small business leaders, and corporate leaders agreed that regulatory reform was a "counterpoise to the threat of working class revolution" occasioned by repeated labor turbulence (Weinstein 1968, xi (quoting Sidney Kaplan, "Social Engineers as Saviours: Effects of World War I on Some American Liberals," *Journal of the History of Ideas* 17 (June 1956): 354-55)). This occurred as a business movement successfully ensured a twentieth century-oriented "state intervention to supervise corporate activity, rather than a [nineteenth-century] movement for the removal of state control over private enterprise" (Ibid., xiii).

Consequently, in a pattern that reproduces and amplifies itself throughout American history in the wake of crises, the halfway, halting, patchworked, and layered implementation structures that were created to accommodate the disparate aims of this Baptist-bootlegger coalition frequently came about in ways advocated by the regulated community. These structures ultimately allowed historically privileged interests to ensure longer-term power, access, and influence to shape subsequent regulatory agendas, enforcement aggressiveness, and enforcement targets. For example, industry associations actually lobbied in favor of what were typically seen as Progressive reforms creating independent regulatory commissions (IRCs)—such as the Interstate Commerce Commission and the Chemistry Bureau (charged with implementing the Pure Food and Drug Act of 1906)—for the purpose of reining in the "malefactors of wealth." Kolko chronicles, for example, how "any measure of importance in the Progressive Era was not merely endorsed by key representatives of businesses involved; rather such bills were first proposed by them" (Kolko 1963, 283; see also Hofstadter 1989; Wiebe 1967). As Weinstein puts it, traditional liberalism favoring free markets morphed into "corporate liberalism" (Weinstein 1968, xi-xiv, 3).

Corporate liberalism, however, also appealed to the enfranchised, well-educated citizens, the rising professional class, and social reform leaders. It did so by "granting them status and influence as spokesmen for their constituents on the condition only that they defend the framework" offered, which favored business-oriented agendas and implementation structures (Weinstein 1968, xiv; see also Sklar 1988; Trachtenberg 2007; Wiebe 1967). For example, the National Civic Federation, a coalition of largely corporate businessmen founded in 1900,established the principle of tripartite—business, labor, and public—representation in politics (Sklar 1988; Weinstein 1968, xv).

Historians often portray the end of the Progressive Era in 1920 as a function of the citizenry's tiring of the moral proselytizing of the social reformers involved (McGerr 2003). However, a less-recognized faction of the Progressive reform movement continued in the form of the "associationalism" movement. Nevertheless, associationalism, too, produced administrative approaches to existing and projected crises that resulted in the advantaging of market over state interests. Dubbed the "associative state" (Hawley 1974), Secretary of Commerce Herbert Hoover most

notably championed this alternative model. Associationalism was premised on government-stimulated voluntary cooperation to address public problems, with direct federal government intervention being a last resort if private and civic volunteerism failed (Kennedy 2005). For the proponents of associationalism, properly educated and self-governing businesses, state and local governments, voluntary organizations, and professional associations informed by data analyses provided by federal agencies would willingly tackle any problems or crises identified (Hofstadter 1989).

Nor did the associationalist elements of the Progressive Era fade away as popularly portrayed. Rather, its public-private partnership components were marbled into the administrative state during the New Deal and in subsequent decades (e.g., Burner 1978; Clemens 2006; Hawley 1974). Hart, for example, argues that due to the "shadow of the welfare state and the warfare state, the associative state has survived" and "advocacy [has occurred] on its behalf in virtually every Administration" since FDR's (Hart 1994, 30). The justifications offered for each iteration of associationalism from the 1960s on have been highly reminiscent of previous administrative reform movements in the wake of alleged crises, stressing as they have gaps in governance capacity to deal with societal problems. These include President Kennedy's expansion of associationalist public-private partnerships to the textile and construction industries in the face of global competition, the Carter Administration's plans for "generic technology" and a government-industry-labor economic revitalization board to compete with international competitors, and the Atari-Democrat fascination with the Japanese Ministry of International Trade and Industry in the face of cooperative mercantilism abroad in the 1980s. Associationalist "heirs" in the Reagan and Clinton Administrations persistently tried "to replace the adversarial relationships [among businesses] that they thought placed American firms at a disadvantage in the world economy" (Hart 1994, 27).

The 1980s and 1990s also witnessed continuing movement toward public-private hybrid organizations pushed by associationalists during the 1920s, including the now discredited Fannie Mae and Freddie Mac (Koppell 2003). These were all efforts to shape a "new kind of governmental agency, one that sought not to regulate. . . but to make private groups more statesmanlike and hence better able to cope with modern conditions and problems" (William Hard, as quoted in Hawley 1981, 3). However, problems quickly arose, especially in the consumer protection policy arena and its fate in Congress. These problems were presaged in patterns over the past four decades in interest-group testimony before congressional committees. Not only is this arguably a surrogate for interest-group policy access, but it also highlights the policy predilections of members of Congress over time because they control witness lists. By whatever measure, prior research on interest groups suggests that the "ordinary task of advocating for protecting consumers is left to a small and shrinking community of organizations and a very few individual voices" (Bykerk and Maney 2010-2011, 653).

Political science researchers have discerned several trends from 1987 to 2006 (from the 100th Congress to the 109th Congress) that illustrate the constitutive effects of public policy and that set the stage for continued layering without

learning (Bykerk and Maney 2010-2011). First, consumer protection remained a consistent issue in Congress during the entire time span. In fact, hearings in the House consistently expanded over this time period, from a low of seven after the Republican Party takeover of the House in 1994 to a high of nineteen in the 106th Congress.

Second, a decided and persistent gap favoring business over consumer group participation in hearings developed during the Republican years, with the 104th Congress sustaining the most significant gap. In the 104th Congress, business interests constituted seventy-four percent of all groups giving testimony. In the years since, consumer group testimony plummeted even further, never exceeding twelve percent of the total number of appearances by all groups and falling to a low of four percent in 2004. Moreover, while there were ninety-one consumer protection groups identified in a study of interest-group testimony between 1970 and 1986, the 1987 to 2006 study identified only thirty-three consumer groups testifying (Bykerk and Maney 1991-1992). What is more, with the dominance of peak associations such as the Chamber of Commerce and the National Association of Manufacturers, it is clear that business perceived a sense of common interest in advancing its agenda.

Third, business participation never fell below fifty-five percent of the total number of interest groups testifying on consumer protection between the 100th Congress and 108th Congress. This means that even if all other participants in those hearings are combined (e.g., labor unions, consumer groups, state attorneys general, think tanks, and independent experts such as university professors), they are still competing for influence with businesses and associations of considerable political clout (e.g., the American Council of Life Insurers, the American Bankers Association, the Independent Bankers Association, the Chamber of Commerce, the National Retail Federation, and the National Association of Manufacturers). Other research indicates that consumer groups did not fare any better with individual legislators during this era at the federal level (Bykerk 2008) and that campaign contributions to state election candidates were sometimes advancing conservative interests (Witko and Newmark 2005).

Finally, scholarship indicates that federal and state agency rulemaking during the 1990s did not escape inordinate business influence in relationship to consumer interests. Especially profound has been the role of business and professional associations at each stage of the rulemaking process. Premised on prior research, there is little doubt that interest-group involvement in the development of rules is deep and influential. Research by Golden (1998), Kamieniecki (2006), and Yackee and Yackee (2006) has found a bias toward business participation in the process. For example, in a study examining over thirty regulations from four different agencies, Yackee and Yackee (2006) found that over fifty-seven percent of public comments came from business interests. Likewise, survey data from studies conducted by Furlong and Kerwin (2005) suggest that businesses, and the trade associations that represent businesses and professions, were (and remain) involved in rulemaking more often than are other groups. While others have found that influence varies across stages of the regulatory process, it nonetheless raises the

specter of a concentrated strategy to advance industry interests in this venue (Kraft and Kamieniecki 2007).

These trends in Congress were the predicate for the self-regulatory preoccupation by its members that led to the 2007-2008 financial crisis and its aftermath. For starters, the 1999 Gramm-Leech-Bliley Act (henceforth, the GLB Act) repealed the Depression-era Glass-Steagall Act (more formally, the Banking Act of 1933). This made it possible for commercial (and not just investment) banks, as well as unregulated hedge funds and over-the-counter (OTC) traders, to participate in highly profitable, exorbitantly opaque, and risky speculative schemes involving the securitization of prime and subprime mortgages. Touted as a system that would diffuse risk among a large number of securities holders and thus create a more "shock-proof" global financial system, securitization had precisely the opposite effect (Helleiner 2011). Within ten years, most of the risks were held by a relatively small number of institutions that were deemed "too big to fail." Indeed, the GLB Act had intensified competitive pressures among these institutions, prompting many of the largest to embrace mortgage securitization as they watched competitors' profits rise.

Dovetailing with this, Congress placed pressures on Freddie Mac and Fannie Mae to buy more subprime mortgages, pressures that interacted in deleterious ways with other regulatory actions reflecting the influence of the financial industry in regulatory structures. Still, private lenders remained the prime underwriters of subprime loans. The financial industry had been creating new financial instruments such as mortgage-backed securities (MBS) and collateralized debt obligations (CDS). These allowed mortgage lenders to worry less about the risks they were assuming in awarding mortgage loans because they sold them to others. They also masked the risks that underwriters were accepting. This occurred as a spate of private firms entered the market, bundling prime and risky subprime mortgages into MBS packages. In turn, these were often subdivided and sold as CDS packages. Concomitantly, the Securities and Exchange Commission (SEC) further reduced risks to investment banks for getting into mortgage securitization schemes by lifting a 12:1 leverage ratio.

While all this was happening, regulators created "market-friendly" accountability structures that "trusted private actors to self-regulate" (Porter 2005). In 2000, Congress ratified a veritable laissez-faire regulatory environment for OTC derivatives trading. In addition, regulators adopted and implemented standards developed by the industry itself when it came to derivatives, accounting, and hedge-fund management. In the process, elected officials and regulators privatized the tasks of valuing market risks and assets to credit-rating firms whose compensation schemes depended on undervaluing risk. Eventually, the opaqueness of the financial instruments created nearly froze lending in the United States and abroad. Not knowing which lenders held risky subprime debt in their portfolios destroyed confidence in the entire system.

Also illustrative of self-regulatory emphases of the era were developments in international financial markets. Paradoxically, these efforts started with the creation of the Basil Accords in 1988, which were subsequently updated between

1998 and 2004. The aim of the accords—collectively known as Basil II—in the wake of financial crises in Mexico (1994) and East Asia (1997-1998) was to create common capital standards to reduce threats to global financial markets. The rub was that these standards did not apply to the primary protagonists in securitization (viz., investment banks, hedge funds, and insurance companies). Among other self-regulatory devices, Basil II also allowed banks to use their own internally generated value-at-risk models to determine their risk reserve, and they created only voluntary standards for credit-rating organizations (Helleiner 2011).

Consonant, as well, the 1990s saw a movement by elected officials worldwide to shift financial regulatory power to their treasury departments. As Roberts (2010) explains, a two-track "logic of discipline" reminiscent of the politics-administration dichotomy again took root as nations shifted financial authority to finance ministers and erected legal restraints on borrowing and spending. This time it was not hordes of immigrants and political machines that required shifting policy decisions to experts who would apply their technocratic skills to public problems in rational ways unavailable to the masses. As one anonymous official in the UK Treasury put it in 1987, "Parliament is incapable of exercising its financial responsibilities. We must do it for them" (Roberts 2010, 47). Meanwhile, in the United States, former Federal Reserve Chair Alan Blinder argued that the country would be better off if more policy decisions were taken out of the "political thicket" and placed with "unelected technocrats" (Roberts 2010, 47).

Given the logic of isomorphism, nations faced with spiraling federal spending, mounting debt, and what some called "fiscal drift" began shifting increasing amounts of fiscal power to their finance ministries. In the short run, the strategy appeared effective as deficits began to decline, which further reinforced lock-in and isomorphism of these approaches. In the long run, however, their flawed logic was that technocrats in finance ministries could constrain the out-of-control spending habits of legislatures and other agencies. However, the power of the treasury was only good as long as a crisis existed (Roberts 2010). Then, in 2007, the finance ministries became the biggest spenders of all in trying to pull the world back from financial collapse. Subsequently, as Roberts points out and the Tea Party and the Occupy Wall Street movements in the United States attest, this approach seemed so antithetical to democratic accountability once things went very badly that populist revolts from the Right and Left subsequently ensued.

"Layering without Learning" in the Post-Financial Crisis Era

Market deregulatory efforts seem to have come disastrously home to roost in the 2007-2008 global financial meltdown. The same aggressive state initiatives that prompt calls of a new Leviathan, if not socialism, in America are accompanied by implementation structures which assure that market interests continue to have access and influence in the distribution of resources, regulatory power, and legal enforcement. Consider, for example, the Bush Administration's Emergency Economic Stabilization Act of 2008. The act created within the Treasury Department the Toxic Asset Recovery Program, which was to purchase troubled

assets and equity from financial institutions to deal with the subprime mortgage crisis. The program subsequently contracted out the bulk of its asset-management functions to between five and ten large private sector asset-management firms. A "bare-bones internal staff of about two dozen people" (rather than career bureaucrats) were hired to oversee how these firms spent the first $250 billion of the recovery plan (Landler and Andrews 2008).

Likewise, much of the economic stimulus package (the American Recovery and Reinvestment Act of 2009) that passed in the early weeks of the Obama Administration involved money for pass-through funding to states and localities and private sector subsidies, albeit with line items for some capacity rebuilding in the federal government. Relatedly, President Obama's Treasury Secretary, Timothy Geithner, relied heavily on public-private partnerships to buy up toxic bank assets. Certainly, grants to subnational governments and contracts to private and nonprofit organizations have strings attached, but reams of intergovernmental and implementation research demonstrate that bargaining and mutual adaptation characterize these efforts (Hill and Hupe 2009).

Next, the Financial Reform Act of 2010 (henceforth, FINREG) was designed to deal with flaws in the existing financial regulatory structure. FINREG is too massive to consider in all its aspects in this article. Clearly, however, the new law represents an improvement over the status quo in a variety of ways. These include the early resolution authority that FINREG provides as an antidote to "too big to fail" institutions, and its closing of gap-ridden existing regulation that caused lapses in oversight which nearly brought the world's global economy to its knees. Still, FINREG also created a significantly increased and unprecedented role for the Treasury Department within government (Cho 2010, A1). For example, Treasury acquired sweeping new power to shape bank regulation, oversee financial markets, and review the regulations issued by a new Consumer Financial Protection Bureau (CFPB). Importantly, however, it did not acquire the ability to regulate Freddie Mac and Fannie Mae, organizations whose lending practices led to the mortgage crisis. Nor does the CFPB have direct regulatory authority for all banks (indeed, it regulates approximately ten percent of banks).

The CFPB inherited regulatory authorities from existing (or "legacy") financial regulatory agencies that failed to perform adequately during the financial crises, but which did not go away. These authorities included not only those under the Truth in Lending Act, but also those under six other financial regulatory statutes. Moreover, the CFPB was located within the Federal Reserve and overseen by a Financial Stability Oversight Council (FSOC) designed appropriately to take a systemic view of the financial markets that was lacking in the past. The council consists of representatives from all the other seven financial regulators. The chairman of the council is, again, the Treasury Secretary.

Each of these design features ensures the continuing access and influence of the legacy financial regulators that failed to perform adequately. To be sure, the CFPB was given an independent budget when Congress guaranteed that it would receive increasing percentages of the budget allocated to the Federal Reserve (rising from ten to thirteen percent). It can also ask Congress for additional funding. However,

the FSOC has the ability to reverse, stall, or drop any regulation issued. Moreover, the location of the CFPB in the Federal Reserve also suggests that consumer protection will face an uphill battle in a culture focused historically on stability and safety of the system rather than on consumer protection. Furthermore, creating the CFPB with a dual mission of protecting consumers while assuring the safety and viability of the system as a whole will make the CFPB susceptible to policy swings in the political appointees who leadthe bureau. Meanwhile, the fragmented and layered structure that requires the CFPB to coordinate (e.g., gain information from legacy regulators) and work with state regulators to advance consumer protection affords a variety of other veto points for the banking industry.

To be sure, advances have been made and others are still evolving. However, they are advances that have, to date, to be wrenched from financial interests largely on their terms. Overall, as one of the most informed observers of the financial scene, Michael Lewis, argues, American financial stress tests are nowhere near as stringent as they should be. He is joined by European critics who have argued that the first two rounds of stress tests conducted by the European Banking Authority in the European Union (EU) were similarly deficient (Taiti 2011). For example, two of the largest Irish banks that survived stress tests failed within months of evaluation, leading to that country's financial bailout by the IMF and the EU. Meanwhile, credit default swaps still remain largely unregulated, rating agencies that were complicit in the mortgage crisis remain largely untouched, and the sinewy bounds between Goldman Sachs Chase and the Treasury Department remain unabated.

Conclusion

This article has examined how the American political system has historically reacted to crises—financial or otherwise—since its founding. It has offered a framework for understanding repeated patterns of privileging business interests, as well how and why American exceptionalist values help privilege those interests because of such features as anti-statism, individual liberty, classic economic liberalism, and the folklore of competitive capitalism as opposed to the reality of administered capitalism since the early 1900s. In the process, it has shown the persistent allure and path dependency of nongovernmental implementation structures; of implementation structures designed to ensure future access and influence for those interests (regulated or otherwise); and of halfway, halting, and patchworked regulatory structures layered atop each other in America's Madisonian system to compromise effectiveness. It has done all of this, it is hoped, in ways that call readers' attention to the fallacy of assuming that today's aggressive federal legislation should be equated with a move toward state dominance of markets or to the rise of socialism inthe United States.

As mentioned earlier, however, one should not assume that a massive conspiracy controlled by business is underway. Equally misguided is assuming that no progress is being made toward addressing the social ills that are negative externalities of markets. Progress *has* been made historically and in contemporary times. However, path dependency means that the alternatives for addressing such market challenges are constrained by, among other things, early choices made; the

sequencing of events and the conjuncture of others; and changing interests and conflicts within and across authority structures in the United States. Only when shifts in secular trends arise, exogenous shocks occur to the system, supporting coalitions disintegrate, or the interests of those advantaged by existing structures shift do new reform movements arise and then go through the cycle depicted in this article. Moreover, in these instances, even massive shocks to the system are insufficient for fundamental change. Policy spaces are too crowded, institutional interrelationships in policy spaces too intertwined, and the dynamics of the legislative process so demanding of concessions to the status quo that nonincremental change disadvantaging business to the advantage of state power is unlikely absent a total breakdown of system legitimacy.

Thus, those expecting, hoping for, or lamenting that the "new interventionism" in the United States means a period of state dominance over markets ignore not only the power of the past to shape the present, but also the powerful, protean, and durable sway of American exceptionalist values. These affect the nation's predilection for networked, cross-sectoral solutions to public problems. Put most simply, market and nonprofit solutions to public problems are an easier "sell" to the American polity, consonant as they are with American exceptionalist values. Such solutions also are marbled throughout a Madisonian system that diffuses responsibility among actors, affords multiple access points to policy making by disparate actors, and necessarily layers new implementation structures atop old ones. The networked state is here to stay, as it has been throughout American history in times of normalcy and in the wake of crises.

References

Balogh, Brian. 2009. *A Government Out of Sight: The Mystery of National Authority in Nineteenth-Century America*. New York: Cambridge University Press.

Burner, David. 1978. *Herbert Hoover: A Public Life*. New York: Alfred A. Knopf.

Bykerk, Loree. 2008. Lobbying Unorthodox Lawmaking. *International Social Science Review* 83(Fall-Winter):120-121.

Bykerk, Loree, and Ardith Maney. 1991-1992. Where Have All the Consumers Gone? *Political Science Quarterly* 106(4):677-693.

———. 2010-2011. Consumer Protection Policy Issues on the Congressional Agenda. *Political Science Quarterly* 125(4):639-655.

Cho, David. 2010. Timothy Geithner's Realm Grows with the Passage of Financial Regulatory Reform. *Washington Post*, July 17: A1.

Clemens, Elisabeth S. 2006. Lineages of the Rube Goldberg State: Building and Blurring Public Programs, 1900-1940. In *Rethinking Political Institutions: The Art of the State*, edited by Ian Shapiro, Stephen Skowronek, and Daniel Galvin, 187-215. New York: New York University Press.

Durant, Robert F. 2009. "Theory Building, Administrative Reform Movements, and the Perdurability of Herbert Hoover," *American Review of Public Administration* 39(4): 327-51.

———. 2011. "Global Crises, American Public Administration, and the 'New Interventionism" Revisited," *Administration & Society* 43(3): 267-300.

Frederickson, George H., and Edmund C. Stazyk. 2010. Myths, Markets, and the "Visible Hand" of American Bureaucracy. In *The Oxford Handbook of American Bureaucracy*, edited by Robert F. Durant, 349-71. Oxford: Oxford University Press.

Furlong, Scott R., and Cornelius M. Kerwin. 2005. Interest Group Participation in Rule Making: A Decade of Change. *Journal of Public Administration Research and Theory* 15(3): 353-70.

Golden, Marissa M. 1998. Interest Groups in the Rule-Making Process: Who Participates? Whose Voices Get Heard? *Journal of Public Administration Research and Theory* 8(2): 245-70.

Hart, David M. 1994. Herbert Hoover's Last Laugh: The Enduring Significance of the "Associative State" in the United States. Paper presented at the Society for the History of Technology meeting, Lowell, MA, and the Northeastern Political Science Association meeting, Providence, RI. [Published 1998. *Journal of Policy History* 10(3): 419-44].

Hawley, Ellis W. 1974. Herbert Hoover, the Commerce Secretariat, and the Vision of an "Associative State," 1921-1928. *Journal of American History* 61(1): 116-40.

———, ed. 1981. *Herbert Hoover as Secretary of Commerce: Studies in New Era Thought and Practice*. Iowa City: University of Iowa Press.

Helleiner, Eric. 2011. Understanding the 2007-2008 Global Financial Crisis: Lessons for Scholars of International Political Economy. *Annual Review of Political Science* 14: 67-87.

Hill, Michael, and Peter Hupe. 2009. *Implementing Public Policy*. 2nd ed. Thousand Oaks, CA: SAGE Publications.

Hofstadter, Richard. 1989. *The American Political Tradition and the Men Who Made It*. New York: Vintage Books.

Howe, Daniel W. 1979. *The Political Culture of the American Whigs*. Chicago, IL: University of Chicago Press.

Kamieniecki, Sheldon. 2006. *Corporate America and Environmental Policy: How Often Does Business Get Its Way?* Palo Alto, CA: Stanford University Press.

Kennedy, David M. 2005. *Freedom from Fear: The American People in Depression and War, 1929-1945*. Oxford: Oxford University Press.

Kingdon, John W. 1999. *America the Unusual*. Australia: Thomson/Wordsworth.

Kolko, Gabriel. 1963. *The Triumph of Conservatism: A Reinterpretation of American History, 1900-1916*. New York: Free Press.

Koppell, Jonathan G. S. 2003. *The Politics of Quasi-Government: Hybrid Organizations and the Dynamics of Bureaucratic Control*. Cambridge: Cambridge University Press.

Kraft, Michael, and Sheldon Kamieniecki. 2007. *Business and Environmental Policy: Corporate Interests in the American Political System*. Cambridge, MA: MIT Press.

Landler, Mark, and Edmund L. Andrews. 2008. For Treasury Dept., Now Comes the Hard Part of Bailout. *New York Times*, October 3: A1, *available at* www.nytimes.com. Accessed October 3, 2008.

Lears, Jackson. 2009. *Rebirth of a Nation: The Making of Modern America, 1877-1920*. New York: Harper Perennial.

Lipset, Seymour M. 1996. *American Exceptionalism: A Double-Edged Sword*. New York: W. W. Norton & Co.

McGerr, Michael. 2003. *A Fierce Discontent: The Rise and Fall of the Progressive Movement in America, 1870-1920*. New York: Free Press.

Meacham, Jon, and Evan Thomas. 2009. We Are All Socialists Now. *Newsweek*, February 16: 23-27.

Pierson, Paul. 2004. *Politics in Times: History, Institutions, and Social Analysis*. Princeton, NJ: Princeton University Press.

Porter, Tony. 2005. *Globalization and Finance*. Cambridge, UK: Polity.

Roberts, Alasdair. 2010. *The Logic of Discipline: Global Capitalism and the Architecture of Government*. Oxford: Oxford University Press.

Schattschneider, E. E. 1975. *The Semi-Sovereign People: A Realist's View of Democracy in America*. New York: Wadsworth Publishing.

Sklar, Martin J. 1988. *The Corporate Reconstruction of American Capitalism, 1890-1916*. Cambridge: Cambridge University Press.

Skocpol, Theda. 1992. *Protecting Soldiers and Mothers: The Political Origins of Social Policy in the United States*. Cambridge, MA: Belknap Press.

Skowronek, Stephen. 1982. *Building a New American State: The Expansion of National Administrative Capacities, 1877-1920*. Cambridge: Cambridge University Press.

Taiti, Nikki. 2011. New EU Bank Stress Test Rules Out Soon, *available at* FT.com. Accessed February 27, 2011.

Trachtenberg, Alan. 2007. *The Incorporation of America: Culture and Society in the Gilded Age*. New York: Hill and Wang.

Weinstein, James. 1968. *The Corporate Ideal in the Liberal State*. Boston, MA: Beacon Press.

Wiebe, Robert H. 1967. *The Search for Order, 1877-1920*. New York: Hill and Wang.

Wilentz, Sean. 2005. *The Rise of American Democracy: Jefferson to Lincoln*. New York: W. W. Norton & Co.

Witko, Christopher, and Adam J. Newmark. 2005. Business Mobilization and Public Policy in the U.S. States. *Social Science Quarterly* 86(2): 356-76.

Yackee, Jason W., and Susan W. Yackee. 2006. A Bias Towards Business? Assessing Interest Group Influence on the U.S. Bureaucracy. *Journal of Politics* 68(1): 128-39.

Prospects for the Future Administration of Elections

Robert M. Stein, Rice University

Introduction

Where, when, and the how American voters cast their ballots has become highly variable. We do not all vote on the same day, we vote at a host of different places, and we do so on a cornucopia of types of voting equipment ranging from paper ballots to sophisticated electronic voting machines. In this context, trying to predict what the future conduct of elections will look like in the next half century is daunting and requires a well-planned strategy. In this essay, I offer both a way of understanding and predicting how elections might be conducted in the future. This essay is not intended to be a prescriptionbut, rather, is intended to provide a conjecture about the future of the administration of elections in democratic systems. Further, this essay is not about how elections are contested, who wins, why, and the consequences of electoral outcomes for governance. Instead, I am concerned with how we conduct our elections, with the impact of our election procedures on who votes and for what offices, and with whether the voters' performances and evaluations of their voting experiences vary with the way elections are conducted.

In the next section, I identify the goals associated with the administration of elections in democratic systems of government. Each goal is briefly described and assessed in terms of current practices, (mostly, but not exclusively in the United States) constraints, and impediments to achieving these goals. In section three, I identify what practices pursuant to these core goals of election administration we might expect in the future. I assume that predicting the administrative conduct of elections in the future requires that we understand the constraints and incentives for achieving the core goals of election administration and the trade-offs between each goal. Within this framework I identify some of the political, economic, social, and technological trends that might impact election practices and their consequences for governance.

The Goals of Election Administration

In democratic systems of governance, we primarily hold elections in order to choose persons to govern on our behalf.[1] Contesting candidates who receive the most votes are selected to govern. Of course, the consequences of elections for governance are myriad and beyond the scope of this paper. I identify five core functions and goals for the administration of elections. These include:

- Assuring that all persons who are eligible to vote can register to vote.
- Assuring that all persons who want to vote are able to vote.
- Assuring that all voter choices are accurately recorded.

1 Other elections direct and enact policy in the form of referenda and initiatives.

- Assuring that all votes are accurately counted.
- Assuring that all voters have a positive and confident voting experience.

Voter Registration

In the United States, the determination of who is eligible to vote is a shared responsibility between the states and the federal government. The Constitution, Supreme Court decisions, and Congress determine who may be eligible to vote in federal elections. The states and their localities have authority to regulate this right by requiring voters to be registered to vote before balloting. Requirements for voter registration directly and indirectly limit who is eligible to vote and thus who have access to the ballot. Several states place limitations on who can register to vote, as in the case of excluding convicted felons from participating in the voting process. The most onerous constraint voter registration requirements place on voting is the requirement to be registered before election day in order to vote. This requirement varies across states but essentially requires that a citizen reside for a minimum period of time (e.g., 30 days before an election) at the same address in order to be eligible to vote.

Voter registration requirements are a relatively recent practice that began in the late nineteenth century (*see* Bensel 2004) to avoid voter fraud. The goal of ballot access for eligible voters is weighed against the competing goal of avoiding election fraud. Wolfinger and Rosenstone (1980) and others (Squire et al. 1987; Rhine 1996) have shown that voter registration requirements suppress voter participation among otherwise eligible voters and particularly among the most residentially mobile portions of the population (the poor and young). Squire et al. (1987) conservatively estimate that turnout would increase by nine percent with election day voter registration, which is currently practiced in ten states. Voter participation might be expected to increase significantly if voter registration requirements were eliminated or if registration were allowed on election day, or whenever one voted.

One such improvement in the voter registration system is the development of a single, uniform, and interactive computerized statewide voter registration list, maintained and operated by the state or federal government. Locally maintained voter registration lists are often inaccurate, out of date, and thus an obstacle to voters who have recently moved to the area. The *Help American Vote Act of 2002* (HAVA) required that states have a statewide voter registration database in place by 2004. To date, less than half of the states have complied with this requirement of HAVA.

Enhancing Voter Turnout

In the 2000 *Current Population Survey*, the modal response to the question, "Why did you not vote in the 2000 Presidential election?" was, "Too busy." Enhancing the likelihood that registered voters will go to the polls may require extensive campaign activities, a fundamental change in attitude toward government, or the removal of obstacles that prevent an individual from voting (e.g., access to transportation).

One way to think about voting is that it is inconvenient or at least competing

with other interests, including work, shopping, school, and recreation. People who might want to vote will not because they choose to engage in some other activity on election day. Many states have endeavored to make voting more convenient by authorizing relaxed requirements for absentee mail-in ballots, permanent absentee mail-in voting, all-mail elections, in-person early voting, and, in the case of overseas military personnel, internet voting.[2]

The evidence is mixed and weak, however, as to whether convenience voting actually enhances voter participation, particularly among those who historically do not vote. Since its adoption in Texas in 1988, for instance, early voting has not been associated with a significant increase in voter turnout in that state, especially among individuals who historically do not vote, such as the younger generations and racial and ethnic minorities (Stein 1998; Stein and Garcia-Monet 1997; Berinsky et al. 2001; Berinsky 2005). Allowing people to vote up to two weeks before election day at a variety of locations that are more proximate to where people work, shop, and travel maybe more convenient for *voters*; but it is only more convenient to those who already vote on election day. Extensive research in a host of states with in-person early voting over the last twenty years has failed to show that early voting significantly increases turnout (Berinsky 2005). Moreover, the evidence is that in-person early voting actually increases the likelihood that those who already vote on election day will be more likely to cast a ballot when they are afforded more days, times, and places to vote.

Of course, the reason for this unfulfilled promise of early voting is simple; voters do not vote early if they do not know for whom they are going to vote. Exit polls have consistently shown that early voters are strong partisans, ideologues, attentive, and informed citizens. They know about early voting, and most importantly, their vote choices have been made well before election day. Election day voters are significantly less likely to have strong partisan and ideological preferences, are less interested in and informed about politics and political campaigns, and are far *more* likely to make their vote decisions just before election day (Stein 1998).

Like in-person early voting, election day polling centers afford citizens the opportunity to vote at any polling place—not just one location. The centers are located at large venues, including shopping malls, hotels, and even car dealerships. To date, the research has shown that affording election day voters the "'conveniences"' of early voting increases turnout between four percent and seven percent (Stein and Vonnahme 2008; Brady and McNulty 2004; Dyck et al. 2005). Most importantly, the positive effect election day polling centers have on turnout is greater for individuals who are infrequent rather than frequent voters. The lessons learned from early voting can be applied to election day voting with the desired effect—increased voter turnout among those who historically are too busy to vote (Stein and Vonnahme 2010; Gimbel and Schuknecht 2003).

Mail-in ballots can also increase voter turnout. For instance, Oregon, which has the most extensive and widespread experience with mail-in voting, moved to an all-mail ballot in 1998. To date, the evidence suggests that voter turnout has increased by ten percent since the adoption of all-mail voting in Oregon (Richey

2 Several states, including Arizona and Michigan, have conducted partisan primary elections on the internet.

2008). In constrast, Richey's analysis and others (Hamner and Traugott 2004; Karp and Banducci 2000; 2001; Kousser and Mullin 2007) show that states that have adopted in-person early voting or permanent mail-in voting have not experienced a significant increase in voter turnout. The lessons learned from states with all-mail elections and election day polling centers seem to offer the best prospects for enhanced voter participation. While voter turnout is a major concern during elections, accurately recording votes presents another challenge to the voting process.

Accurately Recording Vote Choices (The Usability of Voting Systems)

Persistent problems in voting administration are often the result of errors associated with the casting of ballots. An error occurs when what is recorded does not reflect the intent of the voter. On a paper ballot, this may result from "mismarks" when voters fail to properly mark the box, circle, or bubble on their paper ballots for their chosen candidates. The residual vote is defined as the difference between the number of voters who come to the polls to vote and the number of votes cast per ballot contest. When the number of votes cast for any one race is less than the total number of ballots cast, we observe an undercount. Related to the incidence of undercounts are weak and anemic candidate campaigns that fail to inform and mobilize voters. Voting equipment, long lines, and ineffectual poll workers can operate to deter voters from casting a complete ballot (Bernstein et al. 2001).

In contrast, overcounts happen when voters improperly mark or select more than one ballot choice for a single contest. These ballots are invariably voided because voter intent cannot be readily determined (Knack and Kropf 2003). This problem became famous and problematic during the recount of the 2000 presidential ballots in Florida where an assortment of punch card voting systems produced overcounts (Kimall et al. forthcoming). In the aftermath of the 2000 presidential election, Congress enacted legislation that banned some voting systems, most notably punch cards, and funded the purchase of electronic voting machines in federal elections.

New voting technologies, unfamiliar even to the most frequent voter, can lead voters to unintentionally fail to record their preferred vote choice. Overcounts are most often attributed to ballot design, voting technologies, and the interaction of these factors with voter attributes (e.g., education, age, and related physical acuity). Here human factors interact with technology and ballot design to produce errors similar to those observed with the use and operation of consumer products (Greene, Byrne, and Everett 2006; Everett, Byrne, and Greene 2006; Laskowski et al. 2004; Roth 1998).

Using different methodologies (i.e., experimental and aggregate studies), research settings (precincts, cities, and states), as well a mix of different voting technologies, the researchers have found that paper ballots are significantly superior to lever machines, direct-recording electronic (DRE) voting machines, and punched cards in reducing the residual votes (White 1960; Asher et al. 1982; Shocket, Heighberger, and Brown 1992; Nichols and Strizek 1995; Knack and Kropf 2003; Kimball, Owens, and Keeney 2004; Ansolabehere and Stewart 2005). The congruity of these rankings, however, does not provide an unambiguous

explanation for voting technology's effect on the residual vote. Conditions unique to each technology may interact with human factors and local election administration to influence the residual vote. Moreover, the actual impact of voting technology on residual votes may be exaggerated. Voters "may intentionally abstain in a particular race" (Ansolabehere and Stewart 2005, 369) because of disinterest, lack of information, or no preferred choice among the contesting candidates. There are conditions that cannot be attributed to any alternative voting technology nor readily measured with aggregate/precinct level designs. "The incidence of uncounted and spoiled ballots depends strongly and systematically on [the] 'county' in addition to equipment. We conjecture that this county effect is substantially the result of local *institutions* of electoral administration, such as the administration of local polling places or advance instruction to voters" (Ansolabehere and Stewart 2005, 386).

The usability of a voting technology refers to the efficient and timely manner with which voter cast their ballots and voters' confidence and satisfaction with the voting technology (Brennen Center 2006). Usability can directly contribute to voter error and an increase in residual votes cast. The literature on human factors and human computer interactions identifies several principals for the design of voting technologies directed at improving the voting experience. Among the metrics recommended for usability by the Federal Election Assistance Commission are a lower error rate for marking the ballot, efficient operation, and voter satisfaction (EAC 2005).

The tradeoffs between different technologies may be influenced by how elections are administered and where they are held (i.e., the polling place location). There is evidence that the performance of different voting technologies is dependent on the quantity and quality of election day poll workers (Alvarez and Hall 2006; Hall, Monson, and Patterson 2007; Atkeson and Saunders 2007). The number of adequately trained election day poll workers may have a direct impact on voter usability, satisfaction, and confidence with different voting technologies. In addition, the place where one votes, specifically the proximity of the polling place to one's residence or workplace, the availability of parking, and the ease of locating the polling place may structure a voter's experience with different voting technologies. To date, the evidence strongly points to "low tech" solutions to residual votes, a solution that favors paper ballots over more sophisticated electronic voting machines.

Counting Ballots

Counting ballots also poses several problems for election administrators. For instance, there are challenges associated with improperly marked ballots. In addition, poll workers can count votes inaccurately. In large part, the antidote against counting inaccuracies is replication, that is, recounts. However, this is more problematic when voting is conducted on DRE voting machines. Without a paper record of the vote, and only a electronic record, it is difficult to know what is being recounted. Moreover, without a verifiable paper record from DRE voting machines, it is difficult to replicate the individual ballot for recount. All DRE machines provide the voter a summary screen of his or her ballot choices before the ballot is actually cast. Thirty-seven states require the printing of a voter-verified paper

ballot from DRE voting machines. A paper copy of the voter's ballot is available to the voter and to polling place officials. A tangible copy of the ballot cast, however, is not considered a valid ballot and thus is not relied upon in recounts. Recounts on DRE voting systems consist merely of readouts of the originally recorded votes.

Another obstacle to accurate counts of ballots is fraud. Neither electronic nor paper voting systems are immune to fraudulent activity. Fraud with electronic voting may, however, be less obvious. Post-election audits (as opposed to recounts) provide a means of determining fraud. In such audits, election officials examine a chain of events to determine, for instance, whether the number of persons who signed in to vote matches the number of ballots that were actually cast. What is not altogether obvious is whether post-election audits of electronic voting can easily and in a timely fashion detect more sophisticated means of fraudulently tampering with electronic hardware and software.

Accuracy in counting ballots appears to be an ambiguous goal closely linked to other goals of election administration, most notably the accurate recording of intended vote choices and the usability of voting systems. Counting ballots may seem to be easier on electronic voting systems, but recounts without verifiable ballots pose a significant challenge for detecting manipulation of voting systems. It is noteworthy that, to date, no significant incidence of electronic voter fraud has been reported during a national or subnational election.

Voter Performance and Confidence in Election Administration

A core goal of any election and its administration is that voters accept the outcome of the election as fair, honest, and legitimate. The legitimacy of the election is at least necessary for a smooth transition of power and for compliance with the subsequent actions of the new government. Voter confidence in the accuracy with which their vote will be counted has several sources, including partisanship, age, and voter experience (Alvarez and Hall 2008). There is a significant amount of evidence that voters do perceive electronic voting systems as more accurate and usable than other forms but that such systems are also more prone to unintentional and fraudulent errors (Alvarez and Hall 2008:138). Alvarez and Hall conclude, "Americans have a decidedly mixed view of electronic voting" (2008, 153).

Other research (Stein et al. 2008) shows a strong preference among voters for electronic voting systems over paper ballots. Their preference extends to usability as well as confidence that one's vote will be accurately counted. The latter may reflect a growing confidence with either electronic voting or with the ever-widening role of electronic technology in society. A recent study (Byrne et al. 2008) reported that voter preferences were significantly greater for electronic voting systems but that the same voters' performance was *not* markedly better using electronic voting machines than traditional paper ballots. The public's preference for voting systems may not match their usability and performance, posing a serious challenge for the future administration of elections. This study suggests that a greater level of performance does not accompany greater voter convenience.

The Choices and Trade-Offs for Future Election Administration

Voter Registration

Relaxed voter registration requirements such as election day registration or the elimination of voter registration altogether would seem to enhance voter turnout, but the implementation of such changes in the near future seems unlikely. One reason election administrators are unlikely to relax these requirements is due to their commitment to ensure against voter fraud. A way to eliminate voter registration and still retain some prophylactic against voter fraud is to replace voter registration with the establishment of a national voter identification system. Many countries utilize this system to track the movement of citizens, including permanent and temporary residents, for the purposes of work, taxation, and government benefits. Such a system would be similar to that followed by the federal government in assigning Social Security numbers to people at the time of their birth. The maintenance of a national registry would insure that persons attempting to vote would present a current identification that would be valid for their current address and validate their voter eligibility in a specific venue (i.e., state and locality).

The public's support for a national identification system is strong, albeit largely driven by concerns about terrorism. A 2007 poll conducted by Zogby International (2007) reported that seventy percent of respondents supported a national photo identification system maintained by the federal government. Public support for a national ID system, however, might not carry over to a national identification card for voter validation in national and subnational elections.

Convenience Voting

Greater convenience voting is probably not the most efficacious means of assuring that everyone who wants to vote can participate. This is an ironic conclusion since the public prefers convenience when voting, especially the sort that current technology affords (i.e., voting on the Internet). Notwithstanding concerns about the security of ballots cast over the Internet (Jefferson, Rubin, and Simons 2007), there is little evidence to support the hypothesis that greater voter convenience will increase voter participation.

The best administrative practices for enhancing voter turnout seem to be those that focus on election day voters (i.e., late deciders), which make the locations at which to vote on election day less competitive with other activities on election day (e.g., larger venues with ample parking, access to well traveled roadways, a larger number of voting stations/machines and poll workers). The likelihood that these changes will be adopted seems dependent on the costs associated with conducting elections. Since the enactment of HAVA in 2003, the cost of conducting elections has increased significantly, mainly due to HAVA's requirement that older punch card and lever voting systems be replaced with electronic voting systems. Until recently, federal funding covered the cost associated with the purchase of new voting equipment. Without federal funds to replace aging voting equipment, local governments are likely to look for ways to reduce the cost of conducting elections

while still maintaining voter satisfaction previously enhanced with new voting machine technology. There is evidence to suggest that election day polling centers and their degree of accessibility have a significant and negative effect on the cost of conducting elections (Stein and Vonnahme 2010). This is also true for all-mail elections. Moreover, the cost advantage of polling centers might motivate the adoption of election practices that also increase access to the ballot for infrequent voters.

Recounts and Audits

The obstacle to timely, accurate, and fair recounts and audits is not an immature technology. To the contrary, the technology exists to conduct recounts and audits on any voting system, paper or electronic. The problem for the future is the disparity between the public's frequent use and reliance on technology, their skepticism about the accuracy and security of electronic voting systems, and their ability to accurately replicate vote counts. As Alavarez and Hall (2008) suggest, future generations who are more familiar with and trusting of technology will replace voters who are less familiar and trusting of these electronic voting systems.

Predictions

The Short-Term Prediction

The public's demand for greater technological conveniences when voting confront two short-term (i.e., three- to five-year) obstacles: first, the cost of expensive electronic voting equipment and, second, the public's concern with security. A likely compromise to Internet or electronic voting is permanent mail-in voting and all-mail elections. Recall that with permanent and all-mail elections, voters are sent a ballot weeks before election day. Voters do not have to take the initiative to request a ballot. This mode of voting is far less costly that traditional in-person voting, either on or before election day and poses fewer security risks than electronic or web-based voting systems. Moreover, there is evidence to suggest that voters quickly embrace this mode of voting when it is first offered.[3]

One consequence of voting by mail before election day is that it may not be widely used by residentially mobile populations. For many of the reasons discussed above, residentially mobile populations often fail to re-register to vote at their new residence. These voters, who are disproportionately non-Anglo are less likely to request and use permanent mail-in ballots or participate in mail-in elections. Consequently, this mode of convenience voting may not have a uniform effect across the electorate. A mode of voting like mail-in voting may be convenient for residentially stable voters, making it easier for those who already vote.

The Long-Term Prediction

In contrast, long-term trends favor remote web-based voting. The demand for convenience voting will continue to grow as a new generation of voters who are

3 Colorado adopted permanent mail-in voting in 2007. By 2008, 64% of all ballots cast were cast by mail before election day.

comfortable with and reliant on Internet technology replaces older voters who are often less familiar withsuch technology. Objections to this mode of voting on the basis of added cost and security risks will abate. The cost of the technology and its access will grow to the point where even the most mobile of persons will be able to remotely receive and send their ballots while retaining their eligibility to vote. The cost of conducting elections via the Internet or on electronic voting systems is likely to decline as technology and the needed security are enhanced.

References

Alvarez, R. M., and T. E. Hall. 2006. Controlling Democracy: The Principal-AgentProblems in Election Administration. *Pol'y Stud. J.* 34:491-510.

————. 2008. *Electronic Elections: The Perils and Promises of Digital Democracy*. Princeton: Princeton University Press.

Ansolabehere, S., and C. Stewart III. 2005. Residual Votes Attributable to Technology. *J. Pol.* 67:365-89.

Asher, H. 1982. The Effect of Voting Systems on Voter Participation." Paper presented at the Ann. Meeting of the Midwest Political Science Association, Milwaukee, Wisconsin.

Atkeson, L., and K. Saunders. 2007. Election Administration and Voter Confidence: A Local Matter. Manuscript, University of New Mexico.

Bensel, R. 2004. *The American Ballot Box in the Mid-Nineteenth Century*. Cambridge: Cambridge University Press.

Bernstein, R., A. Chadha; and R. Montjoy. 2001. Overreporting Voting: Why It Happens and Why It Matters. *Pub. Opinion Q.* 65(1):22-44.

Berinsky, A. 2005. The Perverse Consequences of Electoral Reform in the United States.*Am. Pol. Res.* 33(3):471-91.

————, N. Burns, and M. Traugott. 2001. Who Votes by Mail? A Dynamic Model of the Individual-Level Consequences of Vote-By-Mail Systems." *Pub. Opinion Q.* 65(2):178-97.

Brady, H., and J. McNulty. 2004. The Costs of Voting: Evidence from a NaturalExperiment. Paper presented at the Ann. Meeting of the Soc'y for Pol. Methodology, Stanford University.

Brennan Center for Justice. 2006. *The Machinery of Democracy: Voting System Security, Accessibility, Usability, and Cost*. The Brennan Center for Justice at New York University School of Law, New York.

Byrne, M., S. P. Everett, K. K. Greene, D. S. Wallach, K. Derr, D. Sandler, and T. Torous. 2008. Electronic Voting Machines versus Traditional Methods: Improved Preference, Similar Performance. Paper presented at the CHI 2008 Proceedings—Measuring, Business, and Voting.

Cassel, C., and L. Sigelman. 2001. Misreporters in Candidate Choice Models. *Pol. Res. Q.* 54(3):643-55.

Dyck, J., and J. Gimpel. 2005. Distance, Turnout and the Convenience of Voting. *Soc. Sci. Q.* 86(3):531-48.

Everett, S. P., M. D. Byrne, and K. K. Greene. 2006. Measuring the Usability of paper ballots: Efficiency, effectiveness, and satisfaction. Forthcoming in *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*. Santa Monica: Human Factors and Ergonomics Society.

Gastwirth, J., A.Krieger; and P. Rosenbaum. 1998. Dual and Simultaneous Sensitivity Analysis for Matched Pairs. *Biometrika* 85(4):907-20.

Gimpel, J., and J. Schuknecht. 2003. Political Participation and the Accessibility of theBallot Box. *Pol. Geography* 22(4):471-88.

Greene, K. K., M.D. Byrne, and S.P. Everett. 2006. A comparison of usability between votingmethods. Forthcoming in *Proceedings of the 2006 USENIX/ ACCURATE Electronic Voting Technology Workshop*.

Hall, T., Q. Monson, and K. Patterson. 2007 Poll Workers in American Democracy: An Early Assessment. Manuscript in author's possession.

Hanmer, M., and M. Traugott. 2004. The Impact of Voting by Mail on Voter Behavior. *Am Pol. Res.* 32(4):375-405.

Hill, J., J. Reiter, and E. Zanutto. 2004 Comparision of experimental and observational data analyses. *Applied Bayesian Modeling and Causal Inference from Incomplete Data Perspectives*. Edited by A. Gelman and Xia Li Meng. Hoboken, N.J.: John Wiley & Sons, Inc.

IBOPE Zogby International. 2007. *Available at* http://www.zogby.com/ news/readnews.cfm? ID=1290.

Jefferson, D., A. Rubin, and B. Simons. 2007 A Comment on the May 2007 DoD report on Voting Technologies for UOCAVA Citizens. *Available at* http://www. servesecurityreport.org/SERVE_Jr_v5.3.pdf.

Karp, J. A. and S. A. Banducci. 2000. Going Postal: How All-Mail Elections Influence Turnout. *Pol. Behav.* 22(3):223-39.

———. 2001. Absentee Voting, Mobilization, and Participation. *Am. Pol. Res.* 29(2):183-95.

Kimball, D. C., C. T. Owens, and K. M. Keeney. Unrecorded Votes and Political Representation. Forthcoming in *Counting Votes: Lessons from the 2000 Presidential Election in Florida*. Edited by Robert P. Watson. Gainesville: University Press of Florida, pp. 135-50.

Knack, S. 1995. Does "Motor Voter" Work? Evidence from State-Level Data. J. Pol. 57(3):796- 811.

———, and M. Kropf. 2003. Voided Ballots in the 1996 Presidential Election: A County-Level Analysis. *J. Pol.* 65:881-98.

Kousser, T., and M. Mullin. 2007. Does Voting by Mail Increase Participation? Using Matching to Analyze a Natural Experiment. *Pol. Analysis* 15(4):428-45.

Laskowski, S. J., M. Autry, J. Cugini, W. Killam, and J. Yen. 2004. Improving the usability and accessibility of voting systems and products. National Institute of Standards and Technology (NIST) Special Publication 500-56.

Leighley, J. E. 1995. Attitudes, Opportunities, and Incentives: A Field Essay on Political Participation. *Pol. Res. Q.* 48(1):181-209.

Magleby, D. 1987. Participation in Mail Ballot Elections. *W. Pol. Q.* 40(1):79-91.

Morgan, S., and C. Winship. 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge: Cambridge University Press.

Nagler, J. 1991. The Effect of Registration Laws and Education on U.S. Voter Turnout. *Am. Pol. Sci. Rev.* 85(4):1393-1405.

Neeley, G., and L. Richardson. 2001. Who is Early Voting? An Individual Level Examination. *Soc. Sci. J.* 38(3):381-92.

Nichols, S. M., and G. A. Strizek. 1995. Electronic Voting Machines and Ballot Roll-Off. *Am. Pol. Q.* 23:300-18.

Oliver, J. E. 1996. Who Votes at Home?: The Influence of State Law and Party Activity on Absenteee Voting and Overall Turnout. *Am. J. Pol. Sci.* 40(2):498-513.

Rosenbaum, P. 1989. On Permutation Tests for Hidden Biases in Observational Studies: An Application of Holley's Inequality to the Savage Lattice. *Annals of Stat.* 17(2):643-53.

Rhine, S. 1996. An Analysis of the Impact of Registration Factors on Turnout in 1992. *Pol. Behav.* 18(2):171-85.

Roth, S. K. 1998. Disenfranchised by design: voting systems and the election process. *Info. Design J.* 9(1):1-8. *Available at* http://www.informationdesign.org/pubs/roth1998.html.

Sigelman, L. 1982. The Nonvoting Voter in Voting Research. *Am. J. Pol. Sci.* 26(1):47-56.

Shocket, P., N. R. Heighberger, and C. Brown. 1992. The Effect of Voting Technology on Voting Behavior in a Simulated Multi-Candidate City Council Election: A Political Experiment of Ballot Transparency. *W. Pol. Q.* 45:521-37.

Silver, B., P. Abramson, and B. Anderson. 1986. The Presence of Others and Overreporting of Voting in American National Elections. *Pub. Opinion Q.* 50(2):228-39.

Southwell, P., and J. Burchett. 2000. The Effect of All-Mail Elections on Voter Turnout. *Am. Pol. Q.* 29(1):72-80.

Squire, P., R. Wolfinger, and D. Glass. 1987. Residential Mobility and Voter Turnout. *Am. Pol. Sci. Rev.* 81(1):45- 66.

Stein, R.M. 1998. Early Voting. *Pub. Opinion Q.* 62(1):57-70.

————, and P. Garcia-Monet. 1997. Voting Early, But Not Often. *Soc. Sci. Q.* 78:657-77.

————, C. Owens, and J. Leighley. .2003. .Electoral Reform, Party Mobilization and Voter Turnout. Paper presented at 61st Ann. Meeting of the Midwest Pol. Sci. Ass'n, Chicago, Ill.

————, and G. Vonnahme. 2008. Engaging the Unengaged Voter: Vote Centers and Voter Turnout. Forthcoming in *J. Pol.*

————, G. Vonnahme, D. Wallach, and M. Byrne. 2008. Voting Technology, Election Administration, and Voter Performance. *Election L. J.*_7(2):123-35.

————, and G. Vonnahme. 2010. The Cost of Elections. Report prepared for the Pew Charitable Trusts.

Teixeira, R. 1992. *The Disappearing American Voter*. Washington, D.C.: Brookings Institution Press.

Tomz, M., and R. P. Van Houweling. 2003. How Does Voting Equipment Affect the Racial Gap in Voided Ballots? *Am. J. Pol. Sci.* 47(1):46-60.

U.S. Bureau of the Census. 2009. Residential Mover Rate in U.S. is Lowest Since Census Bureau Began Tracking in 1948. *Available at* http://www.census.gov/newsroom/releases/archives/mobility_of_the_population/cb09-62.html.

Wand, J. N., K. W. Shotts, J. S. Sekhon, W. R. Mebane, Jr., M. C. Herron, and H. E. Brady. 2001. The Butterfly Did It: The Aberrant Vote for Buchanan in Palm Beach County, Florida, *Am. Pol. Sci. Rev.* 95:793-810.

Wolfinger, R., and S. Rosenstone. 1980. *Who Votes?* New Haven; Yale University Press.

State Trial Courts: *A Virtual Future?*

John M. Scheb II, The University of Tennessee–Knoxville
Colin Glennon, The University of Tennessee–Knoxville
Hemant Kumar Sharma, The University of Tennessee–Knoxville
Dorian Stiefel, The University of Tennessee–Knoxville
Kristin Wagers, The University of Tennessee–Knoxville

Introduction

It is axiomatic that courts of law play an important role at all levels of governance in the United States. As a result, any examination of the future of governance must include consideration of the judicial branch. The obvious question to be addressed at the outset is, "Will courts continue to have a prominent role in our society in the future?" We think the answer to this question is most certainly, "Yes." As traditional social structures continue to evolve, society will rely increasingly on the enforcement of legal norms to maintain order and cohesion and to resolve conflicts. So, courts are certain to have a future, but the broader query is: "What will that future entail? What will courts look like a hundred years from now?"

As a starting point, we must acknowledge the limited scope of our inquiry. Our focus is on the state trial courts, where the great majority of civil and criminal cases in this country are adjudicated. Today, state trial courts are grappling with a number of extremely difficult issues, including (but certainly not limited to): providing security for judges, jurors, and other courthouse personnel; creating effective alternatives to incarceration in the face of overcrowded jails and prisons; developing effective treatment programs for persons with substance abuse issues; reducing the financial and temporal costs of litigation; providing access to justice for those of very limited means; securing stable sources of funding; and adapting to increased use of specialized courts and procedures. All of these issues are precipitating significant changes, not only in the administrative processes of courts, but also in the way that cases are adjudicated.

Together, these issues have precipitated the greatest challenge facing state trial courts in the future, namely, *managing the caseload*. Even as state trial courts provide security, create alternatives to incarceration, develop effective treatment programs, they must also deal with the drastic increases in caseloads and the reality of unchanging or decreasing budgets. We assert that the key challenges driving caseload management are *managing information* and *managing the courtroom workgroup*. Here, we introduce the great challenge of managing the caseload and the constituent challenges of managing the information and managing the workgroup. We then offer visions for the future that address each of the constituent challenges. Finally, we offer three visions for the future that may completely eliminate the

greatest challenge of managing the caseload: a vision of automated justice; a vision of Artificial Intelligence; and a vision of privatization.

The Significant Future Challenges

Caseload Management—the Greatest Challenge

Frequent observation of local courts over the years, interviews with judges and other court personnel, and a review of the court administration literature lead us to conclude that the greatest challenge facing the state trial courts is *managing the caseload*. Fundamentally, many courts are entirely overwhelmed with cases.

This concern is most pronounced in our state court systems. While federal judges adjudicate several hundred thousand cases per year, the fact that state courts handle several million in a year is simply overwhelming.[1] The reality of state trial caseloads becomes inconceivable when matched to the fact that there are just over 31,000 state trial judges in the nation.[2] Further, the problem is not likely to be attenuated in the future—assuming (as we do) that demands on courts will increase while funding fails to keep pace. This issue has been exacerbated in recent years by the economic downturn, which has significantly increased the number of criminal defendants seeking the assistance of public defender offices and the number of civil litigants electing to proceed without assistance of counsel.[3]

A harbinger of the future of courts can be observed in the recent trends toward pretrial diversion, alternative dispute resolution, and drug courts. Ostensibly, these can be viewed as an effort by courts to manage caseloads by diverting cases out of traditional legal processes. Yet the problem is not merely one of volume. Many of the cases that state trial courts deal with today require more than traditional adjudication and, increasingly, courts are being asked to target particular societal problems. Thus, we have seen the development of numerous specialized "problem-solving courts" across the country; specifically, "[t]he U.S. currently has more than 2,000 drug courts, 200 mental health courts, 250 domestic violence courts, 30 community courts, and 500 other models (e.g., homelessness, truancy, teen, and sex offense courts), with dramatic growth expected in the years ahead."[4] Specialized courts have also been innovative in changing how cases are placed on the docket in order to have less complex cases resolved sooner and in carrying out "therapeutic jurisprudence,"[5] through which courts administer treatment services in addition to

1 Robert A. Carp, Ronald Stidham, and Kenneth L. Manning, *Judicial Process in America*, 8th ed. (Washington, D.C., CQ Press, 2011) p. 70.

2 *See* Shauna M. Strickland, Chantal G. Bromage, Sarah A. Gibson, Ashley N. Mason, and William E. Raftery, comps., "State Court Caseload Statistics: An Analysis of 2007 State Court Caseloads," Fig. G (Williamsburg, Va.: National Center for State Courts, 2009), pp. 92-93, *available at* http://www.ncsconline.org/D_Research/csp/2007_files/StateCourtCaseloadStatisticsFINAL.pdf.

3 *See* Richard Zorza, "Access to Justice: Economic Crisis Challenges, Impacts, and Responses," in Carol R. Flango, Amy M. McDowell, Charles F. Campbell, and Neal B. Kauder, *Future Trends in State Courts 2009* (Williamsburg, Va.: National Center for State Courts, 2009), p. 9.

4 Christine Sisario, "Improving Outcomes through Better Data Tracking: The Use of Technology in Problem-Solving Courts and Beyond," in Flango et al., *Future Trends in State Courts 2009*, *supra* , note 3, at p. 92.

5 Carp et al., *Judicial Process in America*, *supra*, note 1, at pp. 65-66.

traditional case proceedings.

Another recent trend that affects the future of courts is the move towards out-sourcing and privatization. Today, "ADR (alternative/appropriate dispute resolution), private judging, security services, facilities management, probation and community-based treatment programs, print-shop operations, court-reporter transcription services, case/cash management information services, e-f[i]ling, imaging, drug testing, public relations, employee assistance and counseling, and payroll services are frequently outsourced by courts across America."[6] In this regard, state trial courts have become increasingly "flat" organizations,[7] and there is no reason to believe that this trend will abate into the foreseeable future.

Finally, there is a trend towards collaboration and engagement. Of course, courts cannot operate in isolation from other agencies of government—or from social institutions like schools, community organizations, or even entities in the private sector. Indeed, the term "governance," as distinct from "government," "incorporates the idea that public, private, and nonprofit organizations working together, rather than government agencies acting unilaterally, are the dominant mode for delivering goods, services, and opportunities to citizens."[8] From the perspective of the courts, good governance entails working together with effective foresight and accountability to make decisions that are efficient, fair, unbiased, consistent, transparent, and attempt to do the most good for the most people. Given these definitions, the greatest challenge of managing the caseload is to manage all of the issues and trends outlined above while simultaneously addressing the challenges of managing information and managing the courtroom workgroup.

The Challenge of Managing Information

One element of the great challenge of caseload management is the *challenge of managing information*. Courts are information-intensive environments, and the scope and complexity of information are only increased as courts move from traditional adjudication toward a role as problem-solvers. Judges, lawyers, clerks, and court administrators routinely generate (and therefore must manage) a tremendous volume of information—including dockets, warrants, orders, opinions, memoranda, pleadings, briefs, reports, transcripts, depositions, subpoenas, etc. It is not uncommon for judicial proceedings to "bog down" as judges muddle through thick case files looking for particular paper records. Further, the advent of computers has not solved the problem. As Bud Borja, Chief Information Officer of the Oregon Judicial Department, has observed, court information systems "are siloed, inflexible, and unable to leverage new technologies or improve business practices. They inhibit our ability to promote new and improved ways to share

6 Gordon M. Griller, "The Growth Of Outsourcing: Courts Are Becoming Flatter," in Flango et al., *Future Trends in State Courts 2009*, *supra*, note 3, at p. 18.

7 *See generally* Thomas L. Friedman, *The World Is Flat: A Brief History of the Twenty-first Century* (New York: Farrar, Straus, and Giroux Publishers, Inc., 2005).

8 Robert F. Durant, "The Twilight of the Neoadministrative State? Crises, American Political Development, and the 'New Interventionism,'" paper presented at the Workshop on the Future of Governance, Howard H. Baker Jr. Center for Public Policy, University of Tennessee, Knoxville, Tenn., Oct. 15, 2010, p. 1.

court data with criminal justice and human service agencies, legal and business stakeholders, and the public."[9]

Electronic Courts

Court administrators have begun to grapple with the information challenge, and already it is not uncommon to hear them speak of "e-Courts." Oregon is one of the more innovative states in this regard, and is moving toward a statewide electronic courthouse that, when fully implemented, "will expand and simplify access to courts, allow electronic document filing and payment of fees, and provide numerous other benefits."[10] Presumably, one of the benefits to be reaped by moving to electronic courts is financial savings. Going paperless has obvious benefits to lawyers and litigants. "E-filing saves time, which in turn reduces transaction costs—costs ultimately borne by clients."[11] One study of courts in Manatee County, Florida, found that the electronic filing of over two million documents saved nearly $1,000,000 in a single year.[12]

Moving from paper documents to e-filing may also benefit courts with respect to the quality of information used in decision making. Jim MacMillan at the National Center for State Courts has observed:

> Many courts are continuing to view and insist that E-filed electronic documents should continue to be functionally the same as their paper and much dumber cousins. Please consider that information entombed in a paper document is now locked as to the accuracy of the moment it is printed. It is essentially a snapshot. This of course results in all sorts of problems as to the information accuracy when that paper document is later read and used. And unfortunately, judges and court staff are relying on the accuracy of that locked paper information to make decisions that affect people's lives.[13]

However, there is still resistance to the movement toward e-Courts. Joseph Jarret, currently the Law Director for Knox County, Tennessee, recalled that when he practiced law in Florida, "despite the fact that the Federal Court went to an

9 Bud Borja, "Oregon eCourt—Improving Judicial Outcomes and Services," in Flango et al., *Future Trends in State Courts 2009*, *supra*, note 3, at p. 87.

10 Paul J. De Muniz, "Building Tomorrow's Courts Today," in Flango et al., *Future Trends in State Courts 2009*, *supra*, note 3, at p. 83.

11 Ben F. Tennille and Corinne B. Jones, "Developments at the North Carolina Business Court," in Carol R. Flango, Amy M. McDowell, Charles F. Campbell, and Neal B. Kauder, *Future Trends in State Courts 2010* (Williamsburg, Va.: National Center for State Courts, 2010), p. 92.

12 R.B. Shore, Mark Singer, and Carole Pettijohn, "The ROI of Emerging Technologies," paper presented at the National Court Technology Conference, Denver, Colo., Sept. 2009.

13 James E. McMillan, "Why the Future Is Not Paper – Second in a Series," Court Technology Bulletin, National Center for State Courts, June 23, 2010, *available at* http://courttechbulletin.blogspot.com/2010/12/why-future-is-not-paper-second-in.html.

exclusive e-filing Public Access to Court Electronic Records (PACER) system, nevertheless, at least one judge insisted that all documents be filed 'the old fashioned way.'"[14]

Judge Lee Haworth, Chief Judge of Florida's Twelfth Circuit, points out that going paperless does not necessarily improve efficiency. He specifically indicates that there is considerable room for technological improvement of the electronic environment:

> [O]ne of my counties, Sarasota, has eliminated paper files. Judges view almost all documents through an intranet that gives access to electronically scanned or e-filed papers. The standard the court has set for clerks in the Twelfth Circuit is that the new technology they are designing for us must meet or exceed the efficiencies of the paper environment. Judges should be able to access documents in case files from the bench or in chambers quickly. However, judges handling large dockets or seeking to find specific documents in cases with many pages report that the current systems actually retard this ability. Timed tests prove critical documents are more speedily located in paper files the old fashioned way. This situation is due primarily to the lack of index and document word search functionalities, a deficiency clerks' vendors are attempting to remedy.[15]

Therefore, the mere reliance on electronic systems may not guarantee improvements in judicial efficiency or reductions in information management challenges. The development of effective, efficient electronic information systems is, without question, one of the most pressing issues facing courts today. How courts respond to the information challenge will to a great extent determine how courts operate in the decades to come. Moreover, the challenge of managing information is only one element of the great challenge of caseload management; the second element is the challenge of managing the courtroom workgroup.

Managing the Courtroom Workgroup

The second element of the grand challenge of caseload management is the *challenge of managing the courtroom workgroup*. An important goal of the courtroom work group is to handle cases expeditiously. Yet, in a society that is increasingly fast-paced, judges are often vexed by the difficulty of assembling all members of the courtroom workgroup together in the courtroom at an appointed time. In a preliminary hearing in a criminal case, for example, the courtroom workgroup consists of the judge, the defendant, the prosecutor, the defense lawyer, witnesses for the prosecution (which often include law enforcement officers), and various ancillary personnel. If any one member of the workgroup is not present, the

14 E-mail from Joseph G. Jarret, Law Director, Knox County, Tenn., Sept. 10, 2010.

15 Interview with Lee Haworth, Chief Judge, 12th Cir. Ct., Sarasota, Fla., June 28, 2010.

proceeding cannot go forward. There are a number of reasons why this happens quite frequently: the judge has taken ill; the defendant has been arrested and is in custody in another jurisdiction; the prosecutor is arguing a motion in another courtroom, and the hearing is taking longer than expected; the assistant public defender is on the phone trying to find her client; the deputy who is supposed to testify has an emergency call; and so forth. Ultimately, while managing information is difficult, managing people is even harder.

This problem is easily observable when one visits courts of limited jurisdiction, such as Tennessee's General Sessions Courts. In the General Sessions Courts, judges are asked to deal simultaneously with numerous cases at various stages of the judicial process. In a setting that more resembles a retail store than a court of law, lawyers, witnesses, and defendants come and go as judges try to make progress through their dockets. Of course, many of the proceedings, such as first appearances, arraignments and routine motions, are very brief and simple . . . once the courtroom workgroup has been assembled.

To alleviate the difficulties inherent to this environment (and perhaps as a precursor to the future of adjudication), courts of limited jurisdiction increasingly conduct hearings with the assistance of remote video. In fact, it is now common to have first appearances in criminal cases conducted remotely from the jail. The defendant appears on video, which is displayed in the courtroom. The defendant likewise sees the judge on a video screen. The judge can explain the charges and make a determination on bail or counsel without having to transport the defendant to the courthouse. From such limited beginnings a new movement is developing—the movement toward "virtual courts" and virtual interactions.

Virtual Courts

By definition, the virtual court is one that exists only in cyberspace.[16] More than a decade ago, Gordon Bermant recognized that "in one configuration or another, the virtual courthouse is inevitable. Its foundation is already under construction, but the scope of the edifice and the design of its interior spaces remain open for negotiation."[17]

Predating Bermant, a scenario involving virtual courts was articulated nearly two decades ago by futurist Jim Dator:

> Same-day hearings, made possible by computers, were first applied to certain traffic cases, and then spread quickly to other areas, and throughout the system. Since these first cases did not require a physical appearance, it became more and more rare for anyone to appear physically in any courtroom. At the same time, improved means of electronic communication seemed to bring

16 "By definition, the virtual courthouse needs to exist nowhere but electronically." Gordon Bermant and Winton D. Woods, "Real Questions about the Virtual Courthouse," 78 *Judicature* 64 (1994), p. 64.

17 Gordon Bermant, "Courting the Virtual: Federal Courts in an Age of Complete Inter-Connectedness." 25 *Ohio N.U.L. Rev.* 527 (1999), p. 528 (emphasis omitted).

> the court to the place of the controversy instead of the parties to the court.[18]

Other scholars such as Paul D. Carrington, a strong proponent of virtual courts, cautioned that the movement described by Dator was not imminent at the turn of the twenty-first century. "Virtual litigation," said Carrington, "is not yet on our doorstep. Neither the profession, nor the courts, nor the litigants—and certainly not this author—are ready for it, nor can they be made ready soon."[19] Yet, only four years later, Michigan became the first state in the Union to establish a "cybercourt" with jurisdiction over business and commercial cases.[20] Because the cybercourt's jurisdiction was made concurrent with already existing tribunals, litigants have a choice between traditional and virtual litigation. In the cybercourt, complaints, answers, motions, briefs, etc. are filed online; hearings take place via videoconference. Evidence is presented through streaming audio or video.

While there are few examples of functioning virtual courts in the United States today, the concept seems to be catching on, albeit with some reservations from the bench and bar. Along these lines, Walt Smith, court administrator for Florida's Twelfth Circuit, observes that

> [a]s the technology improves, [virtual courts] may be a very real possibility. We currently use a lot of video conferencing technology in our courtrooms. However, for anything very important it is my understanding that the attorneys still prefer to come to court or have their witnesses physically appear in court, etc.[21]

Judge Scott Brownell, a trial judge in Florida's Twelfth Circuit, is very interested in the idea of virtual courts but also outlines specific impediments to the realization of such a concept:

> The concept is great. In real life, until virtual courts are a) three dimensional, b) in real time with no delay, and 3) designed on a large enough screen so anyone can see everyone else in the virtual courtroom, (perhaps a separate section of the screen for each person at the "virtual hearing") it will not adequately substitute for real courtrooms. In court a judge makes decisions every few minutes of every day in court. These are decisions that are

18 James A. Dator, "Inventing the Future of the Courts and the Courts of the Future: A Futurist's Perspective," Hawaii Judicial Foresight Congress Proc., Honolulu, Hawaii, Jan. 6-8, 1991, subsequently published in Sohail Inayatullah, ed., *Judicial Foresight in the Hawaii Judiciary* (Honolulu: Judiciary, State of Hawaii, 1994), pp. 31, 36-37.

19 Paul D. Carrington, "Virtual Civil Litigation: A Visit to John Bunyan's Celestial City," 98 *Colum. L. Rev.* 1516 (1998), pp. 1536-37.

20 Anita Ramasastry, "Michigan's Cybercourt: Worthy Experiment or Virtual Daydream?" *available at* http://writ.lp.findlaw.com/ ramasastry/20020206.html.

21 Interview with Walt Smith, 12th Cir. Ct. Adm'r, Sarasota, Fla., July 7, 2010.

> driven by not only what the law requires and what the procedures require, but also based on the needs of this party, or that lawyer, or this emergency, or that lost visitation, or the time to rest, or the impact of this evidence might have on that jury or this witness without asking some questions and reading their response with expressions as well as words. Two dimensional in its present state simply does not permit that level of decision-making.[22]

One of the possible benefits of virtual courts is expanding access to the justice system. More specifically, Douglas A. Blaze, Dean of the University of Tennessee College of Law, believes that cybercourts will improve access to justice in two ways:

> First, I think that lawyers will be able to practice in rural areas more effectively without having to travel to those areas. It will be more efficient for those lawyers and make the lawyers more readily available to rural clients. We have unequal distribution to lawyers in most states and this will help increase the availability of counsel in areas where lawyers are limited. Second, as mentioned above, I think that courts will become more "user friendly" and accessible (and as a result efficient) through the use of virtual or cybercourts.[23]

However, Dean Blaze anticipates "limiting the use of virtual or cyber courts to certain aspects of the adjudicatory processes not to the end game of trials."[24] Similarly, Judge John M. Scheb, a retired appellate judge and law professor in Florida, thinks that cybercourts have a future but is skeptical about virtual trials:

> [Virtual courts] have a potential use in many proceedings, but not for criminal trials where defendants have a constitutional right to face their accusers. In civil litigation matters that require judicial determinations of factual matters it is important for trial judges to see all witnesses in order to assess their credibility. Yet, there are limited phases of civil litigation that are susceptible to being handled in virtual courts.[25]

Judge Mary Beth Leibowitz, a criminal court judge in Knoxville, Tennessee, also foresees virtual courts in specialized instances but not for general use in criminal proceedings. In her view, the manner in which court reporters handle transcripts for criminal court could become virtual, but generally, Judge Leibowitz does not

22 Interview with Scott Brownell, Judge, Fla. 12th Cir. Ct., Bradenton, Fla., June 29, 2010.

23 Interview with Douglas A. Blaze, Dean, Univ. of Tenn. College of Law, Sept. 22, 2010.

24 *Ibid*.

25 Interview with John M. Scheb, Judge, Fla. Ct. App., 2nd Dist. (retired), Sarasota, Fla., July 1, 2010.

envision criminal courts becoming completely virtual.[26] In this regard, Judge Lee Haworth envisions a "hybrid system" in which certain proceedings will take place in an electronic environment, but like other judges we interviewed, he does not foresee, nor does he favor, the total elimination of face-to-face hearings or trials:

> I predict the next step will be virtual hearings viewable by the public where attorneys can appear for certain matters from their office with a video link to a judge in chambers or in a courtroom. However, few persons who have seen a baseball game or other sporting event in person would substitute that experience for watching it on television or a computer monitor. There is a qualitative value to being close and in the physical presence of witnesses, the judge, and the litigants, and having a ringside seat at contested matters enhances the potential that fact finders will pay closer attention and achieve a better result.[27]

Similarly, Jim McMillan, a court management consultant with the National Center for State Courts, believes that "a lot of simple and procedural work can be done virtually, but I seriously doubt that any significant litigation will be done without putting people in the same room for quite awhile."[28]

Perhaps the greatest resistance to the idea of virtual courts exists with respect to jury trials. Gordon Bermant and Winton Woods identify four key issues with this aspect of jurisprudence:

> Can judges and lawyers successfully conduct "virtual voir dire"? Where would the jurors be physically located? Could jurors adequately judge the demeanor of witnesses who were virtually present but physically elsewhere? Is it essential to our civil jury system that jurors physically sit together during trial and deliberate in the same room?[29]

Therefore, it seems that along with concerns for defendants' rights to face their accusers guaranteed in the Sixth Amendment, the role of jurors in a courtroom setting may provide the biggest obstacle for the implementation of virtual proceedings. In this regard, Frederic Lederer notes that "[f]reed of the need for a jury, virtual trials and courtrooms become much easier to institute."[30] However,

26 Interview with Mary Beth Leibowitz, Judge, Knox County Crim. Ct., Knoxville, Tenn., July 6, 2010.

27 Interview with Lee Haworth, Chief Judge, Fla. 12th Cir. Ct., Sarasota, Fla., June 28, 2010.

28 Interview with James E. McMillan, Principal Court Management Consultant, Nat'l Center for St. Cts., Williamsburg, Va., July 6, 2010.

29 Bermant & Woods, *supra* note 16, at p. 66.

30 Fredric I. Lederer, "The Road to the Virtual Courtroom? A Consideration of Today's—And Tomorrow's—High-Technology Courtrooms," 50 *S.C. L. Rev.* 799 (1990), p. 838.

one must also recognize that jury trials are becoming rare events, especially in criminal cases. Indeed, they are viewed by many as failures of the judicial system. As plea bargaining has become the overwhelming norm in criminal trials at the state level[31] and the legal system evolves more toward problem-solving courts and adjudication by non-judicial actors, the problems associated with virtual jury trials become less of a concern. Certainly, one "can more easily imagine arbitration panels or specialized tribunals resolving commercial disputes where a virtual gathering for decision-making would be accepted."[32]

Alternatives That May Eliminate the Greatest Challenge

What are the alternative visions that may eliminate the greatest challenge of caseload management? How can we reconceptualize caseload management for the future? Here, we articulate three visions for the future that may drastically reduce or eliminate caseloads for public courts: (1) the vision of automated justice; (2) the vision of Artificial Intelligence, or AI; and (3) the vision of privatization. Automated justice reduces caseload management by handling every aspect of the case automatically or with minimal human intervention. Artificial Intelligence reduces caseload management by using AI to handle the majority or all of the justice issues, either returning the final analysis to a human for a final decision or even making the final decision. Privatization reduces caseload management for the public courts by off-loading some or all of the court functions to not-for-profit and for-profit institutions.

Automated Justice

Beyond the use of e-Courts and virtual interactions in judicial proceedings, many jurisdictions have begun to use technology for police purposes. For example, according to the Insurance Institute for Highway Safety, as of September 2010, red light cameras were being used in roughly 485 American communities, and more than 68 jurisdictions employed speed detection cameras.[33] When such cameras detect that a driver is running a red light or speeding, he or she simply receives a citation by mail (and/or e-mail). The alleged offender is provided with a link to visit online and watch the video that provides evidence of the violation. He or she is given the option to pay a predetermined fine or to appear in court (or before an administrative board) to contest the charge. Although quite controversial (and certainly unpopular), these procedures appear to meet the essential requirements of due process. That is, the alleged offender receives fair notice and the opportunity to be heard, so long as the accused is not charged additional penalties for challenging the citation and winning on appeal.

---

31 *See* Carp, Stidham & Manning, *Judicial Process in America*, *supra* note 1, at pp. 190, 227.

32 Gordon Bermant, "Symposium: The Powers and Pitfalls of Technology: The Development and Significance of Courtroom Technology: A Thirty-Year Perspective in Fast Forward Mode," 60 *N.Y.U. Ann. Surv. Am. L.* 621 (2005), pp. 644-45.

33 Insurance Institute for Highway Safety, "Communities using red light and/or speed cameras as of September 2010," *available at* http://www.iihs.org/laws/auto_enforce_cities.aspx (accessed Sept. 5, 2010).

It is also true that public spaces in the United States are becoming increasingly subject to video surveillance. Most retail stores, shopping malls, and other places of business have their own cameras. Most police cars are now equipped with dashboard cameras. American cities are also following the example set by London, England, where the entire city is covered by video cameras. While the larger implications for society of such techniques remain under debate, the trend towards increased use of technology in policing citizens is quite clear.

Let us intersect the ideas of traffic cameras, citations by e-mail or mail, and video surveillance. In the not too distant future, it may well be the case that many criminal prosecutions will follow the model of the red light and speed cameras. Shoplifting, assaults, disorderly conduct, prostitution, malicious mischief, incitement to riot, passing worthless checks, and many other forms of criminal behavior can be captured on the increasingly-ubiquitous video cameras. In the future, video surveillance might well be augmented by audio. Highly-sensitive directional microphones could be attached to cameras and could be activated by humans monitoring the cameras or, one might also expect, by computers analyzing what the cameras are recording. Bio-sensors could augment the audio and video surveillance to collect bio-metric information about the individual's identity and current physiological state.

Thus, the inchoate offenses of solicitation and conspiracy, which hinge on verbal communication, could become susceptible to surveillance. Police officers would not have to be on the scene to actually make arrests. Violators could be notified by e-mail, be presented with the video and/or audio evidence against them, and be given an opportunity to contest the charges or resolve them by paying a fine, performing community service, making restitution, etc. Those who wish to contest the charges could be directed to a Website to determine whether they are eligible for appointed counsel and, if so, to the public defender's Website for an online interview. Those who fail to respond within a given time frame would be subject to arrest and processing through the traditional means. As a result, many of the minor offenses that now clog the dockets of courts of limited jurisdiction could be managed technologically and, if necessary, adjudicated with virtual interactions. Imagine a future in which the entire process from arrests to filings to verdicts to consequences to parole determinations is handled automatically.

A third, inverse aspect of technology and surveillance could relate to the use of bio-surveillance combined with genetic indicators. Studies of genetic dispositions could lead to information about the likelihood of committing a particular crime. Embedded bio-sensors examining our DNA and physiology could accurately identify proclivities for many types of crimes, especially crimes consistent with mental and physical disorders. Based on these predicted and automated diagnoses, justice could be meted out in the form of fines, treatment, rehabilitation, and/or counseling.

Artificial Intelligence

How will Artificial Intelligence (AI) affect the development of the courts? We already are witnessing the application of AI to the lawyering function. In the

Economist Technology Quarterly five years ago, there appeared an article containing the following gem:

> Given the choice, who would you rather trust to safeguard your future: a bloodsucking lawyer or a cold, calculating computer? Granted, it's not much of a choice, since neither lawyers nor computers are renowned for their compassion. But it is a choice that you may well encounter in the not-too-distant future, as software based on "artificial intelligence" (AI) starts to dispense legal advice.[34]

Also in 2007, *LiveScience* ran a short article on software designed to assist couples with divorce:

> Divorce is never pleasant, but new software is aimed at making the process a little less harrowing. The computer program combines artificial intelligence, game theory and an electronic or human external mediator to help divorcing couples settle their disputes in a fair and rational manner—and hopefully with fewer gray hairs.[35]

While recognizing the complexity and nuance of judicial decision-making, researchers have already proposed AI models to aid judicial decision-making.[36] Using AI, K.W. Chau was able to correctly predict more than eighty percent of outcomes in construction disputes in a Hong Kong court.[37] One might not be impressed with the twenty-percent error rate; but in an age of widespread skepticism about the ability of courts to find "correct" answers, who is to say that the program erred one-fifth of the time rather than the human judges?

The obvious next step is to introduce AI software into virtual courts. Conceivably, lawyers and judges could be done away with altogether! With respect to the lawyers, such a development might elicit widespread applause. However, can judging really be computerized? In this regard, one immediately recalls the term "mechanical jurisprudence" that legal realists and judicial behavioralists coined to disparage the notion that judging can and should be a purely logical or scientific

34 "AI am the law," *Economist Tech. Q.*, Mar. 20, 2005, *available at* http://www.economist.com/node/3714082 (from the Mar. 12, 2005, print edition).

35 Melinda Wenner, "Divorce Software Designed to Handle Negotiations," *LiveScience*, July 31, 2007, *available at* http://www.livescience.com/4569-divorce-software-designed-handle-negotiations.html.

36 *See, e.g.*, Felipe Lara-Rosano and María del Socorro Téllez-Silva, "Fuzzy Support Systems for Discretionary Judicial Decision Making," in Vasile Palade, Robert J. Howlett, and Lakhmi C. Jain, eds., Knowledge-Based Intelligent Information and Engineering Systems 7th Int'l Conf. Proc., KES 2003, Oxford Univ., Sept. 2003, pt II, *Lecture Notes in Computer Science:* vol. 2774 (Berlin: Springer-Verlag GmbH, 2003), pp. 94-100.

37 K.W. Chau, "Prediction of Construction Litigation Outcome—A Case-Based Reasoning Approach," in Moonis Ali and Richard Dapoigny, eds., Industrial Engineering and Other Applications of Applied Intelligence Systems 19th Int'l Conf. Proc., IEA/AIE 2006, Annecy, France, *Advances in Applied Artificial Intelligence*. (Berlin-Heidelberg: Springer-Verlag GmbH, 2006), pp. 548-53.

enterprise. Can a computer program, even one capable of learning, ever approximate the complex task of dispensing justice or settling disputed legal questions? Perhaps it is facile to suggest that AI could ever be employed to render decisions in virtual appellate courts or reach a verdict in a complex civil or criminal trial. On the other hand, what about routine or preliminary matters in minor civil and criminal cases? Surely the prospect of a much more advanced form AI than exists today being employed in such matters fifty years from now is quite conceivable (and we think quite likely).

Privatization of Adjudication

Private companies could handle some or all of public state trial court functions, up to and including the actual decisions. Based upon the assumption that markets can resolve disputes more efficiently than government, privatization may assure follow-up and efficiencies in processing. At the same time, the private institution's profit motive and/or the desire to perpetuate the organization will assure attention to the details and issues so that they can continue to win business.

Private companies already offer information-management solutions and private solutions such as scheduling software companies, and call-ahead services could facilitate the courtroom workgroup. Many activities of the courts—especially information-management—are already out-sourced to companies, and appellate courts have already upheld the delegation of police powers to private companies.[38] Privatization may be a way of managing public state trial court caseloads by eliminating them entirely.

Virtual Justice: A Legal Dystopia?

The challenges and visions above led us to consider a legal dystopia that may emerge from many or all of the visions, as well as the practical and ethical considerations inherent to these ideas. We therefore recommend significant and additional research and discussion on these issues so that they may be understood, framed, and managed as they emerge. We also recommend ongoing evaluation of future research so that we may all continue to learn from our work so that our previous work informs us as new ideas and opportunities emerge.

Writing in 2006, Shulamit Almog suggested that "[t]he line between the wireless courtroom, based on digital technology, and the courtroom that exists entirely in virtual space may be the line that the law should not cross."[39] Perhaps he and others are alarmed at the following scenario described by Jim Dator:

> In certain cases at first, members of juries too were no longer physically together. Eventually, whole neighborhoods, or communities, or random samples of them, became the juries. In

38 *E.g.*, City of Knoxville v. Brown, 284 S.W.3d 330 (Tenn. Ct. App. 2008) (appeal from the Circuit Court for Knox County, No. 3-649-06, Wheeler Rosenbalm, Judge, No. E2007-01906-COA-R3-CV, July 30, 2008).

39 Shulamit Almog, "Creating Representations of Justice in the Third Millennium: Legal Poetics in Digital Times," 32 *Rutgers Computer & Tech. L.J.* 183 (2006), p. 244.

> some controversies of worldwide import, indeed, the whole world was watching.
>
> During the early stages of this development, judges came to be administrators of decisions reached through experts systems and, later, AI. Then humans—more nearly philosophers than judges—were used primarily to review AI decisions. Now that scarcely seems necessary or possible, given the advances in AI on the one hand, and the virtual merger of humans with cyborgs and the imminent emergence of post-homosapiens, on the other.[40]

In a sense, some might see virtual courts as dehumanizing the judicial process. Would the increased reliance on technology to enforce legal norms and resolve disputes necessarily result in a loss of humane values in the legal system? Perhaps. Certainly this is an issue worthy of considerable thought and discussion, especially in regard to basic constitutional provisions related to a "fair trial," such as the right to "confront witnesses."

Fiscal Considerations

In terms of practical considerations related to governance, it is important to note that a judicial system reliant upon technological innovation could prove costly. As we suggested at the outset, inadequate funding is a major concern for state trial courts today.[41] Inadequate funding not only makes it more difficult for courts to perform their essential functions, but also diminishes their capacity to plan for the future and develop alternative information systems and models of adjudication. One court administrator we interviewed put it in these terms:

> The most immediate problem is creating a stable source of funding for the courts. If we [do not] accomplish this, it will have major ramifications concerning how we operate, our ability to dispose of cases in a timely manner, etc. Interestingly enough, the cutbacks we have already taken have diminished our capacity to (1) spend time looking at what the future courts might look like (i.e., how we can improve); and (2) find the funds to experiment with technology to examine different ways for people to "appear" in court.[42]

The ultimate practical consideration is to constantly ask whether a particular function can be automated or eliminated. If automated, can it be handled by traditional computer systems, virtual systems, or artificial intelligence? If eliminated, what will take the place of the missing function?

---

40 Dator, "Inventing the Future of the Courts and the Courts of the Future," *supra* note 18, at p. 37.

41 "Forty-seven states . . . face shortfalls in their budgets in FY 2009 and FY 2010, influenced primarily by the worst decline in state and local sales taxes in 50 years." Daniel J. Hall, "How State Courts Are Weathering the Economic Storm," in *Future Trends in State Courts 2009*, *supra* note 3, p. 1.

42 Interview with Walt Smith, *supra* note 21.

Ethical Considerations

Ethical considerations underpin every aspect of these visions for the future of the courts. As we address the challenge of managing information by instituting e-Courts initiatives, what are the ethical considerations for backing up the digital data; protecting the data from hackers; and ensuring that all users of the systems are able to use the systems and technologies? As we address the challenge of managing the courtroom workgroup by instituting virtual courts, what are the ethical considerations for maintaining confidentiality in the virtual space; verifying participant identities; and protecting the process from system crashes? For the alternatives that may eliminate caseload management, how do we address issues of due process; rights and responsibilities of AI; and the possible disconnects between public and private interests? We must continue to think about each of these visions and address each of the implications, especially with respect to the impacts on humans, governments, and private institutions.

Recommendations and Conclusions

Based on the foregoing challenges and visions, we identified two recommendations. First, we recommend significant additional research and discussion about the future of state trial courts in general and these challenges and visions in particular. The issues, trends, and challenges are real; what we choose to make of those—and the visions we choose to implement—depends upon ongoing thought and rigorous conversations. Second, we recommend constant and consistent evaluation of the future of courts ideas as the future emerges. We must evaluate future thinking often and rigorously to see how we are progressing as a society towards the vision.[43]

We have good reason to believe that the pressures that have moved courts in the direction of e-Courts and virtual interactions are not likely to abate and, if anything, will only intensify. Information challenges that have precipitated dilemmas for the courtroom workgroup are not likely to subside in the future. In fact, based on our discussions with experienced courtroom personnel, we suspect that these problems will only become more pronounced over time. In the short term, we anticipate continued specialization of court systems, particularly with the advent of alternative styles of dispute resolution.

In addition, we also know that the rate of technological change is increasing. Looking further into the future, then, it seems inevitable that currently unknown technological advancements will allow for virtual courts—and even virtual law enforcement—to become more prevalent. However, we submit that this movement must heed the concerns expounded upon by the myriad legal professionals before it can come to fruition. Whether in one hundred years' time these visions will have completely replaced the notion of traditional courts, we cannot say. We would venture a guess that a century from now, traditional courts will be the exception and not the rule.

---

43 James A. Dator, "World Futures Studies Federation List: The Business Case for Foresight," World Futures Studies Federation, Oct. 1, 2010.

Why People Are Poor/Wealthy: *Powerful Frames for Public Attitudes and Opinions*

Mark D. Harmon, The University of Tennessee–Knoxville

Introduction

Why are people poor, and, conversely, why are people wealthy? Those questions have filled books and journal articles, studies covering fields ranging from public policy to sociology to economics. The questions themselves, however, or more precisely the answers to those questions by respondents to public-opinion polls, represent not only a clear dividing line in the U.S. electorate, but also a strong and resilient frame through which information is sorted, stored, and applied.

Lakoff (2002, 2004) has pointed out that American conservatives take a "stern father" approach to understanding the world around them. The world is a dangerous place. People have bad instincts and must be taught right. People who are successful have achieved such status by moral uprightness and good choices. People who are not well off, conversely, are in such a state because of some personal or moral failing. Liberals or progressives in the Lakoff analysis follow more of a nurturing family model, seeing societal links and multiple causations.

Mediated portrayals of poverty also very likely play a role in public opinion about the causes of poverty. Other researchers have critiqued the failings in mediated portrayals of poverty, specifically how those portrayals skew toward urban, African-American, and personal failings frames. This article will present some of those findings, as well as other results that link mediated portrayals to how people assign causality regarding poverty, personal failings or societal problems. This research, however, will test a final link—how those views of "why are people poor/rich" represent a clear cleavage in both the American electorate and political information sorting. Secondary analyses of polls on the matter will be used to test the strength of that link. That is the purpose of this study.

Literature Review

Mediated Messages about Poverty and Race

Gilens (1996, 519-27) conducted a substantial content analysis of the 1988-1992 images of race and poverty presented in nightly network newscasts and three major news magazines. His study yielded 635 poor people (for 560 of whom race could be determined) in 214 still photos in 182 magazine news stories related to poverty and 1,100 poor people of 1,353 for whom race could be determined in a random subset of 50 out of 534 television news stories. In both media, the poor presented were substantially more likely to be African-American than the actual national percentage of black poor. Further, the most sympathetic impoverished subgroups (elderly and working poor) were underrepresented, while unemployed working-age adults, the least sympathetic group, were overrepresented.

Gould, Stern, and Adams (1981) looked at primetime television entertainment shows and found few images of poverty, and those that existed were of a sentimentalized, simple, and happy deprivation. Another analysis looked at the text of eleven *Newsweek* articles about welfare and found that the articles were dominated by a conservative view that stresses victims' failures (de Goede 1996). Clawson and Kegler (2000) found that even in college textbooks, poverty is race-coded as a "black problem" and that this view is bolstered by stereotypical images of the poor.

Two separate analyses (Cloud 1998; Clawson and Trice 2000) found that the stereotypes about race and poverty cycled through federal politics in the early 1990s. Bill Clinton's 1992 pledge to "end welfare as we know it" and the congressional Republicans' "Personal Responsibility Act" both built on mediated stereotypes about race and poverty. Cloud (1998) concluded that the "family values" language of both such stereotypes constructs the family as the site of all responsibility and change, thus privatizing social responsibility for ending poverty and racism and demonstrating how Lakoff's conservative stern-father model moves easily from media to policy.

Media source as well as story framing may play a role in how much poverty/race stereotyping is adopted by the news consumer. Iyengar (1990) found that when news media frame poverty in general terms about outcomes, the public assigns responsibility for poverty to society. When news presentations present an example of a particular poor person, however, respondents then assign causality to a failing in that individual. Sotirovic (2001) found that viewer use of cable television news and entertainment shows correlated with greater perception of welfare recipients as non-white and young and with higher estimates of federal spending on welfare. This she blames on the contextually poor, event-centered, and overly personalized approach of such programs. Persons who read public affairs content in newspapers or watched more "thematic" stories about welfare and poverty in the non-print media not only had more accurate perceptions of the dimensions of poverty, but also expressed greater support for welfare programs.

Public-Opinion Studies and a Working Theory

One study of early public-opinion polling data (Newman & Jacobs 2007) looked at attitudes toward the poor during the Depression and the subsequent New Deal and found that "the jobless were regarded with suspicion, immigrants should be forced to 'go home,' women belong in the kitchen not on the shop floor. The harsher the economic conditions (by state), the more conservative were public attitudes. Hence New Deal legislative victories accrued despite rather than because of public support" (*Ibid.*, p. 6).

The link between ideology and answers to "why poor" also appears to be cross-cultural. Wagstaff (1983) studied attitudes toward the poor among male and female respondents in Liverpool and Glasgow, using MacDonald's Poverty Scale and the Protestant Ethic Scale. He found that supporters of the British Conservative Party were more likely to blame the poor for their plight, whereas Labour Party supporters were much less likely to do so. Supporters of the Liberal / Social Democratic

Federation (SDF) Alliance fell somewhere in between. Similarly, Pandey et al. (1982) found that respondents in India with a right-wing orientation took more negative attitudes toward the poor than those with a left-wing orientation. Ideas about wealth also are very resilient. Prabhakar (2008) conducted seven focus groups with 58 members of the English public about wealth taxes. He hoped to counteract the "death taxes" frame described by Lakoff with other ways of framing the issue. Participants, however, generally clung to opposition to wealth taxes even when presented with substantial contrary information.

The correlation of "why poor" answers to political ideology can be seen as a logical extension of Attribution Theory. This theory observes that people have a compelling need to explain things and that those explanations tend to break down into causal assertions either internal to the self or external to an outside agent or force. Zucker and Weiner (2006) studied the attribution of causes of poverty among student and non-student samples. In both samples, conservatism correlated positively with individualistic causes and negatively with societal causes.

Beck, Whitley, and Wolk (1999) went one step further and sent a questionnaire to Georgia state legislators, asking them to evaluate ten explanations of poverty. The 74 respondents out of a total of 236 legislators (a 31-percent response rate) represented a good cross-section of the different demographics of the legislature. At significant levels, legislators who were Democrats, women, and people of color viewed low wages and discrimination as more important causes than did their Republican, male, and white counterparts.

Descriptive Data from Polls

Before one re-examines available datasets, however, it would be useful to review the descriptive data, the poll numbers concerning answers to questions about the causes of poverty. Those polls rather consistently show significant numbers of respondents in both the "moral failings" and "social conditions" camps, but with slight majorities or pluralities selecting the moral failings explanation.

The U. S. General Social Survey (Davis, Smith, and Marsden 1990), for example, asked why people are poor. When the option "lack of effort by the poor themselves" was presented, 46 percent of the respondents said that is a very important reason, 45 percent somewhat important, and less than 9 percent not important at all. Furthermore, 39.5 percent said that loose morals and drunkenness are a very important reason, 34.9 percent somewhat important, and 25.6 percent not important.

One specialized poll, a national telephone sampling of more than a thousand U. S. Catholics (Davidson 1995), found 214 respondents blaming poverty on "poor people's own behavior such as not managing their own money," while 761 chose "social conditions such as lack of jobs and low wages." Eighty-three chose the response "Don't Know."

The Pew Forum on Religion and Public Life Survey (2002) conducted a split sample, asking one group the general question of why American children are being raised in poverty, and another group the more specific question of why ten million American children are being raised in poverty. The two groups did not differ

in their responses. Half chose a "failure of the parents as individuals," while 31 percent opted for "social and economic problems," and between 13 and 14 percent volunteering, "Both."

Similar numbers emerged when Global Strategy Group (2005) polled Americans on behalf of the Foundation for Ethnic Understanding. The question was "Do you think poor people in this country are poor because of reasons that are largely under their own control [47 percent] or because of reasons that are largely out of their control [41 percent]?" Ten percent said, "Don't Know," and two percent refused to answer.

One previous nationwide telephone survey (NPR/Kaiser Foundation/ Kennedy School 2001), conducted in English and Spanish, took the unusual step of breaking down reported answers by income group. The sample included 294 respondents with an income of less than the federal poverty level, 613 with an income of between 100 percent and 200 percent of the federal poverty level, and 1,045 with an income above 200 percent of the federal poverty level. Results for the groups were weighted to reflect the actual distribution in the nation.

All were asked, "Which is the bigger cause of poverty today – that people are not doing enough to help themselves out of poverty, or that circumstances beyond their control cause them to be poor?" Some 39 percent of those respondents whose incomes were below the poverty line said that people were not doing enough to help themselves, but 57 percent of those respondents attributed poverty to circumstances beyond the control of the poor. Those barely above the poverty line themselves split, with 46 percent of those respondents citing circumstances, and 44 percent attributing the cause to a failure on the part of poor people to do enough to help themselves. The group comprised of those with incomes at twice the poverty line and higher was the only group for whom a majority placed the onus on the poor themselves, 50 percent to 44 percent.

The three groups did not differ substantially on direct questions about whether a lack of motivation among poor people is a major cause of poverty; slightly more than half called it a major cause, and about a third tagged it as a minor cause. Roughly the same pattern held true on a "decline in moral values" as a cause, with half characterizing it as a major cause and about three in ten identifying it as a minor cause. The differences were clearer when respondents were asked to specify the most important reason for poverty. The poor were more likely to mention drug abuse, medical bills, a shortage of jobs, or jobs that only were part-time or that paid low wages. Those slightly above the poverty line also mentioned low pay and drug abuse but added poor quality schools or declining moral values as reasons. The more economically comfortable group were most likely to mention a lack of motivation or declining moral values, but some also choose poor schools or low-paying jobs.

The public opinion split on reasons for poverty carries over into anti-poverty programs, under the generic term "welfare." Of course, welfare long has held a negative stigma in the U.S. (Gilens 1999) and has fostered persistent myths with little relation to the reality of poverty (Seccombe 2007). Two polls (Kaiser 1994; and NBC and Wall Street Journal 1994) asked fairly similar questions about the

reasons people were on welfare. The Kaiser poll found that 65 percent of the respondents attributed recipients' "choos[ing] not [to] work" as a major reason and that 26 percent identified that same as a minor reason. More than seven in ten thought a major reason people were on welfare was that welfare pays better than some jobs, and 62 percent listed as a major reason that women have more babies in order to get larger checks.

In the NBC poll, 57 percent of the respondents characterized a decline in moral values as a major reason for poverty, and 20 percent termed it a moderately important reason. Fifty nine percent of the NBC respondents thought that the breakdown of the traditional family unit was a major reason, and 20 percent believed that it was a moderate reason. Fifty three percent said a major reason was that welfare pays better than some jobs. Fully half of the respondents listed as a major reason women having babies for the purpose of receiving larger checks, and one in five labeled that as a moderate reason. An insufficiency of jobs and racial discrimination by employers were chosen less often as reasons by respondents to both polls. Poor education scored highly as a reason in the Kaiser poll, but not as much in the NBC poll.

Negative public attitudes toward the poor also appear in historical assessments of anti-poverty programs. Schwarz (1988) compiled several studies about Great Society anti-poverty programs and found that they achieved their objectives and reduced poverty. Such was not the case in public-opinion polls. A study by Americans Talk Issues Foundation (1994) found more of the respondents (31%) had a negative opinion of the 1960s War-on-Poverty programs than those who had a positive opinion (22%). Among those with a negative opinion, 45 percent said that the programs had not reduced poverty, 22 percent said that the programs had made recipients dependent on welfare, and 14 percent complained that the programs had merely created yet another government bureaucracy.

Public assumptions and mythologies about poverty and welfare also dovetail with views on immigration, race, and electoral choices. When presented in one study (Kane, Parsons & Associates 1984) with the statement, "Most refugees admitted to the U.S. wind up on welfare," 45 percent of the respondents agreed and indicated that the asserted opinion was a good reason not to allow refugees to enter the United States. Another 19% thought that the opinion was true but had no relevance; 23 percent thought it untrue; and 10 percent replied, "Don't know."

The General Social Survey (GSS) (Davis, Smith, and Marsden 2007) has asked the question, "On the average (negroes/blacks/African-Americans) have worse jobs, income, and housing than white people. Do you think these differences are . . . [b]ecause most (negroes/blacks/African-Americans) just don't have the motivation or willpower to pull themselves up out of poverty?" More than 54 percent of respondents overall replied, "Yes." One should note that the GSS first asked this question in 1977, then every year or other year starting in 1985. The percentages initially ran as high as 64.7 percent, yes. The percentage generally has declined with passing years but in 2006 was still 49.8 percent of the respondents agreeing with the statement.

If Lakoff is correct about moral politics, then the following three hypotheses will hold true. Hypothesis One is that those who view themselves as conservatives will exhibit the highest levels of associating poverty with personal moral failure, while those viewing themselves as liberal will exhibit the lowest levels of viewing poverty as associated with personal moral failure. Hypothesis Two is that those who self-identify with the Republican Party will exhibit the highest levels of viewing poverty as associated with personal moral failure, while those who self-identify with the Democratic Party will exhibit the lowest levels of viewing poverty as associated with personal moral failure. The inverse also should hold true. Thus, Hypothesis Three is that Republicans and conservatives will be more likely than their Democratic and liberal counterparts to view wealth as a consequence of moral uprightness, including hard work.

Methods

The researcher used keyword searches to find polls in which respondents gave reasons for poverty or wealth. Roper's iPoll archive was very useful in obtaining some of the descriptive data cited in the literature review. In addition, the researcher obtained the Pew Religion and Public Life and the NPR/Kaiser/Kennedy datasets through Roper's iPoll archive.[1] Each was imported into a Statistical Package for the Social Sciences (SPSS) file for data analysis. The General Social Survey, National Race and Politics Survey (Sniderman, Tetlock & Piazza 1991), and the American National Election Survey were available for online analysis through the Survey Documentation & Analysis (SDA) archive.[2], The Association of Religious Data Archives had an additional useful poll, imported into an SPSS file for further analysis.

The previously mentioned hypotheses were tested using both measures of correlation/association and, where possible, multiple regressions. The regressions tested "why poor" reasons against political philosophy and party identification and viable alternative explanations for variance such as age, income, education, and religiosity to test the strength of the relationship.

Findings: Correlations and Associations

The U.S. General Social Survey is available online from 1972 to 2006. Unfortunately, only in 1990 did GSS ask respondents questions about why people are poor. In the GSS self-identifying as a conservative correlated with attributing poverty to lack of effort and loose morals on the part of the poor themselves. Identifying as a liberal correlated with attributing poverty more to poor schools and not enough jobs. These relationships were linear and met a high standard of statistical significance (Table 1).

Party identification held to the same pattern on three of the four proffered reasons why people are poor. Greater identification with the Democratic Party also meant respondents were more likely to attribute poverty to lousy jobs or failing

1 The survey results reported here were obtained from searches of the iPoll Databank and other resources provided by the Roper Center for Public Opinion Research, University of Connecticut.

2 *Available at* http://sda.berkeley.edu/.

schools. Greater identification with the Republican Party meant associating poverty with lack of effort by the poor. Though Republicans were slightly more likely than Democrats to link poverty to loose morals or drunkenness, this tendency failed to achieve statistical significance (Table 2).

The 1991 National Race and Politics Survey presented the statement, "Most people are poor because they . . ." and offered a personal factors reply (do not try hard enough, coded 1) and a social factors reply (do not get the training and education they need, coded 2). Democrats (N=326) and Independents (N=293) leaned toward the social explanation. Both had a mean of 1.87. The 311 Republicans did not lean so heavily in that direction, having a mean of 1.79 (ANOVA, Sum of Squares 1.217, df=2, Mean Square .608, Fisher F-value 4.602, p =.0103). The 167 liberals had a mean of 1.90, compared to a mean of 1.87 for moderates (N=248) and a mean of 1.78 for conservatives (N=224). This was statistically significant in the expected duration (ANOVA, Sum of Squares 1.747, df=2, Mean Square .873, Fisher F-value 6.781, p=.0012).

The NPR/Kaiser/Kennedy School Poll (2001) showed a clear pattern on political philosophy and "why poor" questions—circumstances beyond the control of the poor versus not doing enough to help themselves. Self-identified liberals opted for the circumstances-beyond-control answer by a plurality of 207 to 124. Conservatives chose the not-doing-enough-self-help answer by nearly two to one, 257 to 131. Moderates split fairly closely, with 263 choosing the not-enough explanation and 235 selecting the circumstances-beyond-control factor. The same pattern held true in political party identification. Respondents who identified with the Republican Party replied "not doing enough" by more than two to one, 337 to 164, while Democrats chose "circumstances beyond their control" by a closer margin, 362 to 235. Independents split closely, with 224 responding "not enough self-help" to 216 responding "circumstances beyond their control."

The responses also can be analyzed by this scale, coding 1 for "not enough self-help" and 2 for "circumstances beyond their control." By that measure, declared Republicans averaged 1.33, Democrats 1.61, and Independents 1.49 (ANOVA, Sum of Squares 21.256, df=2, Mean Square 10.628, Fisher F-value 44.972, p <.0001). The means barely budged when the measure changed to which party the respondent felt more closely aligned to his or her views, Republicans 1.33, Democrats 1.61, Neither 1.49 (ANOVA, Sum of Squares 6.400, df=2, Mean Square 3.2, Fisher F-value 13.573, p < .0001). Those who called themselves conservative had a mean of 1.34, liberals 1.63, and moderates 1.47 (ANOVA, Sum of Squares 14.792, df=2, Mean Square 7.396, Fisher F-value 31.133, p <.0001).

The Pew Religion and Public Life Survey (2002) also confirmed the strong association between ideology and "why poor" answers. The scale was from 1 = very conservative to 5 = very liberal. On that scale, respondents who said people were poor because of personal failings averaged 2.69 compared to a much more liberal 3.05 for those who credited society's failures for poverty (t=9.3339, p < .0001).

The results were similar for why children, in general, were being raised in poverty: 2.63 for personal failings; 3.03 for social and economic conditions (t=7.9984, p < .0001). Modifying the question to "ten million American children

in poverty" had little effect: personal failings, 2.63; social/economic conditions, 3.00 ($t=7.5195$, $p < .0001$).

The results also did not change much when the respondents' replies were grouped into a two-by-two table by political party identification. Republicans attributed poverty to personal factors rather than societal ones by better than a three-to-one ratio, 303 to 90. Democrats also blamed personal factors, but at a much smaller ratio, 215 respondents to 134 (Chi-Square p-value < .0001). Democrats split nearly evenly, 93 to 94, on personal-versus-societal reasons for children in poverty, but Republicans stuck to personal reasons, 118 to 59 (Chi-Square p-value = .0014). The numbers were not much different when the wording was modified to "ten million children in poverty." Republicans blamed personal factors, 139 to 65; Democrats opted for societal factors, 81 to 72 (Chi-Square p-value < .0001).

Findings: Regressions

The GSS results from four potential causes were recoded and combined so that high scores were from citing social conditions (poor schools, insufficient jobs), and low scores were from citing individual failings (laziness, moral problems, or drunkenness). This served as the dependent variable. Five independent variables were entered in a multiple regression. Conservative views and strong association with the Republican Party correlated strongly with attributing poverty to individual failings; so did rising respondent income. Education ran in the opposite direction; greater education meant a greater tendency to cite social conditions for poverty. Respondent age was not associated with reasons given for poverty (Table 3).

The 1991 National Race and Politics Survey found that conservative respondents were much more likely than their liberal counterparts to say that most people are poor because they do not try hard enough. Liberal respondents opted for the choice that the poor do not get the training and education that they need. Party identification was not significant, and neither were factors such as age and income. Education fell just shy of a .05 standard of significance (Table 3).

The NPR/Kaiser/Kennedy (2001) poll had the clearest and firmest links between political philosophy or party and "why poor" answers. Liberals and Democrats opted for "circumstances beyond their control" while conservatives and Republicans said the poor are not doing enough to help themselves. These relationships were significant below a .001 standard, while education, income, age, and religiosity were not significant at all (Table 3).

The Pew Religion and Public Life Survey (2002) had one question about why people are poor and another about why children grow up in poverty. Low scores were for giving social reasons, high scores for individual reasons for poverty. The combined score on "why poor" became the dependent variable against the independent variables of education, party preference, age, income, conservative to liberal philosophy, and a religiosity score summed from four measures (church attendance, importance of religion, involvement in church, and prayer). Once again, greater education and being liberal correlated with societal explanations; being Republican correlated with individual reasons given for poverty (Table 3).

Findings: Why Are People Rich?

Answers to the inverse question "Why are people rich?" (even though rarely phrased precisely that way) also reveal a clear cleavage in the American electorate. The Pew Social Trends Survey (Pew Research Center 2008) asked the question "Which of these statements come closer to your own views—even if neither is exactly right: (1) most people are wealthy today because of their own hard work, ambition, or education, or (2) most rich people today are wealthy because they know the right people or were born into wealthy families?"

Those who attributed wealth to hard work scored a mean 2.67 on an ideology scale from 1 (very conservative) to 5 (very liberal). Those who tilted toward connections and family were more liberal at 2.99 ($t=7.3165$, $p<.0001$). Democrats cited connections over hard work 469 to 311, while Republicans declared hard work ahead of connections 330 to 172 (Fisher two-tailed Chi-Square $p<.0001$).

Democrats and Republicans, liberals and conservatives split even more dramatically on the source of success for one particular individual, George W. Bush (Gallup, CNN, and USA Today 2003). The two response options were: (1) mainly based on his own merits, or (2) mainly because of his family's wealth and influence. On a five-point partisan scale, 1 = Strongly Republican to 5= Strongly Democratic, those who cited Bush's own merits averaged 1.9966, while those who credited the subject's family connections leaned heavily toward the Democratic end of the scale, 3.7777 ($t=19.6689$, $p<.0001$). The five-point ideology scale results were similar, means of 2.4389 for merit and a more liberal 3.1494 for family connections ($t=12.1302$, $p<.0001$).

For several years, Gallup also has asked two other questions about wealth: (1) whether distribution of money and wealth in the U.S. is fair or should be more evenly distributed, and (2) whether the government should redistribute the wealth by imposing heavier taxes on the rich (Gallup and USA Today, 2009; Gallup, 2008; Gallup, CNN, and USA Today 2003). Those wanting more even wealth distribution, and those willing to use taxation to accomplish that goal, consistently were more liberal and more likely to be Democrats than those who thought current wealth distribution to be fair and opposed taxation for redistribution (Table 4).

Both Democrats, 453 to 49, and Republicans, 404 to 18, declared that the fundamental purpose of the U.S. Constitution is to protect and serve the interests of all the people regardless of their wealth and power, instead of protecting and serving the interests largely of people who are powerful and rich (National Constitution Center, 2002). A follow-up question asked, "In reality would you say all citizens actually have the same rights and freedoms offered in the Constitution or that citizens who are rich and powerful have more of them?" Republicans split almost evenly between those options, 201 to 216. Democrats, however, by 148 to 347 tilted heavily toward the view that, in practice, the rich enjoy greater rights (Chi-square $p<.0001$). In the same survey, Republicans and Democrats differed substantially in self-reported sympathy and compassion for homeless people ($t=9.4729$, $p<.0001$). Collectively these results support Hypothesis Three.

Democrats and Republicans had some differences on questions of faith, God, and wealth (Time and SRBI, 2006). The biggest split was on the statement

"Christians in America do not do enough for the poor." Republicans disagreed, 140 to 127; Democrats agreed, 137 to 70 (Chi-square p <.0001). Republicans split evenly, 136 to 136, on whether poverty can be a blessing from God; Democrats disagreed, 128 to 82 (Chi-square, p = .0209). Neither Democrats nor Republicans agreed with the proposition "If you pray enough, God will give you what money you ask for." Republicans overwhelmingly rejected the statement, 23 to 250, while a greater proportion of Democrats agreed, 41 to 173 (Chi-square p = .0007).

Partisan preference, however, did not yield statistically significant differences on a broad array of other questions: whether giving away ten percent of one's income is the minimum God expects; if one earns a lot, one should give away most of it and live modestly; Jesus was not rich, and we should follow his example; if you give away your money, God will bless you with more; God is not interested in how rich or poor one is; whether material wealth is a sign of God's blessing; whether poverty is a sign God is unhappy with something in one's life; and attitudes toward churches practicing a gospel of material success.

Discussion

These secondary analyses make a strong case that answers to the questions "Why are people poor?" and "Why are people rich?" demonstrate a clear and compelling cleavage in the American electorate and a major frame for political messages. These datasets show that conservatives and Republicans tend to blame personal failings of the poor for poverty, while liberals and Democrats tend to blame social conditions. These tendencies held up at statistically significant levels (p <.05) in most of the datasets in which party preference and/or liberal-to-conservative orientation were measured. Education was the only alternative tested that had much explanatory power. The more highly educated respondents opted for social explanations for poverty at statistically significant levels in some studies. Analyses of reasons for wealth also connect strongly with party and political philosophy. Collectively, these results support the hypotheses tested in this project.

In the light of these results, the researcher conducted one more secondary analysis, using the American National Election Study 2004 (Krosnik & Lupia 2004). One must caution that this survey never asked about the causes of poverty. It only had a "feeling thermometer," scaled 0 to 100, and used for many groups, including poor people. Nevertheless, one finds a bit of an echo of the results from the other datasets. Political party affiliation, running from Strongly Democratic to Strongly Republican, was associated at highly significant levels with the feeling thermometer regarding poor people. Democrats felt more warmly toward the poor, Republicans were colder toward the poor. Political philosophy, liberal to conservative, was not significant. Income yielded confusing results; respondent income was associated with the feeling thermometer, but household income was not. These results were placed at the bottom of Table 3.

One cannot state how long the current state of party and ideological alignment will last, but these secondary analyses confirm Lakoff's recent observations about how Republican/Democratic and conservative/liberal political orientations mirror a stern father versus nurturing family mindset. These findings also validate the recent

observation by Zucker and Weiner that Attribution Theory may well be at play in how people explain poverty. Democrats and liberals (and the highly educated) lean toward external agents and outside forces. Republicans and conservatives tend to blame the poor for their own plight, seeing individual failings as the primary, even sole, cause of American poverty.

Finally, from an ethical perspective, one must note with alarm how mediated portrayals of poverty have tracked, mirrored, and likely exacerbated this conservative alignment with a "blame the poor" perspective and the racial stereotypes that go along for the ride.

References

Americans Talk Issues Foundation. 1994. Conducted by Center for the Study of Policy Attitudes, University of Maryland, Oct. 13-16, 1994, and based on telephone interviews with a national adult sample of 900 respondents. [USUMARY.94ATI.R26-27]. Retrieved Dec. 9, 2008, from iPOLL Databank, The Roper Center for Public Opinion Research, University of Connecticut, *available at* http://www.ropercenter.uconn.edu/ ipoll.html.

Beck, E. L., D. M. Whitley, and J. L. Wolk. 1999. Legislators' perceptions about poverty: views from the Georgia General Assembly. *J.l of Soc. & Soc. Welfare* 26(2):87-104.

Clawson, R. A., and E. R. Kegler. 2000. The "race coding" of poverty in American government college textbooks," *How. J. of Comm.* 11(3):179-188.

Clawson, R. A. and R. Trice. 2000. Poverty as we know it: media portrayals of the poor. *Pub. Opinion Q.* 64(1):53-64.

Cloud, D. L. 1998. The rhetoric of 'family values': scapegoating, utopia, and the privatization of social responsibility. *W. J. of Comm.* 62(4):387-419.

Davidson, J. D. 1995. Catholic Pluralism Project, 1995, *available at* http://www.thearda.com/ Archive/Cathloic.asp.

Davis, J. A., T. W. Smith, and P. V. Marsden. 2007. General Social Surveys, 1972-2006: [CUMULATIVE FILE] [Computer file]. Chicago: National Opinion Research Center [producer] and Storrs, Conn.: Roper Center for Public Opinion Research, University of Connecticut / Ann Arbor: Inter-university Consortium for Political and Social Research / Berkeley, CA: Computer-assisted Survey Methods Program, University of California [distributors], *available at* http://sda.berkeley.edu.

de Goede, M. 1996. Ideology in the US welfare debate: neo-liberal representations of poverty. *Discourse & Soc'y* 7(3):317-57.

Foundation for Ethnic Understanding. 2005. Conducted by Global Strategy Group, Sept. 29-Oct. 10, 2005, based on telephone interviews with a national adult with oversamples of African-Americans, Jewish-Americans, and Hispanics; total sample of 1,388 respondents. Results are weighted to be representative of a national adult population. [USGLOBES.05RACE.R38] Retrieved Dec. 9, 2008, from iPOLL Databank, The Roper Center for Public Opinion Research, University of Connecticut, *available at* http://www.ropercenter.uconn.edu/ipoll.html.

Gallup. 2008. Conducted Oct. 23-26, 2008, based on telephone interviews with a national adult sample of 1,010 respondents. Interviews were conducted with respondents on landline telephones for respondents with a

landline telephone and on cellular phones for respondents who were users of cell phones only. [USGALLUP.08OCO23.R30]. Retrieved Jan. 1, 2010, from the iPOLL Databank, The Roper Center for Public Opinion Research, University of Connecticut, *available at*, http://www.ropercenter.uconn.edu. proxy.lib.utk.edu:90/ipoll. html.

Gallup, Cable News Network (CNN), and USA Today. 2003. Conducted by Gallup, Jan. 10-12, 2003, based on telephone interviews with a national adult sample of 1,002 respondents. [USGALLUP.03JA10.R10]. Retrieved Jan. 1, 2010, from the iPOLL Databank, The Roper Center for Public Opinion Research, University of Connecticut *available at* http://www.ropercenter.uconn.edu.proxy.lib.utk.edu:90/ipoll.html.

Gallup and USA Today. 2009. Conducted Mar. 27-29, 2009, based on telephone interviews with a national adult sample of 1,007 respondents. Interviews were conducted with respondents on landline telephones for respondents with a landline telephone and on cellular phones for respondents who were users of cell phones only. [USGALLUP.200906. Q08]. Retrieved Jan. 1, 2010, from the iPOLL Databank, The Roper Center for Public Opinion Research, University of Connecticut,.*available at* http://www.ropercenter. uconn.edu.proxy. lib.utk.edu:90/ipoll.html.

Gilens, M. 1996. Race and poverty in America: public misperceptions and the American news media. *Pub. Opinion Q.* 60(4):515-41.

Gilens, M. 1999. *Why Americans hate welfare: race, media, and the politics of antipoverty policy*. Chicago: University of Chicago Press.

Gould, C., D. C. Stern, and T. D. Adams. 1981. TV's distorted vision of poverty. *Comm. Q.* 29(4):309-14.

Kaiser, Henry J. Family Foundation, Harvard School of Public Health. 1994. Conducted by KRC Communications/Research, Dec. 27-29, 1994, based on telephone interviews with a national adult sample of 1,200 respondents. [USKRC.94WELF.R40-45]. Retrieved Dec. 9, 2008, from the iPOLL Databank, The Roper Center for Public Opinion Research, University of Connecticut, *available at* http://www.ropercenter. uconn.edu/ipoll.html.

Iyengar, S. 1990. Framing responsibility for political issues: the case of poverty. *Pol. Behav.* 12(1):19-40.

Kane, Parsons & Associates. 1984. Feb. 2-7, 1984, based on telephone interviews with a national adult sample of 750 respondents. [USKANE.84REF. R09E]. Public Attitudes Toward Refugees And Immigrants. Retrieved Dec. 9, 2008, from the iPOLL Databank, The Roper Center for Public Opinion Research, University of Connecticut, *available at* http://www.ropercenter.uconn.edu/ipoll.html.

Krosnik, J. and A. Lupia, principal investigators. 2004. The National Election Studies. The 2004 National Election Study [dataset]. Ann Arbor,

MI: University of Michigan, Center for Political Studies [producer and distributor]. Retrieved Dec. 15, 2008, *available at* www.umich.edu/~nes.

Lakoff, G. 2002. *Moral politics: how liberals and conservatives think*. Chicago: University of Chicago Press.

Lakoff, G. 2004. *Don't think of an elephant: know your values and frame the debate*. White River Junction, VT: Chelsea Green Publishing.

National Constitution Center. 2002. Conducted by Public Agenda Foundation, July 10-24, 2002, based on telephone interviews with a national adult sample of 1,520 respondents. Fieldwork by Robinson and Muenster Associates, Inc. Funding for the project was provided by the John S. and James L. Knight Foundation. [USPAF.02CONST.R24]. Retrieved Jan. 1, 2010, from the iPOLL Databank, The Roper Center for Public Opinion Research, University of Connecticut, *available at* http://www.ropercenter.uconn.edu. proxy.lib. utk.edu: 90/ipoll.html.

National Public Radio, Henry J. Kaiser Family Foundation, and John F. Kennedy School of Government. 2001. Telephone survey conducted by ICR/International Communications Research, Jan. 4, 2001, and Feb. 27, 2001, among a random representative sample of 1,952 respondents 18 years of age and older. Retrieved Dec. 9, 2008, from the iPOLL Databank, The Roper Center for Public Opinion Research, University of Connecticut, *available at* http://www.ropercenter. uconn.edu/ipoll.html.

NBC News and The Wall Street Journal. Conducted by Hart and Teeter Research Companies, June 10-14, 1994, based on telephone interviews with a national adult sample of 1,502 respondents . [USNBCWSJ.061794. R11A1 to R11A9]. Retrieved Dec. 9, 2008, from the iPOLL Databank, The Roper Center for Public Opinion Research, University of Connecticut, *available at* http://www.ropercenter.uconn.edu/ipoll.html.

NBC News. + June 8-12, 1995. Retrieved Dec. 9, 2008, from the iPOLL Databank, The Roper Center for Public Opinion Research, University of Connecticut, *available at* http://www. ropercenter.uconn.edu/ipoll.html.

Newman, K. S. and E. Jacobs. 2007. Brothers' keepers? *Society*. 44: 6-11.

Pandey, J., Y. Sinha, A. Prakesh,and R. C. Tripathi. 1982. Right-Left political ideologies and attribution of the causes of poverty. *Eur. J. of Psychol.* 12:327-31.

Pew Forum on Religion and Public Life, Pew Research Center for the People and the Press. 2002. Conducted by Princeton Survey Research Associates, Feb. 25-Mar. 10, 2002, based on telephone interviews with a national adult sample of 2,002 respondents. [USPSRA.032002.R47F1] Retrieved Dec. 9, 2008, from the iPOLL Databank, The Roper Center for Public Opinion Research, University of Connecticut,*available at* http://www.ropercenter.uconn.edu/ipoll.html.

Pew Research Center for the People and the Press 2008. Conducted by Princeton Survey Research Associates International, Abt SRBI, Sept. 9-14, 2008, based on telephone interviews with a national adult sample of 2,982 respondents (2,250 respondents were interviewed by landline telephone and 732 by cell phone, including 254 who had no landline telephone. [USPSRA.091808.R34HF1]. Retrieved Jan. 1, 2010, from the iPOLL Databank, The Roper Center for Public Opinion Research, University of Connecticut, *available at* http://www.ropercenter.uconn.edu. proxy.lib.utk.edu:90/ipoll. html.

Prabhakar, R. 2008. Wealth taxes: stories, metaphors and public attitudes. *Pol. Q.* 79(2):172-78.

Schwarz, J. E. 1988. *America's hidden success: a reassessment of public policy from Kennedy to Reagan.* New York: Norton.

Seccombe, K. 2007. *"So you think I drive a cadillac?": welfare recipients' perspectives on the system and its reform.* Boston: Allyn & Bacon.

Sniderman, P. M., P. E. Tetlock, and T. Piazza. 1991. 1991 National Race and Politics Survey. Conducted by Survey Research Center of the University of California, Berkeley. Random Digit U.S. telephone sample, CATI. Data Collection Period: Feb. 1-Nov. 21, 1991. Target population was defined as all English-speaking adults 18 years of age or older, residing in households with telephones, within the 48 contiguous states. The number of completed interviews was 2,223. The response rate was 65.3 percent. Retrieved from Computer-assisted Survey Methods Program, University of California [distributors], 2007, *available at* http://sda.berkeley.edu.

Sotirovic, M. 2001. Media use and perceptions of welfare. *J. of Comm.* 51(4):750-74.

Time and Schulman, Ronca, & Bucuvalas, Inc., SRBI. 2006. Conducted June 27-29, 2006, based on telephone interviews with a national adult sample of 1,003 respondents. [USSRBI.20063868.Q27I]. Retrieved Jan. 16, 2010, from the iPOLL Databank, The Roper Center for Public Opinion Research, University of Connecticut, *available at*. http://www.ropercenter.uconn.edu.proxy.lib.utk.edu:90/ipoll.html.

Wagstaff, G. F. 1983. Attitudes to poverty, the protestant ethic, and political affiliation: a preliminary investigation. *Soc. Behav. & Personality*. 11(1):45-47.

Zucker, G. S. and B. Weiner. 1993. Conservatism and perceptions of poverty: an attributional analysis. *J. of Applied Soc. Psychol.* 23(12):925-43, *available at* http://onlinelibrary. wiley.com/doi/10.1111/j.1559-1816.1993.tb01014.x/pdf.

Table 1
General Social Survey 1990:
Political Philosophy by Reasons Why People Are Poor
(scaled 1 very important, 2 somewhat important, 3 not important)

Personal:	Lack of Effort*	Loose Morals**	Social:	Poor Schools*	Lousy Jobs**
Extremely Liberal	1.86	2.34		1.64	1.55
Liberal	1.86	2.08		1.61	1.66
Slightly Liberal	1.67	1.90		1.82	1.80
Moderate	1.59	1.86		1.88	1.88
Slightly Conservative	1.56	1.85		1.95	1.94
Conservative	1.54	1.72		1.95	1.92
Extremely Conservative	1.53	2.00		1.89	1.62
* ANOVA Sum of Squares 13.617, df=6, Mean Square 2.269, Fisher F-value 5.548, p=.000.			* ANOVA Sum of Squares 14.876, df=6, Mean Square 4.207, Fisher F-value 4.207, p=.003.		
** ANOVA Sum of Squares 22.230, df=6, Mean Square 3.705, Fisher F-value 5.934, p=.000.			**ANOVA Sum of Squares 11.942, df=6, Mean Square 3.574, Fisher F-value 3.574, p=.0016.		

Table 2
General Social Survey 1990:
Political Identification by Reasons Why People Are Poor
(scaled 1 very important, 2 somewhat important, 3 not important)

Personal:	Lack of Effort*	Loose Morals**	Social:	Poor Schools*	Lousy Jobs**
Strong Democrat	1.75	1.98		1.68	1.61
Weak Democrat	1.63	1.90		1.84	1.77
Independent Leans Democratic	1.75	1.82		1.79	1.82
Independent	1.61	1.77		1.86	1.82
Independent Leans Republican	1.57	1.85		1.94	1.89
Weak Republican	1.55	1.87		1.95	1.95
Strong Republican	1.54	1.78		2.03	2.13
* ANOVA Sum of Squares 7.741, df=6, Mean Square 1.290, Fisher F-value 3.193, p=.004.			* ANOVA Sum of Squares 13.525, df=6, Mean Square 2.254, Fisher F-value 3.863, p=.001.		
**ANOVA Sum of Squares 5.193, df=6, Mean Square .866, Fisher F-value 1.370, p=.223.			** ANOVA Sum of Squares 26.856, df=6, Mean Square 4.476, Fisher F-value 8.383, p=.000.		

Table 3
National Surveys, Multiple Regression on Reasons Given for Poverty (Individual versus Social) and Political/Other Variables

Survey/Variable	*B*	*Std. Error*	*Std. Beta*	*t*	*Sig. (p)*
1990					
GSS/Liberal-Conservative	-.186	.042	-.163	-.4.474	.000
GSS/Party ID	-.103	.028	-.133	-.3.672	.000
GSS/Education	.060	.021	.101	2.845	.005
GSS/Income	-.056	.016	-.120	-3.405	.001
GSS/Age	-.003	.004	-.027	-.795	.427
1991					
Race/Liberal-Conservative	-.059	.021	-.127	-2.895	.004
Race/Party ID: D to R	-.001	.019	-.001	-.029	.977
Race/Education	.026	.013	.084	1.932	.054
Race/Income	-.006	.004	-.057	-1.320	.188
Race/Age	-.001	.001	-.127	1.296	.195
2001					
NPR/Liberal-Conservative	-.082	.033	-.124	6.299	.000
NPR/Party ID: R to D	.287	.051	.283	5.648	.000
NPR/Religiosity	-.021	.047	-.021	-.449	.654
NPR/Education	.018	.015	.059	1.194	.233
NPR/Income	-.065	.048	-.066	-1.340	.181
NPR/Age	.002	.001	.064	1.334	.183
2002					
Pew/Conservative-Liberal	-.036	.014	-.066	-2.604	.009
Pew/Party ID	.106	.013	.201	8.020	.000
Pew/Religiosity	.004	.003	.029	1.334	.182
Pew/Education	-.031	.006	-.114	-5.050	.000
Pew/Income	-.004	.004	-.023	-1.004	.315
Pew/Age	.001	.001	.031	1.435	.151
***ANES (Measured Poor People in a Feeling Thermometer rather than Poverty Reasons)**					
2004					
ANES/Liberal-Conservative	.853	.591	.068	1.442	.150
ANES/Party ID: D to R	-1.322	.492	-.154	-3.287	.001
ANES/Education	-.236	.481	-.020	-.490	.624
ANES/Resp. Income	-.453	.146	-.158	-3.104	.002
ANES/HH Income	.019	.163	.006	.116	.908
ANES/Age	.052	.043	.046	1.217	.224

Table 4
Gallup Polls on Fairness of Wealth Distribution, Taxes for Redistribution

Poll	Reply	Mean	N	Std. Dev.	Significance
March 2009/	Fair	2.34	375	.849	t=11.772
Ideology scaled	More Equally	3.05	549	.933	p <.0001
	Redistribute	3.08	466	.943	t=11.0003
	Not Redistribute	2.43	487	.881	p <.0001
Fair/More Equally by Party: Republicans 181/86, Democrats 64/276; Chi-square p<.0001					
Redistribute/Not by Party: Republicans 69/225, Democrats 249/92; Chi-square p <.0001					
October 2008/	Fair	2.30	395	.882	t=11.7888
Ideology scaled	More Equally	3.08	544	1.079	p <.0001
	Redistribute	3.20	432	1.069	t=13.0196
	Not Redistribute	2.37	525	.903	p <.0001
Fair/More Equally by Party: Republicans 229/87, Democrats 49/287; Chi-square p<.0001					
Redistribute/Not by Party: Republican 46/283, Democrats 253/85; Chi-square p <.0001					
January 2003/	Fair	2.54	312	.874	t=6.7341
Ideology scaled	More Equally	2.97	629	.916	p <.0001
January 2003/	Fair	2.05	311	1.368	t=12.8686
Party scaled	More Equally	3.40	622	1.565	p <.0001

Student Article

The Science System, Innovation, Technology, And Economic Growth: *A Review of the Current Economic Literature*

Cristina Reiser, The University of Tennessee–Knoxville

Introduction

Current scholarship that assesses the impacts of science, technology, and innovation (STI) on economic growth employs the National Innovation Systems (individually and collectively, NIS) framework—a systemic model that considers a variety of actors as part of a much larger and complex system of innovation (Freeman 1987; Lundvall 1992; Nelson 1993). Within this system, not only are the individual roles of each actor important for advancing the economy, but equally important are the interactions between the actors. According to the Organization for Economic Co-operation and Development (OECD), the NIS approach is defined as one:

> that stresses that the flows of technology and information among people, enterprises and institutions are key to the innovative process. Innovation and technology development are the result of a complex set of relationships among actors in a system, which includes enterprises, universities and government research institutes (OECD 1997, 7).

Two of the chief subsystems within an NIS are the "republic of science" (i.e., the science system) and the "realm of technology" (including industrial and entrepreneurial innovators) as defined by Dasgupta and David (1994). The distinction between science and technology is determined by the goals and reward structures of the institutes under study:

> It is the nature of the goals accepted as legitimate within the two communities of researchers, the norms of behavior especially in regard to the disclosure of knowledge, and the features of the reward systems that constitute the fundamental structural differences between the pursuit of knowledge undertaken in the realm of Technology and the conduct of essentially the same inquiries under the auspices of the Republic of Science. Loosely speaking, we associate the latter with the world of academic science, whereas Technology refers to the world of industrial and military research and development (Dasgupta and David 1994, 495).

Within the science system, researchers aim to create knowledge and diffuse it as quickly as possible. The reward for a scientist's discoveries is respect and prestige among her fellow scientific community members (Merton 1942). In the United States, much like many other economies, the science system network consists mainly of institutes of higher education (OECD 1996) but also includes federal research laboratories, nonprofit organizations, and industrial basic research labs.[1] Conversely, industry and entrepreneurs are primarily concerned with maximizing profits, and they achieve financial gains by creating new patentable products and processes. Most of the literature thus associates industry and entrepreneurs with the "realm of technology."[2], [3]

Although the republic of science, the realm of technology, and the relationship between the two are important according to the NIS framework, it is nonetheless productive to explore the economics literature on each institution individually. This article provides a synthesis of the literature on the impact of the science system as it relates to the innovative capability of the economy.[4]

The science system has three major roles in the NIS (OECD 1996). First, it generates new knowledge in the form of codified information, such as theories and natural laws published in scientific journals, new methodologies presented at industry-related conferences, or new instruments created in the laboratory (Salter and Martin 2001). According to economic growth theory, the economy can sustain permanent economic growth so long as the knowledge stock continues to grow. Second, the science system transmits existing knowledge to the future scientific labor force through education. This is because most of the science system resides within institutes of higher education. Furthermore, innovators cannot successfully exploit codified knowledge unless they (or their work force) understand the information that is presented to them. Therefore, innovators will need to employ properly trained scientists to exploit the current state of knowledge. Third, owing to its unique set of values and norms, the science system motivates scientists to diffuse their research findings as quickly and as widely as possible. How this

1 In the United States, according to the National Science Board, the largest performers of basic research in terms of financial expenditure are the higher education sector and federal research laboratories (National Science Board 2010). For 2007, the two combined accounted for 73.2 percent of all basic research expenditures in the economy with the rest being performed by industry (14 percent) and nonprofit organizations (12.8 percent) *Id*.

2 This is not to say that other sectors of the economy are not important for innovation. For example, consumers are considered to be an important source of new ideas within the innovation literature (Schmookler 1966; von Hippel 1978).

3 Innovators encompass a broader range of agents, including firms, entrepreneurs, federal laboratories, and even some members of the higher-education community. Although it would be an overgeneralization to assert that "all universities are part of the science system" or that "all firms focus solely on innovation," much of the current literature does make this generalization due to data limitations, for simplicity, or because the majority of particular sectors focus on one or the other. For instance, of total R&D expenditures by universities in 2007, 76 percent was for basic research; for total R&D expenditures by industry in 2007, 96 percent was for applied research and development (National Science Board 2010).

4 Salter and Martin (2001) provide one of the most comprehensive and appealing reviews of the impact of scientific research on the economy. They suggest six main economic benefits of publicly funded research. These are (i) increased stock of information, (ii) introduction of new instrumentation and methodologies, (iii)supply of skilled graduates, (iv) creation of professional networks, (v) technological problem solving, and (vi) creation of new firms.

knowledge is diffused is one of the more important elements in the NIS approach to understanding economic growth. So long as the science system successfully creates, transmits, and diffuses knowledge, an economy enjoys an increased capability to innovate as innovators exploit scientific research outcomes output in their own production processes.

Definitions and Concepts

Types of Research

The science system invests most of its resources in research—basic, applied, and use-inspired.[5] Traditionally, basic research (pure science) is defined as research aimed at expanding the current state of fundamental knowledge even if there is no commercial benefit to the research. The economic benefits of basic research occur when "the results of research can be used to predict the results of trying one or another alternative solution to a practical problem" (Nelson 1959, 299). Scientific knowledge allows the researcher to form credible expectations about the outcomes of different solutions—without actually investigating them individually.

Applied research is defined as determining how best to use the current stock of scientific knowledge for commercial purposes. The case for linking basic research and applied research was set by Stokes, who argued that "although a great deal of research is wholly guided by one or the other of the goals of understanding [basic] and use [applied], some studies of great importance show that the successive choices of research are influenced by both these goals" (Stokes 1997, 12). An adaptation of Stokes's quadrant is reproduced in the next section. Stokes argued that government funding should be geared towards use-inspired research rather than choosing either basic or applied research.

R&D, Technology, and Economic Growth

An economy's aggregate production function defines the ability of such an economy to produce goods and services (output) using current inputs (e.g., machinery and labor). In its most general form, output per person (GDP per capita) simply depends on inputs per person (e.g., machines per capita) and technical progress. Accordingly, the long-run economic growth rate depends on changes in inputs and/or technical progress and on how effective such inputs are in producing output. In theory, it is assumed that physical capital exhibits *diminishing marginal product*. This means that, holding all else constant, capital per worker grows as it becomes *less and less effective* at producing additional output. Eventually, growth in capital does not lead to growth in income per capita, and the economy reaches a "standstill" with zero growth.[6] Neoclassical growth theorists posit that new technologies can get the economy out of the zero-growth slump by making

5 Industry tends to invest in applied research, development, and the deployment stages of technology. However, a discussion of such is beyond the scope of this article.

6 In the most general sense, investing in physical inputs is limited to the space available in which these inputs are employed. In the extreme, one can imagine the aggregation of all landmasses on Earth and how, eventually, the building of and employment of machinery, buildings, and people will cause crowding. Inevitably, the addition of another input to production will not produce as much additional output as it would have before. However, if technological progress occurs, an economy will be able to use a given amount of capital more efficiently (i.e., growth in GDP per capita can occur without growth in capita/labor).

current inputs more productive. For example, one unit of physical capital per capita can now produce two units of output rather than one. The major point is that neoclassical growth theory tells us that an economy cannot rely solely on an input that exhibits diminishing marginal product to sustain long-run economic growth if technical progress does not exist.

Neoclassical models, however, fail to explain where technical progress comes from. Since then, economists have attempted to explain other sources of long-run economic growth that do not rely on an exogenous (i.e., external) source of technical progress. This set of literature is known as endogenous growth (i.e., generated from within) theory, which posits ways in which the economy creates growth based on the decisions made by agents within that economy. The most relevant subset of endogenous growth theory models is largely based on R&D as another input into production, alongside physical capital (Aghion and Howitt 1992; Grossman and Helpman 1994; Romer 1990). The defining characteristic of the R&D-based models of endogenous growth is the two-fold effect that research has on output—making it such a productive input that it outweighs the diminishing marginal product of capital.

Successful research not only raises the efficiency in which output is produced (e.g., by improving existing intermediate products or processes), but it also raises the productivity of the entire scientific labor force (e.g., knowledge transfers between scientists make it easier to conduct future research). Conceptually, the following example explains the latter effect: consider a room of scientists who conduct research. When one more researcher is added to the room, she adds to the existing stock of knowledge by producing an improved method of production. Subsequently, this knowledge spills over to other scientists, who then proceed to create even more knowledge and improved processes—and the cycle continues. These R&D-based models conclude that the economic growth rate is defined by the growth in science inputs (e.g., number researchers or research funding) and how large knowledge spillovers are.

R&D, Technology, and Economic Growth Globally

The same thought process carries over to a global economy context—knowledge is not only transferred within an industry but also across industries, across time, and across the globe (Feenstra 1996; Grossman and Helpman 1991). Therefore, the productivity of a country should depend on both its own R&D investment activities and the R&D activities of other countries (if R&D spillovers exist). Furthermore, productivity is influenced positively by the rate of knowledge diffusion (i.e., how quickly knowledge spillovers are diffused) and whether the receiving country has a threshold level of absorptive capacity (i.e., the level of sophistication necessary to grasp and exploit the knowledge) (Keller 2004). The absence of technological and scientific knowledge spillovers has been one explanation for the large productivity differentials among countries—if spillovers remain geographically proximate to where knowledge production takes place, then other countries will not be able to exploit it.

However, globalization has made diffusion easier, facilitating large spillovers capable of extending beyond traditional geographical limits and occurring between more developed, innovating countries and less developed, adopting countries. These spillovers tend to be somewhat asymmetric or, in many cases, unidirectional (Branstetter 2001). That is, larger and more advanced economies tend to create knowledge spillovers that diffuse to other economies (both developed and developing), but the reverse is not true. According to Keller, "For most countries, foreign sources of technology account for ninety percent or more of domestic productivity growth. At present, only a handful of rich countries account for most of the world's creation of new technology" (Keller 2004, 752).

As developing economies recognize the importance of strong scientific and technological infrastructures and continue to invest in improving such an infrastructures, they are becoming more competitive with developed economies like the United States, which implies that the net effect of economies in transition on global social welfare can be positive or negative. That is, negative effects caused by increased competition from foreign sources (e.g., job loss for unskilled labor, lowering of real wages, and loss of market share) may or may not be outweighed by the positive impacts of global knowledge spillovers (e.g., more products, lower prices, job creation for highly skilled labor, and increases in real wages for highly skilled labor).

Certainly worth mentioning is the recent debate about whether the offshoring of R&D is potentially damaging to the U.S. economy (Bhagwati, Panagariya, and Srinivasan 2004; Samuelson 2004; Tassey 2010). Pure theory of international trade shows that trade benefits the global economy by allowing countries to specialize in the manufacture and export of products that they can most efficiently produce. For instance, the United States, whose economy once was based largely on manufacturing, has moved to being a knowledge-intensive economy. Even though the transition from a manufacturing- to a knowledge-based economy creates losses in unskilled employment, the benefits of increased high-skill employment, higher variety of goods, and lower prices are seen to outweigh the losses. Nonetheless, firms in the United States have been increasingly sending R&D activities offshore (Tassey 2010). This transfer reduces costs and raises productivity, which a is good thing. Tassey explains, however, that there are potentially substantial costs from this behavior. Offshoring R&D replaces domestic R&D labor with foreign labor and allows foreign R&D centers to have a co-location advantage—(host countries will benefit more from the R&D centers located within their own borders since spillovers are geographically localized.) Furthermore, as more and more R&D functions are moved offshore, a global pool of skilled labor grows, causing downward pressure on wages for high-skill jobs (Samuelson 2004). As countries advance their scientific, technological, and innovation-enhancing infrastructures, they create competition in technology-intensive goods, a market in which the United States has historically enjoyed a near monopoly. Whether or not these benefits will outweigh any negative impact on skilled wages is unknown.

Tacit and Codified Knowledge

Within the literature, there are two distinguishable types of knowledge—codified and tacit. Codified knowledge is expressed in a format that is compact, standardized, and relatively easy to transfer (Cowan, David, and Foray 2000). Codified knowledge is produced, for instance, when scientists conduct basic research and develop new theories, models, instruments, etc. The most common channels through which codified knowledge diffuses are publications and conferences; as a result, innovators have relatively easy access to codified knowledge. Tacit knowledge, on the other hand, is knowledge that is "person-embodied" and that cannot be articulated or codified, such as skills or know-how (Polanyi 1958). Tacit knowledge becomes embodied within scientists through experience. They develop practical problem-solving skills and become highly specialized workers in their fields of study. This embodied knowledge can be diffused through joint research collaboration between the science system and firms or by firms hiring trained scientists.[7] Significantly, tacit knowledge is just as important as codified knowledge for innovation, but tacit knowledge is not always as readily accessible. The ability of an innovator to fully understand and successfully exploit codified knowledge requires some threshold level of scientific understanding within her organization (Cohen Levinthal 1989).

Furthermore, the science system largely comprises universities, whose primary focus is on educating current science students. Successful students are the backbone of the scientific labor force—but only after they have obtained their degrees and launched their careers. This means that codified knowledge and tacit knowledge are diffused not only in a geographic sense, but also in an inter-temporal sense. That is, the tacit skills being learned by current students will be utilized either when they enter the science system as qualified scientists or when they enter the scientific work force ashighly skilled, specialized employees within industry. The codified knowledge learned through current textbooks and articles will be used as the foundation on which they will build their new theories and develop new knowledge.

The Science System, Knowledge Accumulation, and Economic Growth

The characteristics of scientific knowledge play a fundamental role in understanding the contribution of that knowledge to growth in any given economy. First, the potential future economic benefits of scientific knowledge are highly uncertain and unpredictable. Advances in knowledge may be exploited immediately, or it may take many years, possibly decades, for industry to begin to apply the knowledge. Similarly, advances in knowledge may have no commercial application. This reality may disincline risk-averse firms to invest in basic research. Second, knowledge is generally characterized by nonrivalry and nonexcludability; in short, it is a public good (Arrow 1971).

Nonrivalry of knowledge means that, once it has been created, one person's use

7 Brökel and Binder (2007) discuss how Polanyi's concept of tacit knowledge has diverged into two definitions in the literature. Whether or not tacit knowledge is actually transferrable is subject to debate and has implications on the extent of knowledge spillovers. First, if tacit knowledge is truly inherent in an individual and is completely uncodifiable, then it cannot be transferred to someone else, even through interaction. Second, if tacit knowledge can be "taught" through apprenticeships or interaction with experts, then interpersonal interaction represents an important method for the transfer of such knowledge.

of it does not diminish its usefulness to others. As a result, the cost of producing a piece of information is zero for each subsequent user of that knowledge. For example, scientists A and B can use a chemical formula at the same time; or once scientist A has used the formula, scientist B can use it without incurring additional costs. A counter-example is a private good such as an apple, which cannot simultaneously be consumed by more than one person. Or, if person A eats an apple, the cost of producing another apple is greater than zero. Nonexcludability means that the producer of knowledge cannot (or chooses not to) prohibit anyone from using it.[8], [9] In proceeding with the example of the chemical formula, a scientist who publishes his research findings makes them available to others. These characteristics imply that once knowledge is created, it is costless to access and does not vanish when others use it. Thus, an institution that invests in basic research produces knowledge that can be freely accessed by other institutions—even if these other institutions did not invest in the science themselves and pay nothing to access it.[10]

According to economic theory, a profit-maximizing firm will not want to invest resources in science if it will not earn money (uncertainty, time lags) and/or if others use it without paying for it. Put another way, a profit-maximizing firm will free-ride on the investments made by other firms. The clear outcome of this profit-maximizing mentality is meager investment in basic research and a parallel tendency to regard science as proprietary and thus discourage its diffusion. Consequently, the public-good properties of knowledge provide a rationale behind a government's investment in basic research. Theoretically, governments can either subsidize investment in science or conduct it themselves. Globally, most national governments provide basic research funding to the science system.[11] Furthermore, this is why most firms use the science system as an external source of knowledge in lieu of conducting basic research in-house.

Still there are arguments that, even if the assumptions of nonexcludability and nonrivalry hold, there are supplementary reasons for firms to conduct basic research that outweigh the associated disincentives (Rosenberg 1990). First, a firm's decision to invest in basic research often is based on the anticipation that returns from the investment with cover costs, regardless of the spillover to other firms. Second, a firm's decision to conduct basic research in-house can lead to more productive applied research and development opportunities. Third, since the commercialization of science is a long-term process, large successful firms that believe they will have a strong market presence in the future are often inclined to invest in basic research today. Finally, firms that invest in basic research often create a specialized R&D department that has the means to fully understand and exploit external sources of knowledge. As Rosenberg contends, in-house science is

8 The classic example is national defense.

9 The science system has unique institutional norms such that scientists are motivated to publicly display their research findings as quickly as possible. For a detailed discussion of the norms of science, please refer to section titled "Mertonian Norms of Science and the Diffusion of Codified Knowledge," below.

10 See Kealy and Al-Ubayadli (2001), who argue that knowledge is not a public good.

11 See *supra* note 1.

a requisite for firms hoping to tap into the network of scientists conducting basic research in other sectors. Or, put another way, for firms to exploit the knowledge created by universities and laboratories, they must exceed a threshold level of in-house capability.

This last point leads to the debate over whether it really is "costless" for a firm to use knowledge. As stated by Rosenberg:

> [S]uch knowledge is regarded by economists as being "on the shelf" and costlessly available to all comers once it has been produced. But this model is seriously flawed because it frequently requires a substantial research capability [in-house]. . . . The cost of maintaining this capability is high, because it is likely to require a cadre of in-house scientists who can do these things. And, in order to maintain such a cadre, the firm must be willing to let them perform basic research (Rosenberg 1990, 171).

Thus, at issue is whether or not knowledge really is a non-rival good. For example, a firm would need to employ scientists who have the ability to understand a chemical formula, and since most scientists are driven by esteem and recognition, the firm would most likely need to grant their scientists the freedom to publish. Similarly, although a chemist, economist, businessperson, and auto technician could all access the formula at zero cost, it is likely that only the chemist could understand it entirely or adapt it for commercial use. Furthermore, for the chemist to understand the formula, he must have invested in education and attained experience in the field of chemistry. In short, accessing the formula might be free, but using it and converting it into something of economic value has an associated cost. This discussion suggests that science is not purely a public good, but rather a quasi-public good (Kealy and Al-Ubayadli 2001).

Science, Innovation, and Economic Growth—The Empirical Evidence

Nonpatent Citation Studies

Intangible knowledge is most often codified in the form of research findings published in peer-reviewed scientific journals. One method of measuring the importance of science to innovation and thus to economic growth is to use nonpatent citation analysis. Economists Adam B. Jaffe, Bronwyn H. Hall, and Manuel Trajtenberg created a large database of more than three million patents granted by the United States Patent and Trademark Office (USPTO) (Hall, Jaffe, and Trajtenberg 2001). All patents must cite the sources used for their creation, including but not limited to, other patents (known as patent citations), conference proceedings, scientific articles, and nonscientific articles. Any reference not made to other patents is referred to as a nonpatent citation (NPC).

By studying these NPCs, one can determine "backward citations" (i.e., those references Patent A cites) and "forward citations" (i.e., all subsequent patents

that cite Patent A). Economists attempting to measure a patent's dependence on scientific knowledge use backward NPCs, which reference previous scientific articles in scholarly journals, as proxies for knowledge.[12] It is important to note that the usefulness of patents as indicators of innovation has been debated within the literature. The foremost issue is that patents really are more representative of inventions and not innovations. An invention is the initial occurrence of an idea and so does not necessarily require entry into a market. An innovation is the realization of an invention, and so is a derivative of an invention. In this sense, an invention does not have economic value until it becomes an innovation. Many patented inventions never make it to the market and thus are never commercialized. Furthermore, using patent citation analysis inherently assumes that science is simply an input into technological advancement (i.e., a one-way linear model). Consequently, such analysis doesn't capture the idea that feedback loops exist and that technology can also advance science.

Recently, a study used scientific NPCs as a measure of the closeness between science and ten fields of technology and examined how this interaction is correlated with both technological productivity (patents per capita) and a country's relative technological specialization (i.e., the technological domain in which it specializes) (van Looy et al. 2003). The focus was on eight of the largest European economies, with data compiled from 1992 to 1996.[13] The results indicated that science and productivity are positively correlated for newly emerging technologies or ones that initially relied on science (biotechnology, pharmaceuticals, organic fine chemistry, and semiconductors) but not for those that began independently from science. The findings of this study suggest that science improves productivity for a particular set of technological domains. The important implication for this is that there should not be a one-size-fits-all policy governing science and technology. For those industries characterized by science intensity, policy should be directed toward spurring more collaboration between the science system and industry. For others, a different approach geared more toward innovation policy, and not science policy per se, should prevail.

Surveys of Innovative Firms

The most often used method of assessing the importance of scientific research to technological progress involves asking innovative firms directly if their innovations relied on science. The method begins by defining which firms are considered innovative (i.e., did they introduce a new or improved product or process innovation?). Next, the innovative firms are asked a series of questions related to their business organization, the sources of their ideas and information, how they developed the innovation, etc. Currently, most industrialized economies conduct bi-annual or annual surveys. The most well-known survey is the Community

12 Forward citations are used most often as indicators of the quality or value of a patent (i.e., the most forward citations a patent has, the higher its value).

13 The technological domains include biotechnology, pharmaceuticals, organic fine chemistry, semiconductors, agriculture and food, chemistry, optics, information technology, basic materials chemistry, telecommunications, and analysis, measurement, and control technology. European economies include Belgium, the Netherlands, Germany, France, the United Kingdom, Denmark, Sweden, and Finland. van Looy et al. 2003.

Innovation Survey (CIS), which is carried out by members of the European Union, Iceland, and Norway. Similar surveys are conducted in Japan, Latin American economies, South Korea, and Switzerland. Unfortunately, the United States is one of the few economies that does not use a standard innovation survey. The closest is the National Science Foundation's Survey of Industrial R&D and Innovation, which includes a much broader set of firms, including both innovating and non-innovating companies. Private surveys in the United States, such as the well-known Yale Survey and the Carnegie Mellon Survey, have been used to gauge the importance of science and innovation.

Mansfield (1991, 1998) was the first to survey firms about the reliance of their technological innovation on public research output. Firms in seven manufacturing industries were asked whether or not new products and processes would have been developed at all or without substantial delay in the absence of recent academic research (occurring within fifteen years). The 1991 conclusions show that, on average, eleven percent of product innovations and nine percent of process innovations could not have been developed without recent academic research. Furthermore, the results indicate how the importance of science varied across different types of industries. For instance, firms in information-processing, pharmaceuticals, and metals industries relied heavily on academic research, whereas petroleum and chemical industries did not.

In his 1998 update, Mansfield found that, in the absence of academic research, an average of fourteen percent of new products and eleven percent of new processes across industries could have not been developed without substantial delay. These figures transformed into new product sales of forty-four billion (5.1 percent of total sales) and seventeen billion in cost savings (two percent of total costs). There are several extensions of the Mansfield approach to analyzing a firm's reliance on public-sector research. Using data from the Fourth Mannheim Innovation Survey given to German manufacturing firms, Beise and Stahl (1999) produced findings similar to Mansfield's. Of the R&D intensive firms surveyed, 15.9 percent relied on public research institutions in Germany. For non-R&D-intensive firms, 6.2 percent utilized public research. The authors estimated an increase of DM19.4 billion (4.5 percent of total sales) in industry sales of new product innovations that relied on public research. Further, university research and public research institutes accounted for 78 and 63 percent, respectively, of the increase in industry sales. Beise and Stahl also added a statistical specification that estimated the propensity of a firm anto utilize public research. The authors found that the larger the firm in terms of employees and R&D intensity, the more likely the firm will use public research outputs.

Malo (2009) modified the Mansfield approach by focusing on the importance of public research in combinatorial chemistry to new pharmaceutical drug discovery. By asking firms to rank, using the Likert Scale,[14] the importance of different scientific and technological sub-fields of the drug-discovery process, Malo found that public research in basic science fields of organic chemistry, genomics,

14 The Likert Scale is a psychometric scale most often used in questionnaire surveys. In such suveys, respondents are asked to choose a subjective rating (e.g., strongly agree, agree, disagree, etc.) for a given statement.

and biochemistry ranks as the most important.[15] Further, his results indicate that scientific publications ranked first as a source for basic research for new entrants, whereas hiring was ranked number one for larger firms.

Some of the literature finds that public research is not as important as other sources of information for innovating firms. Monjon and Waelbroeck (2003) use data from the European CIS survey of French manufacturing firms to estimate the propensity of an innovating firm to develop a new product, which they call a radical innovation, over improving existing products, which they term incremental innovation. The results indicate that a firm is more likely to produce radical innovations if it utilizes external information sources, including the vertical chain of supply (customers and input suppliers) and patent publications. Increased reliance on information from universities and public research organizations actually decreases the probability of radical innovation, although these coefficients are not statistically significant. The authors caution that the insignificance of the coefficients does not necessarily mean that the information source is not important; rather, it could mean that the information source is important both to radical and incremental innovation.

Furthermore, Cohen, Nelson, and Walsh (2002) analyzed survey data from the 1994 Carnegie Mellon Survey of R&D managers at manufacturing firms and found that, across all industries, public research is used much less frequently for the development of ideas that issued from a firm's own vertical chain of production (customers and input suppliers). However, nearly sixty percent of firms in the pharmaceutical, petroleum, steel, machine tool, semiconductor, and aerospace industries used basic research as a source of ideas. Further, if firms did use basic research, most of them obtained it from publications in the form of research findings and instruments and methodologies or by interacting at informal meetings, instead of by using designed prototypes.

The Science System as an Educator of the Future Scientific Work Force

Tacit Knowledge and Absorptive Capacity

Knowledge-based endogenous growth models hypothesize that the accumulation of knowledge leads to perpetual growth (Aghion and Howitt 1992; Grossman and Helpman 1994; Romer 1990). In nearly all of the endogenous growth models, it is implicitly assumed that the firm has the ability to use the codified knowledge generated by the science system. However, firms may not possess the particular expertise necessary to exploit this external knowledge effectively. In other words, a firm's labor force must possess a requisite level of tacit skill if the firm is to exploit external knowledge. This is what is known as a firm's "absorptive capacity" (Cohen and Levinthal 1989). As Polanyi states of the crucial link between codified and tacit knowledge, "explicit (codified) knowledge must rely on being tacitly understood and applied" (Polanyi 1966, 7).

15 For example, one of the survey questions from Malo (2009) is "What is the relevance of public research (i.e., knowledge) in combinatorial chemistry to combinatorial drug discovery along a seven point Likert Scale where 1= highly irrelevant and 7= highly relevant?"

A well-educated scientific labor force not only generates and commercializes new ideas, but also has the inherent ability to build upon the existing stock of knowledge. The science system is not only a research haven, but because it primarily consists of institutes of higher education, its chief mission is to transmit current scientific knowledge to the future scientific labor force. In essence, the science system creates a specialized labor force by educating current students through lectures, by posing problems, and by providing laboratory experience.

Educating the future work force in science is essential for two main reasons. First, to advance the current state of knowledge, scientists must be able to fully understand it. That is, students must be fluent in the current state of science if they are to build upon it. Second, after these students have entered the scientific labor force, they can (i) create technological innovations themselves, and/or (ii) help increase the absorptive capacity of an innovating firm (Salter and Martin 2001).

Institutional Norms of the Science System and Their Impact on the Diffusion of Knowledge

According to the endogenous growth theory, the fundamental feature of knowledge is its positive spillover effects, which result from knowledge's public-good properties of nonexcludability and nonrivalry. The theory contends that new knowledge spills over, or flows, from those who have invested in and produced it relative to other parts of the economy. According to the national innovations system approach, the innovative capability of an economy depends on how effectively this knowledge flows between and among various economic agents OECD 1997). Indeed, how new scientific knowledge and the knowledge embodied within researchers diffuse across various sectors is a major determinant of technological progress.[16],[17]

Distinct diffusion channels exist for codified and tacit knowledge. The science system generates new knowledge in the form of ideas, laws, and theories that can be broken down and codified in an understandable written form. Although primarily disseminated through scientific publications, this new codified knowledge can also diffuse through nonscientific publications, seminars, and even informal interactions between scientists and industrialists.

At the same time, the science system also educates, which effectively "generates" tacit knowledge, as professors and scientists transfer the existing state of knowledge to current students through books, lectures, and hands-on learning experiences (e.g., laboratory experiments and internships). Hands-on learning also builds and

16 This is not to say that the only important linkage is that between industry and the science system. Inter-industry, customer-industry, and nonprofit-industry relationships are also important but beyond the scope of this review.

17 Of course, users of scientific knowledge include not only firms, but also the science system itself (indeed, this is the argument behind its positive spillovers). This statement merely reflects the general truism that firms utilize knowledge to innovate. The science system utilizes knowledge to improve upon it and to create more knowledge. The former sector has a more direct effect on economic growth, as technological innovation is what directly contributes to economic growth. Science is an input into technological innovation.

refines the tacit skills embodied within scientists. This tacit knowledge is further diffused to technologists through face-to-face interaction (e.g., conferences, informal meetings). In these settings, the scientist transfers his embodied knowledge to the firm/entrepreneur. This embodied knowledge is also transferred when scientists depart the science system to join industry. One could reasonably argue that the science system's capacity to transfer knowledge is just as important as its creation of the knowledge in the first place (OECD 1996).

The science system is unique among the actors within an economy that economists traditionally study. In particular, agents within the science system are not driven to maximize profits, as is the case within the realm of technology. Furthermore, a scientist's output is critically reviewed and analyzed—in effect, vetted—by her peers within the science system. This review process establishes the credibly of newly created knowledge and attests to its value and relevance to the current state of science and technology. This particular norm of the science system is what allows codified knowledge to create the positive spillovers that are so important to economic growth. If scientists were driven only to maximize profits, they would opt to keep their research findings secret to prevent others from exploiting them. Instead, the norms governing the science system motivate the producers of knowledge to diffuse their research findings as quickly and widely as possible.[18] The following section discusses how the unique institutional norms of the science system encourage the diffusion of both types of knowledge and, in the process, create the spillovers necessary for economic growth.

Mertonian Norms of Science and the Diffusion of Codified Knowledge

It is worth noting that scientists are not bound by law to reveal their research findings; they can—and sometimes do—opt to keep their work secret (Callon 1994). However, since the scientific revolution, most scientists have conformed to an evolving set of norms termed the Mertonian norms of science or represented by the acronym CUDOS (communalism, universalism, disinterestedness, and organized skepticism) (Merton 1942). The most relevant of these in terms of effects on the economy are communalism and organized skepticism. Communalism is the idea that scientists are inclined to share their work freely with their peers to gain recognition and prestige, rather than cloak their findings in secrecy in hopes of later exploiting those findings to gain financial reward. Organized skepticism refers to the idea that research findings are extensively reviewed, evaluated, and critiqued by a scientist's peers before those findings are deemed valid.[19]

Codified knowledge is so accessible largely because of the prevailing incentives that reward scientists for freely sharing their research. Motivated by the recognition and the esteem associated with the pursuit of ground-breaking research, scientists within the science system want to create knowledge and diffuse it as quickly and

18 As stated earlier, universities are the major component of the science system. It is important to note that universities do patent some of their research findings, leading to the conclusion that there is some research that is kept secret. In general, however, universities are the biggest producers of easily accessible codified knowledge.

19 David (2004) provides a review of the origins of open science.

widely as possible. The unique institutional norm of open science not only motivates quick dispersion of new knowledge, but also reduces replication of research and provides credibility to research findings by inviting public critique (Arrow 1971; Nelson 1959).

Codified knowledge is diffused chiefly through publications,[20] but the growth of information and communication technologies has expedited and simplified the diffusion process and dramatically expanded the audience (Roberts 2000). Indeed, the global reach of digitized videos, images, and publications has eliminated many geographical and intertemporal boundaries. As a consequence, codified knowledge is no longer necessarily localized to a particular region. The new scientific discovery published in an American scholarly journal now is available globally in PDF format on the Internet. Firms in Japan can immediately access this new knowledge and then capitalize on it by creating commercial applications of such knowledge.

Social Networks and the Diffusion of Tacit Knowledge

Innovating firms that tap publications and other external sources for ideas often rely on scientists who embody the tacit skills necessary to exploit the knowledge contained within these codified sources. However, firms do not necessarily have to employ scientists; they can interact with them through research collaboration and various other formal and informal channels.

Skills inherent in one person (i.e., tacit knowledge) cannot easily be transferred person-to-person in a codified form, but tacit knowledge can be diffused to industry by the movement of researchers from the science system to industry, through firms' consultation with trained problem solvers, or through structured research collaboration (OECD 2005). These tacit channels are most important in the transference of new discoveries that cannot be codified quickly, and they help explain why the more innovative firms locate near universities or other research laboratories (Arundel, van de Paal, and Soete 1995; Singh 2005; Zucker, Darby, and Armstrong 2002). The same explanation accounts for the clusters of innovative firms that exist in particular regions, including Boston's Route 128 (home of MIT and Harvard), Silicon Valley (home of Stanford University), and Oak Ridge, Tennessee (home of the Oak Ridge National Laboratory and proximate to the University of Tennessee) (Anselin, Varga, and Acs1997; Saxenian 1994). Being located close to the source of knowledge—both codified and tacit—helps facilitates the diffusion of that knowledge.

Evaluating Localization Effects through Analysis of Nonpatent Citations

The diffusion of knowledge is measured in terms of the geographical distribution of spillovers or by surveying firms on the relative importance of various sources of knowledge. Codified knowledge is diffused primarily through publications; so evaluating the spillover effect (in terms of geographic distance) of a piece of codified knowledge involves analysis of patent citations (Arundel and Geuna 2004). If the spillover is localized (i.e., users of the knowledge are located proximate to the source

20 Over time, the science system has established an institution that transfers codified knowledge in a systematic way.

of the knowledge), one can infer that tacit knowledge served as an important source of information (Anselin, Varga, and Acs 1997, 2000). By contrast, firms that rely more on codified (versus tacit) knowledge would have little incentive to locate close to the source of the information. In his seminal work, Jaffe (1989) found that patents are most likely to be granted in research-intensive states (i.e., those states with the largest amounts of funding for university research). Through an analysis of patent citations, Jaffe, Trajtenberg, and Henderson (1993) found that patents from a particular state or statistical metropolitan area are more likely to cite existing patents or published findings produced locally.

Sorenson and Fleming (2004) analyzed 16,728 U.S. utility patent citations to compare the diffusion of knowledge from published, versus unpublished, materials. As one might expect, the researchers found that published articles diffuse further and faster than those that are not published, which supports the view that open science makes an important contribution to the spillover effect. Similarly, the diffusion of knowledge through publication is more prominent than the diffusion through social networks (Sorenson and Singh 2007), which further affirms the importance of the norms of open science.

Results of Surveys of Innovative Firms

Most often, surveys of firms simply ask which sources of knowledge are the most important to the firms' innovative capability. For example, in the case of the PACE Survey, firms are given a set of different information sources (e.g., publications, conferences, interpersonal interactions, trained scientists) and asked to rank them, using the Likert Scale, in terms of importance. Following the discovery that knowledge spillover tends to be localized, the economics literature has sought to determine why. The explanation cited most often is that firms are more effective at exploiting codified knowledge if they have direct access to the local producers of that knowledge.

Survey results provide mixed evidence on the importance to industry of various sources of knowledge. Malo (2009) conducted a global survey of top R&D managers who specialize in combinatorial chemistry and asked them to rank the importance of different sources of basic scientific knowledge. New firms ranked publications as the most important, ahead of research contracts and conferences. Large established firms ranked hiring, patents, and publications as most important, in that order.

For external sources of information, public research is ranked as the most important relative to competitors, R&D firms, and joint ventures, according to a random draw of the 1994 Carnegie Mellon Survey results (Cohen et al. 2002). Of different diffusion channels, the survey ranked publications, conferences, and informal contacts as the most important public research sources. Furthermore, it was found that more-codified sources of public research, such as publications and conferences, contributed more to the completion of existing R&D projects. Tacit sources, such as informal contacts and recent hires, were positively correlated to the introduction of new R&D projects. The 1993 PACE Survey of Europe's largest industrial firms shows that external sources of tacit knowledge are ranked higher

than external sources of codified knowledge. These sources include hiring trained scientists and engineers, having informal personal contacts, and pursuing research contracting (Arundel and Geuna 2004).

Conclusion

This article summarizes the current economics literature on the different mechanisms through which the science system affects the economy. The science system produces new knowledge upon which innovation builds, creates a skilled scientific labor force that will continue to advance the current state of knowledge, and helps firms/entrepreneurs create innovation. Further, the system diffuses knowledge quickly and widely, generating the many spillovers that contribute to economic growth. Ten conclusions arise from this analysis:

> According to knowledge-based endogenous growth models (e.g., those that use science, R&D, or human capital as an input for production), perpetual economic growth can result exclusively from the positive spillovers created by cumulative knowledge, regardless of advances in exogenous technical progress. Advancement in knowledge affects production in two ways: by reducing the cost of future research and by increasing the productivity of other inputs (e.g., researchers, engineers, intermediate goods). Consequently, an economy that has a continuously growing knowledge base will enjoy ever-increasing GDP per capita growth over the long run.

1. Scientific knowledge is either codified (explicit) or tacit (implicit). Codified knowledge is that which is expressed in a compact and standardized format, such as in publications and presentations (Cowan, David, and Foray 2000). Tacit knowledge, by contrast, is "person embodied" and develops over time through experiences (Polanyi 1958). These two forms of knowledge are highly complementary, and most scholars agree that innovators trying to exploit knowledge require both.

2. Views diverge on the characteristics of codified knowledge, and this divergence has important growth and policy implications. First is the proposition that knowledge is a public good that creates the positive spillovers necessary for economic growth (Arrow 1971). Accordingly, profit-maximizing agents will lack the incentive to invest in science because they (i) cannot appropriate the full returns of their research findings and (ii) they can freely capitalize on research conducted—and funded—by others. This provides a rationale for government-funded science. Others argue that knowledge is not a public good—although access to knowledge may be costless, understanding and exploiting

it is not. In this view, science is not characterized by large spillovers, and the rationale for government-funded science is thus weakened.

3. Rosenberg (1990) summarizes several alternative incentives that explain why a firm will invest in science, even if it is a public good. First, large firms are not concerned about capturing all of the returns generated by their own research. Indeed, all that matters is that their returns cover their costs and generate a profit. Second, in-house research teams allow a firm to operate on the frontiers of knowledge, and this can point them in the direction of exploitable commercial opportunities. Third, an in-house science team is capable of understanding—and exploiting—external sources of knowledge.

4. Econometric studies that attempt to quantify the impact of science on the economy are fraught with difficulties, and, as a result, few if any such studies exist. However, a few econometric studies that calculate a rate of return on firm-level basic research investment show that there is a large premium on basic research over applied research and development. That is, the rate of return on one dollar spent on basic research is much higher than the rate of return if the same dollar is spent on applied research (Lichtenberg and Siegel 1991; Griliches 1979).

5. Publicly funded basic research is shown to complement firm-level basic research. Thus, a one-dollar investment in basic research by the government creates additional investment in basic research by private firms.

6. An often-used methodology to date that relates the importance of science to innovation involves nonpatent citation studies, which indicate that innovation has increasingly relied on science. Predictably, science improves productivity for particular sets of technology sectors. The important implication is that there should not be a one-size-fits-all policy governing science, technology, and innovation. In particular, for those industries characterized by science intensity, policy should be directed toward more collaboration between the science system and industry, which would accomplish quicker diffusion of knowledge between firms and the science system.

7. Surveys provide mixed results on the important sources of innovation ideas. High-technology firms most often rely on scientific knowledge, while less high-tech firms find that input from the vertical chain of production is the best source of insight into new commercial opportunities. High-, medium-, and low-technology firms all report that one of the most important benefits of the science system is the educated scientific work force that the system creates.

8. University-industry collaboration is becoming increasingly important for innovation. In economies across the globe, university and industry scientists are collaborating much more frequently on both research (as evidenced by co-authorship on scientific articles) and on innovation (as evidenced by co-inventors on patent applications). Science and technology policies are beginning to focus on ways to motivate such collaboration within one's own country and with scientists from other countries.

9. The science system diffuses information in the forms of codified and tacit knowledge; but to some extent, the two sources are inextricably linked. Indeed, evidence consistently suggests that those who are most likely to exploit research findings published in a journal (codified knowledge) also live and work geographically proximate to where the knowledge was produced. The explanation is that, for firms to make full use of codified knowledge published from region A, they must also rely on the tacit knowledge of the publishing researcher to fully grasp the science. This is why innovative firms often locate near universities or government research laboratories.

References

Acosta, M., and D. Coronado. 2003. Science-technology flows in Spanish regions: an analysis of scientific citations in patents. *Res. Pol'y* 32(10): 1783-1803.

Aghion, P., and P. Howitt.. 1992. A model of growth through creative destruction. *Econometrica* 60(2): 323-51.

Anselin, L., A. Varga, and Z. Acs. 2000. Geographic and sectoral characteristics of academic knowledge externalities. *Papers in Regional Sci.* 79: 435-48.

Anselin, L., A. Varga, and Z. Acs. 1997. Local geographic spillovers between university research and high technology innovations. *J. Urb. Econ.* 42: 422-48.

Arrow, K. 1971. Economic welfare and the allocation of resources for R&D, in *Essays in the theory of risk bearing*. Chicago: Markham Publishing Company.

Arundel, A., and A. Geuna. 2004. Proximity and the use of public science by innovative European firms. *Econ. Innovation & New Tech.* 13(6): 559-80.

Arundel, A., G. van de Paal, and L. Soete. 1995. PACE Report: Innovation strategies of Europe's largest firms: Results of the PACE survey for information sources, public research, protection of innovations, and government programmes. Maastrict: MERIT, University of Limberg.

Audretsch, D., and P. Stephan. 1996. Company-scientist locational links: the case of biotechnology. *Am. Econ. Rev.* 86(3): 641-52.

Beise, M., and H. Stahl. 1999. Public research and industrial innovations in Germany. *Res. Pol'y* 28(4), 397-422.

Bhagwati, J., A. Panagariya, and T. Srinivasan. 2004. The muddles over outsourcing. *J. Econ. Persp.* 18(4): 93-114.

Branstetter, L. 2001. Are knowledge spillovers international or intranational in scope?: Microeconometric evidence from the U. S. and Japan. *J. Int'l Econ.* 53(1): 53-79.

Brökel, T., and M. Binder. 2007. The regional dimension of knowledge transfers—a behavioral approach. *Industry & Innovation* 14(2): 151-75.

Brooks, H. 1994. The relationship between science and technology. *Res. Pol'y* 23: 477-86.

Callon, M. 1994. Is science a public good? fifth Mullins lecture, Virginia Polytechnic Institute, 23 March 1993. *Sci., Tech., & Hum. Values* 19(4): 395-424.

Cohen, W. M., R. R. Nelson,and J. P. Walsh. 2002. Links and impacts: The influence of public research on industrial R&D. *Mgmt. Sci.* 48(1), 1-23.

Cohen, W. M., and D. A. Levinthal. 1989. Innovation and learning: the two faces of R&D. *Econ. J.* 99: 569-96.

Cowan, R., P. A. David, and D. Foray. 2000. The explicit economics of knowledge codification and tacitness. *Indus. & Corp. Change* 9(2): 211-53.

Dasgupta, P., and P. A. David. 1994. Toward a new economics of science. *Res. Pol'y* 23(5): 487-521.

David, P. A. 2004. Understanding the emergence of 'open science' institutions: functionalist economics in historical context. *Indus. & Corp. Change* 13(4): 571-89.

David, P. A., B. H. Hall, and A. A. Toole. 2000. Is public R&D a complement or substitute for private R&D? A review of the econometric evidence. *Res. Pol'y* 29(4-5): 497-529.

Diamond, A. M. 1997. Does federal funding "crowd in" private funding of science? *Contemp. Econ. Pol'y* 17(4): 423-31.

Fabrizio, K. R. 2009. Absorptive capacity and the search for innovation. *Res. Pol'y* 38: 255-67.

Feenstra, R. C. 1996. Trade and uneven growth. *J. Dev. Econ.* 49(1): 229-56.

Freeman, C. 1987 *Technology policy and economic performance*. London: Pinter.

Gibbons, M., and R. Johnston. 1974. The roles of science in technological innovation. *Res. Pol'y* 3(3): 220-42.

Griliches, Z. 1979. Issues in assessing the contribution of research and development to productivity growth. *Bell J. Econ.* 10(1):, 92-116.

Griliches, Z. 1986. Productivity, R&D, and basic research at the firm level in the 1970's. *Am. Econ. Rev.* 76(1): 141-54.

Grossman, G. M., and E. Helpman. 1994. Endogenous innovation in the theory of growth. *J. Econ. Persp.* 8(1): 23-44.

Grossman, G., and E. Helpman. 1991. *Innovation and growth in the global economy*. Cambridge: MIT Press.

Hall, B. H., A. B. Jaffe, and M. Trajtenberg. 2001. The NBER patent citation data file: lessons, insights and methodological tools. National Bureau of Economic Research (NBER) Working Paper No. 8498, *available at* http://www.nber.org/papers/w8498.

Hall, B. H., J. Mairesse, and P. Mohnen, P. 2009. Measuring the returns to R&D. NBER Working Paper No. 15622, *available at* http://www.nber.org/papers/w15622.

Jaffe, A. B. 1989. Real effects of academic research. *Am. Econ. Rev.* 79(5): 957-70.

Jaffe, A. B., M. Trajtenberg, and R. Henderson. 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. *Q. J. Econ.* 108(3): 577-98.

Kealey, T., and O. Al-Ubaydli. 2001. A critique of science and R&D-based models of endogenous growth. *Knowledge, Tech. & Pol'y* 13(4): 37-48.

Keller, W. 2004. International technology tiffusion. *J. Econ. Literature* 42(3): 752-82.

Klevorick, A. K., R. C. Levin, R. R. Nelson, and S. G. Winter. 1995. On the sources and significance of interindustry differences in technological opportunities. *Res. Pol'y* 24(2):185-205.

Kline, S. J., and N. Rosenberg. 1986.. An overview of innovation. In R. Landau and N. Rosenberg, eds., *The positive sum strategy: Harnessing technology for economic growth.* Washington, D.C.: National Academy Press, 275-305.

Lichtenberg, F. R., and D. Siegel. 1991. The impact of R&D investment on productivity—new evidence using linked R&D-LRD data. *Econ. Inquiry* 29(2): 203-29.

Link, A. N. 1981. Basic research and productivity increase in manufacturing: Additional evidence. *Am. Econ. Rev.* 71(5): 1111-12.

Link, A., and J. Rees. 1990. Firm size, university based research, and the returns to R&D. *Small Bus. Econ.* 2(1): 25-31.

Lööf, H., and A. Broström. 2008. Does knowledge diffusion between university and industry increase innovativeness? *J. Tech. Transfer* 33(1): 73-90.

Lundvall, B., ed. 1992. *National systems of innovation: Towards a theory of innovation and interactive learning*. London: Pinter.

Malo, S. 2009. The contribution of (not so) public research to commercial innovations in the field of combinatorial chemistry. *Res. Pol'y* 38(6): 957-70.

Mankiw, N. G. 1995. The growth of nations. *Brookings Papers on Econ. Activity* 1995(1): 275-326, *available at* http://www.jstor.org/stable/2534576.

Mansfield, E. 1991. Academic research and industrial innovation. *Res. Pol'y* 20(1): 1-12.

Mansfield, E. 1998. Academic research and industrial innovation: An update of empirical findings. *Res. Pol'y* 26(7-8): 773-76.

Mansfield, E. 1980. Basic research and productivity increase in manufacturing. *Am. Econ. Rev.* 70(5): 863-73.

Mansfield, E., and J. Lee. 1996. The modern university: Contributor to industrial innovation and recipient of industrial R&D support. *Res. Pol'y* 25(7): 1047-58.

Maurseth, P. B., and B. Verspagen. 1998. Knowledge spillovers in Europe and its consequences for systems of innovation. ECIS Working Paper No. 98.1. Eindhoven, The Netherlands: Eindhoven Centre for Innovation Studies, *available at* http://cms.tm.tue.nl/Ecis/Files/ papers/wp1998/ eciswp1.pdf

McMillan, G. S., F. Narin, and D. Deeds. 2000. An analysis of the critical role of public science in innovation: The case of biotechnology. *Res. Pol'y* 29:(1): 1-8.

Merton, R. K. 1942. Science and technology in a democratic order. *J. Legal & Pol. Soc.* 1: 115-26.

Monjon, S., and P. Waelbroeck. 2003. The nature of innovation and the origin of technological spillovers: An econometric analysis on individual French data. *Brussels Econ. Rev.* 46(3): 87-106, *abstract available at* http:// econpapers.repec.org/article/bxrbxrceb/ y_3a2003_3av_3a46_3ai_3a3_ 3ap_3a87-106.htm.

Narin, F., K. S. Hamilton, and D. Olivastro. 1997. The increasing linkage between U.S. technology and public science. *Res. Pol'y* 26(3): 317-30.

National Science Board. 2010. *Science and engineering indicators: 2010*. Arlington, Va.: National Science Foundation.

Nelson, R. R., ed. 1993. *National innovation systems: A comparative analysis*. New York: Oxford University Press.

Nelson, R. R. 2004. The market economy and the scientific commons. *Res. Pol'y* 33(3): 455-71.

Nelson, R. R. 1959. The simple economics of basic scientific research. *J. Pol. Econ.* 67(3): 297-306.

Nelson, R. R. and R. Levin. 1986. The influence of science, university research and technical societies on industrial R&D and technical advance. Policy Discussion Paper, Series No. 3, Research Program on Technology Change. Yale University, New Haven, Conn.

Nelson, R. R., and N. Rosenberg. 1993. Technical innovation and national systems. In R. R. Nelson, ed. *National innovation systems: A comparative analysis*. New York: Oxford University Press, 3-21.

Organization for Economic Co-operation and Development (OECD). (1996). The knowledge-based economy. OECD Working Papers. Paris: OECD, available at http://www.oecd.org/dataoecd/51/8/1913021.pdf.

OECD. 1997. National innovation systems. OECD Working Papers. Paris: OECD, available at http://www.oecd.org/dataoecd/35/56/2101733.pdf.

OECD. 2005. OSLO manual: Guidelines for collecting and interpreting innovation data. 3d ed. Paris: OECD, *abstract available at* http://www.

oecd.org/document/23/0,3746, en_2649_34409_35595607_1_1_1_1, 00.html.

Polanyi, M. 1958. *Personal knowledge: Towards a post-critical philosophy*. London: Routledge & Kegan Paul Ltd.

Polanyi, M. 1966. The logic of tacit inference. *Philosophy* 41(155): 1-18.

Roberts, J. 2000. From know-how to show-how? questioning the role of information and communication technologies in knowledge transfer. *Tech. Analysis & Strategic Mgmt*. 12(4): 429-43.

Robson, M. T. 1993. Federal funding and the level of private expenditures on basic research. *S. Econ. J.* 60(1): 63-71.

Romer, P. M. 1990. Endogenous technological change. *J. Pol. Econ*. 98(5 pt. 2): S71-S102.

Rosenberg, N. 1982. How exogenous is science? In N. Rosenberg, *Inside the black box: Technology and economics*. New York: Cambridge University Press, 141-62.

Rosenberg, N. 1990. Why do firms do basic research (with their own money)? *Res. Pol'y* 19(2): 165-74.

Salter, A. J., and B. R. Martin. 2001. The economic benefits of publicly funded basic research: a critical eview. *Res. Pol'y* 30(3): 509-32.

Samuelson, P. A. 2004. Where Ricardo and Mill rebut and confirm arguments of mainstream economists supporting globalization. *J. Econ. Persp.* 18(3):135-46.

Saxenian, A. 1994. *Regional advantage: Culture and competition in silicon valley and route 128*. Cambridge, Mass.: Harvard University Press.

Schmookler, J. 1966. *Invention and economic growth*. Cambridge, Mass.: Harvard University Press.

Singh, J. 2005. Collaborative networks as determinants of knowledge diffusion patterns. *Mgmt. Sci.* 51(5): 756-70.

Sorenson, O., and L. Fleming. 2004. Science and the diffusion of knowledge. *Res. Pol'y* 33(10): 1615-34.

Sorenson, O., and J. Singh, J. 2007. Science, social networks and spillovers. *Industry & Innovation* 14(2): 219-38.

Stokes, D. E. 1997. *Pasteur's Qquadrant: Basic science and technological innovation*. Washington D.C.: Brookings Institution Press.

Tassey, G. 2010. Rationales and mechanisms for revitalizing US manufacturing R&D strategies. *J. Tech. Transfer* 35(3): 283-333.

Tijssen, R. J. W. 2002. Science dependence of technologies: evidence from inventions and their inventors. *Res. Pol'y* 31(4): 509-26.

van Looy, B., E. Zimmermann, R. Veuglars, A. Verbeek, J. Mello,and K. Debackere. 2003. Do science-technology interactions pay off when developing technology? an exploratory investigation of 10 science-intensive technology domains. *Scientometrics* 57(3): 355-67.

von Hippel, E. 1978. A customer-active paradigm for industrial product idea generation. *Res. Pol'y* 7(3): 240-66.

Zahra, S. A., and G. George. 2002. Absorptive capacity: a review, reconceptualization, and extension. *Acad. Mgmt. Rev.* 27(2): 185-203.

Zellner, C. 2003. The economic effects of basic research: evidence for embodied knowledge transfer via scientists' migration. *Res. Pol'y* 32(10): 1881-95.

Zucker, L. G., M. R. Darby, J. Furner, R. C. Liu,and H. Ma. 2007. Minerva unbound: knowledge stocks, knowledge flows and new knowledge production. *Res. Pol'y* 36(6): 850-63.

Zucker, L., M. R. Darby, and J. S. Armstrong. 2002. Commercializing knowledge: university science, knowledge capture, and firm performance in biotechnology. *Mgmt. Sci.* 48(1): 138-53.

Student Article

Explosive Trace Detection Searches in Airports

Benjamin Signer, The University of Tennessee–Knoxville

Introduction[1]

If you are a frequent air traveler, the dance is old hat to you by now. You get to the airport at least an hour early. Check in at the desk if you have not done so online. Check a bag? No, thank you. At thirty dollars a bag, you will just use your carry-on. Then you join the queue at security. Eventually, a Transportation Security Administration (TSA) officer scans your identification and boarding pass, and you take off your shoes in anticipation of the x-ray machine. You submit everything metal to the x-ray conveyor (laptop in a separate case, of course) and shamble barefoot and beltless through the metal detector, praying that you will pass the officer's disinterested scrutiny. If all goes well, you collect your belongings, re-dress, and head to your gate.

Since the attempted Christmas Day bombing on the flight from Amsterdam to Detroit, though, the TSA has added another move to the dance.[2] In an attempt to detect explosives on a person that are not detectable by metal detectors, officers have begun swabbing passengers' hands for traces of explosive material while they wait in line before being scanned by the metal detectors. By conducting the tests on passengers prior to their reaching the metal detectors, the TSA hopes to increase safety without increasing wait times at security checkpoints. By conducting the search before a passenger passes through the metal detector, however, security personnel are treading onto a legal issue that has not been fully resolved.

The legality of searching passengers by having them pass through a metal detector is well established. Courts have long styled the metal detector as the symbolic portal that, once passed through, lowers a person's expectation of privacy, but hand swabbing of passengers occurs before they cross this point of lowered expectations. How will courts respond to this? Is this an acceptable incremental decrease in our already lowered expectation of privacy at airports or an impermissible intrusion on what remains of our privacy expectations?

This article discusses the technology behind the new hand-swab searches as well as some alternative airport security methods. It then covers exceptions to

1 In connection with the preparation of this article, I would like to thank Scott Broyles for his invaluable technical assistance and my father, Greg Signer, for his editorial assistance.

2 Doug Guthrie and George Hunter, *Suspected 'underwear bomber' arraigned on terror conspiracy charges*, The Detroit News, Dec. 17, 2010, *available at* http://www.detnews.com/article/20101217/METRO/12170394/1409/ metro/Suspected-%E2%80%98underwear-bomber%E2%80%99-arraigned-on-terror-conspiracy-charges.

the Fourth Amendment warrant requirement, existing airport search law, and extrapolates the likely result of a court case in which the legality of an airport search is at issue.

Technical Information

Airport security was initiated primarily to keep passengers from carrying weapons onto a plane. However, the 1960s hijacker intent on diverting planes and passengers to make a political statement has evolved into today's terrorist willing to blow up planes and passengers (and often him or herself) to make a political statement. Preventing the smuggling of explosive devices onto airplanes has become as important as, if not more important than, keeping weapons off planes, especially since airplane cockpits have been secured. In response, the technology of explosive detection also has evolved. Today, there are many methods of explosive detection, each with its own pros and cons.

The Explosive Trace Detector

The TSA is currently using a piece of technology referred to as the Explosive Trace Detector ("ETD") to test the hands of some passengers. The ETD involves the acquisition and analysis of microscopic traces of explosive compounds.[3] According to Scott Broyles, an official with the Safe Skies Alliance, "The ETD is a tried and true system which has been used in aviation security since prior to TSA's inception. It is considered to be very accurate and dependable." The ETD is normally used for screening carry-on luggage at security checkpoints and sometimes for screening luggage to be placed in the cargo compartment of passenger aircraft.[4]

The TSA has begun to deploy ETDs randomly to screen passengers' hands while they wait to pass through metal detectors as part of its effort to make security methods less predictable and, it believes, unavoidable. "There seems to be no doubt regarding the technology's ability to perform reliably and accurately. I am not aware of another technology or search method that could serve as an alternative while providing the same accurate results," says Broyles.[5]

Organic Methods of Explosive Detection

A crucial part of a court's analysis of a search method is its effectiveness and invasiveness compared to other search methods. The method with which the general public is probably most familiar is the use of specially trained canines. A dog's olfactory bulb is roughly forty times larger than a human's and has around

3 Lisa Thiesan et al., Survey of Commercially Available Explosives Detection Technologies and Equipment 2004, a research report prepared for the National Law Enforcement and Correction Technology Center, a Program of the National Institute of Justice, U.S. Department of Justice (2004), at 24, *available at* https://www.ncjrs.gov/pdffiles1/nij/grants/208861.pdf.

4 E-mail interview by the author with Scott Broyles, Chief Operating Officer, National Safe Skies Alliance (Jan. 2011).

5 *Id*.

200 million smell receptors.[6] Dogs can detect scents in concentrations 100 million times lower than those that are detectible by humans.[7] There are downsides to using trained dogs, though. Dogs can cause a disturbance when they interact with small children or cause discomfort for passengers and employees who may be allergic to or who may fear dogs. Like people, dogs are also capable of becoming tired or bored, which decreases their effectiveness.

Some biotech scientists are reportedly working on a technique for detecting explosives with honeybees; but this system is still in development, and little is known about it. Bees reportedly are capable of working for days before they have to be returned to their hive.[8] Researchers at Colorado State University are also working on breeding bomb-detecting plants. By manipulating the plants' DNA, they can cause the plants to change color in the presence of certain explosive trace chemicals or pollutants. However, this formidable flora is still three or four years away from deployment.[9]

Machine Olfaction

The method of machine olfaction detection, sometimes called an "electronic nose," works much like an organic nose. Scent molecules strike a sensor that generates a signal that is analyzed by a computer. In an electronic nose, the computer is an actual computer running a series of complex algorithms, while in an actual nose, a brain acts as the computer. These devices can detect a single chemical with certainty. The downside is that the more components a substance has, the less effective an electronic nose is at detecting them.[10] So an electronic nose could identify the sulphur fumes from rotten eggs but would struggle with the sublime and complex aroma of my award-winning chili. This means that a machine olfaction system may be relatively ineffective at detecting even a simple fertilizer bomb, which is a binary explosive.

Chemiluminescence

Chemiluminescence is the creation of light resulting from certain chemical reactions. Most explosives contain nitrogen. Nitrogen molecules can be excited to emit infrared radiation. The amount of emitted infrared radiation is directly proportional to the amount of the original explosive material that was present. Chemiluminescence systems are small, some weighing as little as seven pounds, but can cost up to $60,000 for each detector. A significant disadvantage of the

6 Stanley Coren, How Dogs Think (2004).

7 Natural History Museum of Los Angeles County, *Smell*, http://web.archive.org/web/20080801101136/http://www.nhm.org/ exhibitions/dogs/formfunction/smell.html (last visited Jan. 20, 2011).

8 British Broadcasting Corporation, Hot picks: UK tech start-ups, http://newsvote.bbc.co.uk/mpapps/pagetools/ print/news.bbc.co.uk/1/hi/technology/6972526.stm (last visited Jan. 21, 2011).

9 John Roach, Bomb sniffing plants to the rescue, MSNBC, Jan. 27, 2011, http://cosmiclog.msnbc.msn.com/_news/ 2011/01/27/5936102-bomb-sniffing-plants-to-the-rescue (last visited Jan. 20, 2011).

10 Larry Senesac and Thomas G. Thundat, *Nanosensors for trace explosive detection*, 11 Materials Today 28, 31 (2008).

chemiluminescence system is that it is unable to detect a non-nitrogen based explosive.[11]

Mass Spectrometry

Mass spectrometry works by ionizing molecules and passing them through a filter, which allows them to be identified based on their charge-to-mass ratio. All molecules have different weights and fragmentation patterns that makes them readily identifiable. Mass spectrometry is extremely reliable in identifying an unknown substance. Unfortunately, mass spectrometry units are neither very portable (the smallest versions weighing in at around seventy-five pounds) nor cheap (the cheapest versions costing around $135,000). The units also take a relatively long time to analyze a sample.[12]

Optical Dynamic Detection

Optical dynamic detection systems are currently under development by Princeton University and the U. S. Department of Energy under the direction of the U. S. Department of Homeland Security. Optical dynamic detection works by using a laser to excite the electrons in molecules in a suspicious package. A second laser then passes through the package analyzing its spectrum. One advantage of the optical dynamic detection system is that it is able to identify where an explosive is in a package. This system is not ready for use yet, however, and one of its drawbacks is that can only work on packages, not people.[13]

The Law

The Fourth Amendment Generally

The Fourth Amendment to the U.S. Constitution prevents unreasonable searches and seizures by the government. In most circumstances, this means that a judge must pre-approve a search and issue a search warrant based upon a showing of probable cause that the target of the requested search has or is engaged in criminal conduct and that the search will result in the discovery of evidence of such criminal wrongdoing. The purpose of a warrant is to prevent the second-guessing of the reasonableness of a search and to allow an impartial judge to substitute his or her judgment for the searching officer's.[14]

A search is "ordinarily unreasonable in the absence of an individualized suspicion of wrongdoing."[15] There are certain situations, however, in which a search is still reasonable and lawful, even without prior judicial approval based upon probable cause. Perhaps the best-known example is a search incident to a lawful arrest.

11 Thiesan, *supra* note 3, at 33.

12 *Id.* at 39.

13 Tudor Vieru, *New Explosive Detection Method Created*, Softpedia, April 9, 2010.

14 U.S. v. Martinez-Fuerte, 428 U.S. 543, 566 (1976).

15 City of Indianapolis v. Edmond, 531 U.S. 32, 37 (2000).

Anyone who has ever seen a suspect arrested either on television or in real life has seen the cops do a quick frisk and search of the detainee's person. This search is allowed because courts have readily accepted that a police officer has the right to ensure his or her own safety by searching detained persons.[16] Guards are also free to require identification and search people entering federal office buildings to help protect the building.[17] Police are also allowed to set up road checkpoints to check for intoxicated drivers to help protect the motoring public.[18] These exceptions to the warrant requirement fall into two main categories, the administrative search exception and the special needs search exception.

The Administrative Search Exception

An administrative search exception is generally used to allow government to conduct inspections of a commercial entity in a well-regulated industry. Applying this exception requires a finding of three criteria:

> "First, there must be a 'substantial' government interest . . . ." "Second, the warrantless inspections must be 'necessary to further [the] regulatory scheme.'" Third, the regulatory statute authorizing administrative inspections "must perform the two basic functions of a warrant: it must advise the owner of the commercial premises that the search is being made pursuant to the law and has a properly defined scope, and it must limit the discretion of the inspecting officers."[19]

As can be seen from the wording of the criteria, the administrative search exception was originally contemplated to deal with regulatory searches. For example, the county health inspector does not need to apply for a warrant to search a restaurant located within his jurisdiction. Even though administrative regulations may have the same purpose as penal laws, they do not lose their validity "simply because the government has chosen to 'address a major social problem *both* by way of an administrative scheme *and* through penal sanctions.'"[20]

In the 1960s, the hijacking of commercial airliners started occurring largely as a means to bring attention to the hijackers' political agenda, although a few hijackers were merely run-of-the-mill criminals. The federal government moved quickly to secure the nation's burgeoning passenger air industry by installing magnetometers to screen passengers before they could board an airplane and to inspect containers

16 Terry v. Ohio, 392 U.S. 1 (1968).

17 Dickerson v. Napolitano, 604 F.3d 732, 750-51 (2d Cir. 2010).

18 Martinez-Fuerte, *supra* note 14.

19 Kyle P. Hanson, *Suspicionless Terrorism Checkpoints Since 9/11: Searching for Uniformity*, 56 Drake L. Rev. 171, 178-79 (2007).

20 *Id.* (emphasis in original).

at airports.[21] Lacking any other schema for interpreting the new searches, courts evaluated them under the administrative search exception.[22] "Given the high potential for harm to people and property, airport searches were found to be reasonable as long as people could avoid the searches by choosing not to fly, thus reducing the intrusion of the screenings to a minimal level."[23]

The Special Needs Search Exception

Over time, though, the courts have articulated another exception to the warrant requirement—the special needs exception:

> [W]here a Fourth Amendment intrusion serves special governmental needs, beyond the normal need for law enforcement, it is necessary to balance the individual's privacy expectations against the Government's interests to determine whether it is impractical to require a warrant or some level of individualized suspicion in the particular context.[24]

There is an additional step to the analysis in that the search must be reasonably effective at advancing the special governmental need.[25]

What is a special governmental need beyond the normal need for law enforcement? Generally, in deciding whether the special needs exception should apply, courts must first examine "whether the search '"serve[s] as [its] immediate purpose an objective distinct from the ordinary evidence gathering associated with crime investigation."'[26] The next step is to determine the reasonableness of the search. To determine whether the search was reasonable, the courts balance several competing considerations, including "(1) the weight and immediacy of the government interest, (2) 'the nature of the privacy interest allegedly compromised by' the search, (3) 'the character of the intrusion imposed' by the search, and (4) the efficacy of the search in advancing the government interest."[27]

In *U.S. v. Martinez-Fuerte*, the special governmental need was to secure the border against illegal immigrants.[28] The Border Patrol established checkpoints on a California highway near the Mexican border. Agents slowed all passing cars for a quick visual inspection and diverted some to another area for a secondary search.

---

21 Scott McCartney, *The Golden Age ofFlight*, WALL ST. J., July 22, 2010, *available at* http://online.wsj.com/article/SB10001424052748704684604575380992283473182.html?mod=googlenews_wsj.

22 Hanson, *supra* note 19, at 179.

23 *Id.* at 179-80.

24 Nat'l Treasury Employees Union v. Von Raab, 489 U.S. 656, 665-66 (1989).

25 *See* Mich. Dep't of State Police v. Sitz, 496 U.S. 444, 454-55 (1990).

26 MacWadea v. Kelly, 460 F.3d 260, 268 (2d Cir. 2006).

27 *Id.* at 269 (citations omitted).

28 Martinez-Fuerte, *supra* note 14.

The court upheld the searches based on the need to interdict the flow of illegal immigrants and the relatively small intrusion that the drivers experienced.[29]

However, the special need exception can, at times, straddle a fine line. In *Michigan Department of State Police v. Sitz*,[30] the court upheld checkpoint searches for intoxicated drivers, while in *City of Indianapolis v. Edmond*,[31] a checkpoint search for illegal drugs was deemed improper. The distinction between the two searches made by the courts was that the government has a need to protect drivers from their intoxicated peers that goes beyond the enforcement of criminal laws, whereas searching cars for drugs can only serve the purpose of advancing a criminal investigation.[32] If Indianapolis had only searched the cars of drivers who showed the effects of drug use, the city might also have prevailed in its case.

Though the special needs exception is currently in vogue for explaining the permissibility of antiterrorism checkpoint searches, not everyone agreesthat the exception is warranted. According to Ric Simmons, "the special needs doctrine does not provide much in the way of principled guidance to courts struggling to evaluate antiterrorism searches. . . . [T]he Supreme Court has . . . struggle[d] to find a principled distinction between searches for special needs and searches for a law enforcement purpose."[33] Simmons points to a growing circuit split among the federal circuit courts of appeals concerning the exception. The Eleventh Circuit has held that antiterrorism searches do not serve a special need. The Second Circuit has held the opposite, perhaps giving greater weight to the need, given thelocation of the court in New York City. Other circuits require a terrorist threat to reach some level of specificity before concluding that antiterrorism searches are justified.[34]

Airport Searches in Particular

Until the security changes brought about by the terrorist attacks on September 11, 2001, airport searches consisted almost entirely of magnetometer scans of passengers and x-ray scans of their luggage. Scanning a passenger with a metal detector or their luggage with an x-ray device is a search within the meaning of the Fourth Amendment. Requiring a warrant for such searches, though, would "exalt form over substance." The Fourth Circuit noted in *U.S. v. Epperson* that no judge would deny a warrant for such a search, given the overwhelming governmental interest in air travel safety and the corresponding minimal invasion of passengers' privacy.[35]

---

29 *Id.* at 552.

30 Sitz, *supra* note 25.

31 Edmond, *supra* note 15.

32 Sitz, *supra* note 25, at 451-52; Edmond, supra note 15, at 453-454.

33 Ric Simmons, *Searching for Terrorists: Why Public Safety is Not a Special Need*, 59 Duke L. J. 843, 884 (2010).

34 *Id.*

35 U.S. v. Epperson, 454 F.2d 769, 771 (4th Cir. 1972).

The *Epperson* court articulated a two-step test for determining the constitutionality of an airport search (abandoning the much more complex flamenco test). The search must first be justified at its inception.[36] This is where the administrative or special needs exception comes into play. Though given the prevailing winds of jurisprudence, the special needs exception seems likely to provide the basis these days. The next step is to determine whether the search is "reasonably related in scope to the circumstances which justified interference in the first place."[37] In *Epperson*, a search solely to discover weapons to prevent air piracy was fully justified given the relatively unobtrusive nature of passing through a metal detector. The court reasoned that, "The use of the [magnetometer], . . . unlike frisking, cannot possibly be 'an annoying, frightening, and perhaps humiliating experience,' . . . because the person scrutinized is not even aware of the examination."[38] As the Second Circuit noted in *U.S. v. Edwards*, "More than a million Americans subject themselves to [magnetometer searches] daily; all but a handful do so cheerfully, even eagerly, knowing it is essential for their protection."[39]

Subsequent searches based on the findings from the magnetometer may or may not be valid, depending on the procedure that the screening officers utilize. If a passenger triggers an alert from a metal detector, he or she is not automatically subject to a frisking. The screening officers must first give the passenger a chance to remove whatever metal might have been forgotten on his or her person and to pass through the detector again. Security agents must exhaust other efficient and unobtrusive means before a more invasive search may be undertaken.[40] For instance, merely observing a suspicious bulge in a passenger's jacket is not enough to justify a search.[41] Only after a passenger is unable to explain his or her continued triggering of a magnetometer is the frisking a valid search.[42] Once a search is justified, though, anything discovered by an officer conducting such a search is fair game regardless of whether it was what the officer was searching for.[43]

Some courts consider a consent theory when deciding the validity of an airport search. In *People v. Kuhn*,[44] a New York court held that a passenger's consent was established by the absence of coercive tactics by the security officers, the absence of any official compulsion in the manner of the security officers in requesting

36 *Id.*

37 *Id.*

38 *Id.*; James L. Buchwalter, *Validity of Airport Security Measures*, 125 ALR 5th 281 (2005).

39 498 F.2d 496, 500 (2d Cir. 1974).

40 U.S. v. Albarado, 495 F.2d 799, 808 (2d Cir. 1974).

41 People v. Erdman, 69 Misc. 2d 103 (N.Y. 1972).

42 Epperson, *supra* note 35 at 772.

43 *Id.*

44 306 N.E.2d 777 (N.Y. 1973).

permission to search the passengers, and the fact that passengers had not been chosen arbitrarily for screening.

However, as the Ninth Circuit has pointed out recently in *U.S. v. Aukai*,[45] the validity of an airport search does not depend on consent. "[W]here an airport screening search is otherwise reasonable . . . all that is required is the passenger's election to attempt entry into the secured area of an airport."[46] Under current TSA rules, that point is when a passenger either passes through a magnetometer or places his bag on the conveyor for the x-ray machine. Although the government argued during oral arguments in *Aukai* that this point of election could occur sooner if the TSA chooses to modify their rules, the court declined to speculate on the constitutionality of this argument.[47] The Ninth Circuit was quick to note, however, that while the constitutional validity of airport searches does not depend on consent, the scope of such searches is not limitless. The search must be no more extensive nor intensive than necessary given the state of current detection technology.[48]

Probable Outcome of a Court Case

Consider a hypothetical court case. David Defendant is attempting to fly from Knoxville, Tennessee, to Washington, D.C., to participate in one of the 'capital's famous pancake breakfasts, Tennessee Tuesdays, sponsored by the state's two United States Senators. David works as a mechanic but is also an avid gardener and uses a lot of fertilizer in his avocation. David had to work the morning of his flight and was in such a rush to get to the airport that he neglected to wash his hands before leaving. It should come as no surprise then that he had diesel fuel and traces of fertilizer on his hands. David also suffers from glaucoma, which he treats with marijuana. The possession and use of marijuana is a crime under Tennessee and federal law.

David arrives at Knoxville's McGhee Tyson Airport with plenty of time before his flight, as any responsible traveler should. David checks his baggage at the counter for his airline. As a smart traveler, David knows to carry important medications on his person or to pack them in his carry-on bag. Considering his marijuana to be a medical necessity, David tucks it into his pocket. Obviously, David is on a collision course with the American criminal justice system.

In line for the security screening, one TSA officer checks David's identification and ticket, and a second officer asks David if he minds having his hands swabbed and tested for explosives. Not wanting to rock the boat and not knowing that he had anything suspicious on his hands David consents. Unfortunately for David, ammonium nitrate, a commonly used fertilizer, becomes a potent and powerful

45 497 F.3d 955 (9th Cir. 2007).

46 *Id.* at 961.

47 *Id.* at 962 n.9.

48 *Id.* at 962.

explosive when combined with diesel fuel.[49] The hand swab detects the ANFO (ammonium nitrate / fuel oil) traces. The TSA officer alerts an officer from the airport authority, who conducts a more invasive search of David and his carry-on bag. The airport authorities pull David's checked bag aside and search it as well. The search turns up no other evidence of a bomb, and David is able to easily explain away the presence of ANFO on his hands. Unfortunately for David, the search also uncovers the' marijuana in his pocket.

David is arrested and charged in federal court with possession of marijuana. David hires a prominent local defense attorney, Bertie Barrister, to defend him. Barrister, recognizing a novel legal issue, files a motion to exclude the marijuana on the basis that the hand swab was an improper invasion of David's Fourth Amendment right against unreasonable searches. The deputy U.S. Attorney handling the case files a motion supporting the search and subsequent seizure of the marijuana.

The district court will first determine if the search was justified at its inception.[50] To do this, the court will need to determine what non-law enforcement interest the search served. In this case, that is clearly an antiterrorism security interest. The next step in determining whether the search was justified at its inception is to look to the *MacWade* factors cited favorably recently in *Dickerson v. Napolitano*.[51] Weighing the factors listed earlier, the weight of the government's interest in searching passengers is quite heavy. Recent attempts to destroy passenger airliners by such sartorial scoundrels as the Underwear Bomber or the Shoe Bomber have served to highlight the immediacy of the threat that airliners face from smuggled explosives.[52]

The privacy interests infringed by the search, however, are rather minimal. The palm of one's hand is the part of the body that most comes into contact with the world. It is likely that a passenger has touched dozens, if not hundreds, of other things on his way to the airport. The character of the intrusion is slight. Simply touching someone's hand does not seem terribly invasive; millions of Americans consent to such contacts everyday. The search is apparently very effective in detecting explosives, and no other available technology presents a less intrusive or more effective alternative. In the light of the minimal intrusion into a not very private area and the weight and immediacy of the threat, most, if not all, courts would find this type of search justified.

The next step in the two-step analysis is to determine whether the search was reasonably related in scope to the circumstances that justified the search in the

49 *See, e.g.*, Jo Thomas , *Jury to Be Picked in 2d Oklahoma Bomb Trial*, N.Y. Times, Sept. 29, 1997, at A17.

50 Epperson, *supra* note 35, at 771.

51 604 F.3d 732, 750-51 (2d Cir. 2010).

52 *See, e.g.*, Doug Guthrie and George Hunter, *Suspected 'underwear bomber' arraigned on terror conspiracy charges*, Detroit News, Dec. 17, 2010, *available at* http://www.detnews.com/article/20101217/METRO/12170394/1409/metro/Suspected-%E2%80%98underwear-bomber%E2%80%99-arraigned-on-terror-conspiracy-charges; *Judge denies bail to accused shoe bomber*, CNN, Dec. 28, 2001, http://archives.cnn.com/2001/US/12/28/inv.reid/.

first place.[53] The circumstances that justified the search would be the threat of a passenger smuggling explosives onto an airliner. People who are likely attempting to smuggle explosives onto a plane have probably come into recent contact with explosives. The ETD search is meant to alert the security officers to people who have handled explosive materials recently. ETD is not terribly useful for some purpose other than explosive detection. So using ETD is reasonably related to detecting people smuggling explosives.

The final issue to deal with is whether Defendant attempted to enter into a secured area of the airport.[54] As discussed earlier, courts agree that individuals have a lowered expectation of privacy when they enter a secured area. Whether David did or not may rest in part on whether the TSA has updated its policy manual to address searches prior to the point at which passengers reach a magnetometer or x-ray. Unfortunately, that information is not publicly available. If the TSA has updated its manual, then the court will have to deal with the question that the Ninth Circuit put off in *Aukai* regarding moving the metaphorical security portal. Given that the physical location of the magnetometer can vary from airport to airport, the courts may readily accept that expectations of privacy are lowered when a person lines up to pass through the magnetometer. A passenger will still have notice that he or she is now subject to a search, especially if there are signs informing the public that hand swabbing can occur in line are posted. The hand swab should also not be any more coercive than the magnetometer or x-ray.

Now that we have established that the hand swab was a valid search, we must decide whether the subsequent search of Defendant's person was justified. Extrapolating from the rules for subsequent searches based on magnetometer searches, security officers must first exhaust other less invasive means of search before patting down a person.[55] To the best of my knowledge, McGhee Tyson does not have any other means of explosive detection.[56] It is possible that security officers at McGhee Tyson could call for a bomb-sniffing dog from the Knoxville Police Department or the Knox County Sheriff's Office. However, bomb dogs may not always pinpoint where on a person's body explosives might be. So all a dog may be able to do is confirm the presence of an explosive trace on Defendant, which the ETD has already done. This means that there is no other less invasive search technique that security officers can use short of a frisk. Therefore, the search of David's person is valid along with the discovering of his contraband. The defense's motion to suppress fails.

A Word on Full-Body Imaging Scanners

Though the scope of this article primarily concerns swabbing hands for explosive traces, it would be remiss to not at least mention the newly implemented

53 Epperson, *supra* note 35, at 771.

54 U.S. v. Aukai, 497 F.3d 955, 961 (9th Cir. 2007).

55 Albarado, *supra* note 40, at 808.

56 Actually, to the best of my knowledge, McGhee Tyson does not utilize ETD at all, but for the sake of the hypothetical let's just assume that it does.

backscatter x-ray machines, given the amount of attention that such devices have received recently in the media. Despite the uproar over the use of backscanner x-ray machines, most analysts seem to conclude that searches using such devices are constitutional. I think that it is a much closer call than that.

As established earlier, part of the analysis for determining the validity of a search is weighing the relative intrusion of the search versus the effectiveness of the search in accomplishing its goal. Because the full-body imaging scanners create something akin to a nude department store mannequin based on the subject's body, they would have to be very effective to overcome the privacy hurdle.[57] Security experts debate the effectiveness of these machines, though. A terrorist can easily fool the scanners by hiding his contraband in a body cavity.[58] Rafi Sela, an Israeli security expert who helped design the security at Ben Gurion International Airport, has said, "'I don't know why everybody is running to buy these expensive and useless machines. I can overcome the body scanners with enough explosives to bring down a Boeing 747.'. . .'That's why we haven't put them in our airport.'"[59] If the scanners are indeed ineffective, then courts should not be willing to trade privacy for them.

Conclusion

The struggle to prevent terrorists from waging terror on commercial air flights is a constant escalation between evolving tactics and technology. The years since the September 11th terrorist attacks have seen airport search technology improve from simple metal detectors to explosive detection systems and full-body imagers. As airport searches evolve, we must continue to weigh the value of such searches against our value of privacy and protection against unreasonable searches. Many commentators accuse current security practices of being nothing more than security theater. It is up to the courts to weigh these searches against our rights. If security is, indeed, theater, then judges must be critics of, and not active participants in, the show.

57 *See, e.g.*, Tuan Nguyen, *New X-Ray Scanner at Airports Sees Through Clothes but Not Skin*, Daily Tech, Dec. 5, 2006, *available at* http://www.dailytech.com/New+XRay+Scanner+at+Airports+Sees+Through+Clothes+but+Not+ Skin/article5204.htm.

58 Sheila MacVicar, *Al Qaeda Bombers Learn from Drug Smugglers*, CBS Evening News, Sept. 28, 2009, *available at* http://www.cbsnews.com/stories/2009/09/28/eveningnews/main5347847.shtml.

59 Sarah Schmidt, *Full-body scanners 'useless,' air security expert says*, Vancouver Sun, Apr. 23, 2010, *available at* www.infowars.com/full-body-scanners-useless-air-security-expert-says/.

THE UNIVERSITY of TENNESSEE
KNOXVILLE

Howard H. Baker Jr. Center for Public Policy

The University of Tennessee • 1640 Cumberland Avenue • Knoxville, Tennessee 37996
865-974-0931 • 865-974-8777 (fax) • *www.bakercenter.utk.edu* • e-mail: *bakercenter@utk.edu*