Military and Strategic Affairs

Volume 3 | No. 3 | December 2011

INSS
המכון למחקרי ביטחון לאומי
THE INSTITUTE FOR NATIONAL SECURITY STUDIES
INCORPORATING THE JAFFEE CENTER FOR STRATEGIC STUDIES
TEL AVIV UNIVERSITY
אוניברסיטת תל-אביב

Military and Strategic Affairs

Volume 3 | No. 3 | **December 2011**

CONTENTS

Military and
Strategic Affairs

The purpose of *Military and Strategic Affairs* is to stimulate and enrich the public debate on military issues relating to Israel's national security.

Military and Strategic Affairs is published three times a year within the framework of the Military and Strategic Affairs Program at the Institute for National Security Studies. Articles are written by INSS researchers and guest contributors. The views presented here are those of the authors alone.

Graphic Design: Michal Semo-Kovetz, Yael Bieber
Tel Aviv University Graphic Design Studio

The Institute for National Security Studies (INSS)
40 Haim Levanon • POB 39950 • Tel Aviv 61398 • Israel
Tel: +972-3-640-0400 • Fax: +972-3-744-7590 • E-mail: info@inss.org.il

Military and Strategic Affairs is published in English and Hebrew.
The full text is available on the Institute's website: www.inss.org.il

The Strategic Uses of Ambiguity in Cyberspace

Martin C. Libicki

Strategic ambiguity has an honored place in the mores of statecraft. The studied unwillingness of states to say what they have done (or would do) coupled with the lack of proof that they have done it (or would do it) liberates other states. They can argue that something was done, but if their purposes so dictate, they can pretend that it was not done. The degree of doubt can vary: from thorough (no one is sure what has happened or would happen) to nominal (no one is fooled). In either case, however, those who did it have provided a fig leaf, however translucent, that other states can adopt.

Examples of Strategic Ambiguity in Physical Space

One time-honored example is Israel's refusal to admit (or deny) that it has nuclear weapons. No reputable analyst believes that Israel does not have nuclear weapons. But since Israel has never announced whether it has any, other states are free to pretend that Israel has not crossed the nuclear barrier. This is convenient for states that would be pressured by their people to respond with nuclear programs of their own were Israel's status overt. It also helps states that could not ship certain classes of exports to Israel were Israel's status more open.[1] At the same time, no sane country behaves as if Israel lacked a nuclear retaliation capability.

A parallel ambiguity concerns the putative US use of Predator attack flights and cruise missiles against al-Qaeda members in countries such as Yemen or Pakistan. Official policy is to deny that such flights take place. When Yemen's leader claimed that these were Yemenite operations, very few analysts were fooled. But at least until recently, the leaders of

Dr. Martin C. Libicki is a Senior Management Scientist at the RAND Corporation.

these countries did not have to contend with admitting that sovereignty violations were taking place, with at least their tacit permission.

Another longstanding example is US policy towards Taiwan's independence. The United States has declared both that it opposes a Taiwanese declaration of independence and any attempt to resolve the status of Taiwan by force. The United States does not recognize Taiwan as a state and so has no mutual aid pact with it. However, if Taiwan declared independence and China decided to take the island, would the United States intervene on Taiwan's side? It is clearly in the US interest for China to think so in order that China does not start a war. But it is almost as clearly in the US interest for Taiwan to think otherwise, so that Taiwan does not provoke China into starting a war. Assume the odds of a US intervention are literally a coin toss and perceived that way on both sides of the Straits. If so, Taiwan may well calculate that the expected value from declaring independence is negative (whereas it would have been positive if the US were definitely coming to help), due to the fact that the United States might decide not to intervene. Similarly, China could conclude that the expected value of a cross-Straits invasion is also negative because the United States might intervent. Anything less ambiguous could well prompt one or the other to do something foolish.

Cyberspace is Tailor-Made for Ambiguity

Cyberwar is, literally, inside work. When hackers enter a computer system to misdirect its workings, the direct results are often literally invisible to the outside world. Depending on how such systems have been misdirected, the indirect results may be invisible as well. True, the results of a cyber attack on a power grid that turns off the lights can be viewed even from space. But without further investigation and revelation, it will not be clear whether a blackout was a deliberate attack, or the result of human error, bad software, or (most frequently) Mother Nature. Even if it were clear that a system misbehaved because it had been attacked, exactly who attacked may be shrouded in mystery. Finally, even if the fact and the author of the cyber attack were clear, the purpose may be quite obscure: after all, cyberwar alone cannot kill anyone, or even break very much (but see Stuxnet), much less seize territory or change a regime (and whereas cyberwar can facilitate other applications of force, it is those other applications that are more visible). Nearly all intrusions are meant to steal information or "rent" the capacities of the target machine (as in a bot) and otherwise leave the

system alone. Deliberate attacks can often be framed as attempts to mislead people (e.g., false radar images) or their equipment (see Stuxnet). In the latter cases, obviousness is self-defeating; once it is clear that you have successfully deceived a system, the system's administrators are unlikely to allow the system to operate as it has.

Is Stuxnet an Exception?

One would imagine that a cyber attack that actually broke something might have passed the point where everyone could be try to hide its existence. The Stuxnet worm was discovered in June, 2010, and its target was identified as an Iranian nuclear facility in September. The earliest suspicions tagged the Bushehr reactor as its target,[2] and the Iranians denied that any such reactor was affected. Within a few weeks, the Natanz centrifuge plant was identified (more plausibly) as its target. Initial Iranian denials were contradicted in late November, 2010, the day that assassins killed two Iranian nuclear scientists, and when Ahmadinejad admitted that there was a worm that had caused a great deal of trouble, which was then taken care of.[3] How badly did Stuxnet, in fact, hurt Iran's nuclear development? Statistics from the IAEA would indicate that it may have led to the premature retirement of 10 percent of Iran's centrifuges and thus, at most, it bought the worm's creators several months reprieve from the data at which Iran would have enough nuclear material to build its first bomb.[4] Other reports quote officials predicting that the earliest that Iran can (as of early 2011) assemble such material would be 2015, a delay of several years.

There is a lot more (apart from what it accomplished) that is currently unclear about Stuxnet.[5] One question is how it got into Natanz in the first place; suspicions that the worm's designers received witting or unwitting help from Russian contractors appears to have soured Iran's working relationship with them.[6] More important is exactly who wrote and released the worm. Was it an individual (its sophistication says otherwise)? Was it Israelis – as suggested by several clues internal to the code – but who knows that these clues were not planted to mislead suspicion? Was it Americans? Was it both, working together?[7] Or, was it the Chinese?[8] With all the ambiguity, it is no wonder that Iran has yet to retaliate (at least in any noticeable way). That noted, Syria did not respond to the strike on its suspected nuclear facility, and Iraq did nothing but complain when its Osiraq reactor was bombed – and there was no ambiguity who did it in both cases. Conversely, Iran's strong ties to Hamas and Hizbollah suggest that

it may have had ways of expressing its displeasure that were unavailable to Syria (in 2007) or Iraq (in 1981). Furthermore, Iran has yet to make much of a big deal about the incident; likening it to an act of war after months of silence and denials would be quite a volte-face.

The advantages of using Stuxnet rather than airpower to degrade Iran's nuclear capability are fairly clear (assuming the worm, in fact, did as its designers hoped): comparable effect, and induced distrust among its victims as to which of its suppliers or supplies may still be contaminated, but with less condemnation (indeed, perhaps a sneaking admiration) and fewer strategic risks.

The Uses of Ambiguity

The working hypothesis is that a cyber attack used in lieu of kinetic methods creates more ambiguity in terms of effects, sources, and motives. Thus, if cyber attacks work – and this is a tremendous if – they change the risk profile of certain actions, and usually in ways that make them more attractive options. What follows are some hypothetical uses of cyber attacks.

One, cyber attacks may be used by a victim of small scale aggression to indicate its displeasure but with less risk of escalation than a physical response would entail. In late 2010, for instance, North Korean forces shelled a South Korean island, killing two civilians and two service members. A retaliatory cyber attack that disrupted an important industrial facility (ignoring the fact that North Korea is not well digitized and has nearly zero network connections to the rest of the world) could have conveyed displeasure. North Korea, if it wanted to respond, would have had to (1) admit that one of its facilities had been hacked, and (2) take steps to indicate why it was South Korea, and only South Korea that was at fault (it could be the United States or even Japan, and China). Conversely, if North Korea did not react publicly, it stood a good chance of limiting the number of people with a good idea of why some facility ceased working. This introduces another advantage of cyber warfare over physical combat: although being attacked may be a source of pride (e.g., you can play David to the enemy's Goliath), being hacked primarily means that you ventured into cyberspace with inadequate attention to maintaining control over your systems. Victimhood is not something worth boasting about. Thus, states that can hide having been attacked may well do so, thereby saving face – but doing so also making an obvious response less likely. They could,

of course, respond in kind and so a tit-for-tat struggle that started in the physical worlds ascends (or descends) into the virtual one. But that course may be safer all around than coming to blows.

Two, a state rich in cyber warriors may also use the threat of cyberwar to deter the potential target against support proxy war fighters: e.g., Israel could threaten Iran with cyber attacks if Israel is attacked by Hizbollah, a group with known links to Iran.[9] In this situation, Israel may not want to make such a threat public. A public threat would allow Hizbollah to coerce Iran by claiming a desire to wreak the sort of mischief that would prompt Israel to strike Iran in cyberspace. But there are private ways to convey the threat, and such a threat has logic. The usual problem with cyber deterrence is that attribution (of the starting attack) is a problem, but a physical attack – say, Hizbollah rockets striking Israel – would be obvious. Conversely, although a state like Iran may not fear a direct Israeli attack even in response to a Hizbollah attack (no such attack materialized in 2006, for instance), it may fear a cyber attack given the clear superiority of Israeli hackers over Iranian ones. Such superiority mitigates (although it does not erase) the fear that having declared the intention to carry out a cyber attack, Israel would have no accessible targets in Iran; even if the success of any one attack is uncertain, the odds that enough will succeed and hurt are sufficiently good. Iran's blaming the United States afterwards may be a problem for the United States but make things easier for Israel. Escalation into violence is not really an option for Iran given Israel's conventional combat dominance (at least if the battle were close to Israel). More to the point, Iran would have to admit its systems had been conned and make a convincing case that it knew who did it. Finally, while Israel is more wired than Iran, again, with Israel's cyber capabilities, that fact may not be enough to turn the tide towards Iran's favor should it strike back.

Three, cyber attacks can be used by one state to affect the outcome of conflict in another state without having to make any sort of visible commitment, even an implied one. Consider the civil war in Libya. If Libya's military was sufficiently wired so that cyber attacks could conceivably make a difference in its capabilities,[10] then Western hackers, by disabling the central government's forces, could conceivably tilt the direction of the fight. If the rebels won, Western governments would be better off as a result. Rebel forces, at worst, would have no way of knowing they had received assistance, and that may be just as well (particularly regarding

the more jihadist of Libya's rebels who greet the intervention of US forces by switching sides). Or, hints could be offered (e.g.: if this capability fails tomorrow, you will know why). Conversely, if the government won, it may suspect that its information systems were tampered with by Western forces, but it may not be able to prove as much. It may complain, but if Libya were expected to blame its shortfalls on the West, then such complaints, in the absence of evidence, would have little force. More to the point, it may not want to claim as much if it wants to pretend afterwards that it has no reason to make enemies of the West all over again. If the civil war drags on, the West can pretend that it had made no prior help and thus had made no commitment to escalate its assistance (even if hints were dropped to the rebels, they would have an even harder time proving to others that Western hackers were offering assistance, since unlike the government, they would likely have no access to the tampered computers). The greatest problem in offering such assistance is the possibility of getting caught, but if the target of the attacks is on the outs with the rest of the world, it is unlikely that it will get much help tracing the attacks. So attractive is such assistance (at least from the helper's perspective) that it may be a routine feature – on both sides – of any conflict where the outcome is uncertain and networks matter to war fighting capabilities. And again, admitting that one's systems have been hacked is always at least a little embarrassing.

Four, cyber attacks do not need to be directed towards adversaries, although the risks of making new enemies if the source of the cyber attacks are discovered are obvious. Consider a situation in which two neutral states are inching towards war that one might prefer not take place. Suppose that a third state is capable of introducing faults into both sides' surveillance and/or command-and-control systems that raise doubts whether they have pierced the fog and overcome the friction enough to undertake military operations. If systems go haywire, either target state is more initially likely to blame the other for its woes (if they understand that such woes were obvious *and* induced rather than non-obvious or accidental) rather than a third party; chances are that the initial presumption is likely to color their forensic activities and conclusions. Furthermore, there is a good chance that such blame will be kept private given the embarrassment involved. Yet risks exist in such maneuvers; such machinations may drive states towards war if one side or both comes to convince itself, for instance, that the cyber attacks from the other side are precursors to an immediate movement of forces, or are indications that their foes' forces are not just posturing.

A variant on this technique is to use cyber attacks to disable a capability in a state whose leadership is reluctant to use it anyway (either because the leadership feels itself to be on shaky political ground vis-à-vis its excitable populace, or because the leadership is exercised by a consensus among factions[11]). Once such a capability is found inoperative, the political leadership announces to its military leaders that it has no option but to stand down. Perhaps the military unearths evidence that a third party was behind such an incapacity – the political leadership, relieved at not having to act, may deem such evidence inconclusive or not credible it in the first place.

Five, ambiguity may be useful in declaratory policy, one that indicates how a state would respond to a cyber attack. Ambiguity has both costs and benefits. The cost is that others may think they can get away with attacks that they would have forborne if they had understood that reprisals would follow. But the benefit is that the target state may not want to strike back, particularly if it lacks the confidence to attribute the attack. A state that fails to strike back because it is unsure may not lose stature in its own eyes – attribution really is difficult. Yet if the attacker (and others) come to believe that such a state *did* know but pretended otherwise for fear of a full-scale fight, then any threat to retaliate rings hollow – and not just in cyberspace. If a state leans too far forward in promising reprisals in response to cyber attacks and cannot deliver, its ability to deliver against all other threats may be further doubted.

Conclusion

Cyberwar's many tactical ambiguities lend force to a strategy built on strategic ambiguities. There may be many cases in which an aggressor state does not want what it has done it to be obvious. Even the target state in some cases may conclude that pretending as much (even if it must turn a blind eye to the evidence) has advantages over trying to clarify matters or even claiming clarity in absence of the real thing.

But the downside to strategic ambiguity should be noted. States may arrogate the right to carry out all sorts of mischief in cyberspace on the belief that they will never be called into account. The lack of accountability, however, is inherently dangerous. Sometimes it is unwarranted (the state is only fooling itself), and even if warranted, it provides hackers a degree of freedom that history suggests is dangerous in and of itself.

Notes

1 By contrast, legislation had to be passed in 2006 to permit the United States to share civilian technology with India, which like Israel is a non-signatory to the Non-Proliferation Treaty, but unlike Israel, a declared nuclear power. See Peter Baker, "Signs India Nuclear Law: Critics Say Deal to Share Civilian Technology Could Spark Arms Race," *Washington Post*, December 19, 2006, www.washingtonpost.com/wp-dyn/content/article/2006/12/18/AR2006121800233.html.

2 Robert McMillan, "Was Stuxnet Built to Attack Iran's Nuclear Program?" IDG News, taken from *PCWorld*, September 21, 2010.

3 William Yong, Alan Cowell, "Bomb Kills Iranian Nuclear Scientist," *New York Times*, November 30, 2010.

4 Joby Warrick, "Iran's Natanz Nuclear Facility Recovered Quickly from Stuxnet Cyberattack," *Washington Post,* February 16, 2011. See also the report by the Institute for Science and International Security, http://media.washingtonpost.com/wp-srv/world/documents/stuxnet_update_15Feb2011.pdf.

5 What *is* most clear about Stuxnet is how it worked because the worm was captured alive, so to speak, in the wild before it could self-destruct (which it should have done if it was unable to find a specific programmable logic device that met certain preset parameters associated with a particular type of centrifuge).

6 "The Stuxnet Worm: A Cyber-Missile Aimed at Iran?" *Economist*, September 24, 2010, www.economist.com/blogs/babbage/2010/09/stuxnet_worm.

7 William Broad, John Markoff, David Sanger, "Israel Tests on Worm Called Crucial in Iran Nuclear Delay," *New York Times*, January 15, 2011.

8 Jeffrey Carr, "Stuxnet's Finnish-Chinese Connection," December 14, 2010, blogs.forbes.com/firewall/2010/12/14/stuxnets-finnish-chinese-connection/.

9 Many observers take issue with the characterization of Hizbollah as a puppet of Iran. Yet there is a difference between Hizbollah acting only on Iran's orders, and Iran having enough influence on Hizbollah to discourage it from unwise actions.

10 An influential article reviewing the possibilities of Western intervention in Libya mentioned electronic warfare in the form of communications jamming, but nothing about cyber warfare. See Thom Shanker, "U.S. Weighs Options, on Air and Sea," *New York Times*, March 6, 2011, http://www.nytimes.com/2011/03/07/world/middleeast/07military.html.

11 If the fact that China's stealth fighter surprised Hu Jintao when meeting with Secretary of Defense Gates is any indication, its military is not absolutely beholden to its political leadership and thus the country's effective leadership may also be somewhat of a coalition.

Unraveling the Stuxnet Effect:
Of Much Persistence and Little Change in the Cyber Threats Debate

Myriam Dunn Cavelty

Cyber threats have been on the security political agenda for a number of years. Since RAND researchers John Arquilla and David Ronfeldt suggested in 1993 that "cyberwar is coming!"[1] cyberwar has become the most prominent buzzword in the debate surrounding computers, national security, and cyberspace. Being at the mercy of well-publicized events and occurrences, interest in the topic used to flare up whenever anything involving the aggressive use of computers hit the news, only to disappear again when other issues took over the limelight.

This changed in 2010. In particular, it was Stuxnet, the sophisticated computer worm written to sabotage systems that control and monitor industrial processes, that stirred up the international community in major ways and catapulted the cyber topic into the sphere of public fears and to the top of everybody's threat list. As a result, more and more countries consider cyber attacks to be one, if not *the* major future security threat.

But how justified is this assumption? And what has Stxunet really changed in the debate?

This article aims to provide a balanced picture of the phenomenon of cyberwar. It will show how and why the meaning of "cyberwar" has evolved from the narrow conception referring exclusively to military interaction to its broad meaning, which has become detached from "war" and encompasses almost every activity linked to the aggressive use of computers. In particular, it will distinguish between different forms of cyber conflict in order to lay the ground for a levelheaded threat assessment.

Dr. Myriam Dunn Cavelty is head of the New Risk Research Unit at the Center for Security Studies in Zurich, Switzerland.

It further shows that there is probably less change and more persistence in the cyber threat debate at large than is currently acknowledged. The threat image has been quite solid since the late 1990s, and Stuxnet has not changed this to any substantial degree. The same can be said for the countermeasures that are planned or envisaged.

Contexts and Meanings of Cyberwar

The importance and emergence of the concept of cyberwar can best be understood in the larger context of the information revolution, which has shaped – and is still shaping – perceptions of opportunities and dangers. In particular, the technologies of the information revolution and related organizational innovations in the 1980s and 1990s seemed to alter the nature of conflict and the kinds of military structures, doctrines, and strategies needed. Thus, it seemed to imply the rise of a "new" kind of warfare in which the factor of information was to grow more and more important. This development was facilitated (if not driven) by the end of the Cold War and the ensuing reorientation in terms of enemies, strategic thought, and defense spending.

It was the second Persian Gulf war of 1991 that created a watershed in military thinking about cyberwar. That conflict was seen by military strategists (mainly American) as the first of a new generation of conflicts where victory is no longer ensured only by physical force, but also by the ability to win the information war and to secure "information dominance." As a result of the conflict, strategists began to publish scores of books on the topic.[2] The reaction to the technological developments after the Gulf War also manifested itself in the publication of new doctrinal papers that institutionalized the information component.

The debate was initially characterized by a great deal of euphoria. Soon after, however, more attention was given to the risks associated with this development. Specifically, the formulation of strategies that no longer aimed at enemy capabilities but directly targeted the opponents' flow of information highlighted the relatively high vulnerability of networked US troops. As the debate over attacks on potential hostile information systems progressed, the possible dangers to civilian data networks were also increasingly discussed. The US as the only remaining superpower was seen as predestined to become the target of asymmetric warfare. Widespread fear took root in the strategic community that those likely to fail against

the US war machine might instead plan to bring the US to its knees by striking against vital points at home, namely, critical infrastructures.[3] The concept of critical infrastructure includes sectors such as information and telecommunications, financial services, energy and utilities, and transport and distribution. It also includes a list of additional elements that vary across countries and over time.[4] Most of these sectors rely on a spectrum of software-based control systems for their smooth, reliable, and continuous operation.

With the growth and spread of computer networks into more and more aspects of everyday life, the object of protection moved from being perceived to be limited proprietary (governmental, mainly military) networks to encompass the whole of society – or rather, its way of life provided by the uninterrupted sub-structure of technology.[5] On this basis, a comprehensive threat image with two interrelated sides evolved. First, an inward-looking perspective sees the very connectedness of infrastructure systems as posing dangers, because perturbations within them can cascade into major disasters with immense speed and beyond our control. Advances in information and communication technology have thus augmented the potential for major disaster in critical infrastructures by vastly increasing the possibility for local risks to mutate into systemic risks. Second, an outward-looking perspective focuses on the increasing willingness of malicious actors to exploit vulnerabilities without hesitation or restraint. Because critical infrastructure systems combine symbolic and instrumental values, attacking them becomes integral to a modern logic of destruction that seeks maximum impact.

In addition, the cyber dimension reformulates space into something no longer embedded in place or presence. The "enemy" becomes a faceless and remote entity, a great unknown that is almost impossible to track. This results in two significant characteristics of the threat representation. First, the protective capacity of space is obliterated; there is no place that is safe from an attack or from catastrophic breakdown in general. Second, the threat becomes quasi universal because it is now everywhere.

A Cyber Phenomenology

It comes as little surprise, then, that cyber threats are feared the way they are. Nonetheless, every observer cannot help but notice how unspecified the threats actually are. By leaving its military confines, the concept became

greatly blurred: cyberwar has come to refer to basically any phenomenon involving a deliberate disruptive or destructive use of computers.

Such conceptual vagueness is not helpful if we are to understand what goes on in "cybered" conflicts[6] and what kinds of countermeasures are actually needed for what kind of phenomena. Bruce Schneier, an internationally renowned security technologist and author, differentiates between cyber vandalism, which includes the defacing of websites; cyber crime, which includes theft of intellectual property, extortion based on the threat of Distributed Denial of Service attacks (DDoS) attacks, fraud based on identity theft, and so on; cyber terrorism, e.g., hacking into a computer system to cause a nuclear power plant to melt down, a dam to open, or two airplanes to collide; and cyberwar.[7] Schneider uses "cyberwar" to refer to the use of computers to disrupt the activities of an enemy country, especially deliberate attacks on communication systems.

Schneier's classifications construct a cyber threat escalation ladder – from rung to rung, the potential effects as well as the scope and the intensity become more severe. The last few years have shown that cyber espionage and cyber sabotage are missing from this ladder. More important, however, is that the lines of demarcation between the different activities are greatly blurred. When a particular detrimental event occurs, it is often difficult to determine whether it is the result of a malicious attack, a failure of a component, or an accident. And although their goals are different, the tools and tactics used by armies, terrorists, and criminals in cyberspace are very similar, if not the same. This means that knowing who is behind an attack and what kind of phenomenon it constitutes is a major difficulty when it occurs.

Then again, just because it is difficult does not mean that such a differentiation is not necessary: the opposite is true. First, the advantage of a "severity of effects" view is that it helps policymakers prioritize in theory, which is highly needed. Only computer attacks whose effects are sufficiently destructive or disruptive should be regarded as a national security issue – and should therefore earn the attention needed for something existentially threatening. Attacks that disrupt nonessential services or that are mainly a costly nuisance are not.[8] Second, a narrow and precise definition also helps to circumvent other dangers inherent in calling something "war," like exculpating the victims of an attack from their own responsibility for the consequences of their negligence in terms

of computer security or creating pressure to retaliate against hackers, real or imagined.[9] Third, it clearly shows where the center of gravity lies: with careful computer forensics. Each and every occurrence must be carefully investigated. As Schneier notes:

> Just as every shooting is not necessarily an act of war, every successful Internet attack, no matter how deadly, is not necessarily an act of cyberwar. A cyberattack that shuts down the power grid might be part of a cyberwar campaign, but it also might be an act of cyberterrorism, cybercrime, or even – if it's done by some fourteen-year-old who doesn't really understand what he's doing – cybervandalism. Which it is will depend on the motivations of the attacker and the circumstances surrounding the attack...just as in the real world.[10]

Threat Assessment

That said, how endangered are we? Conflicts in cyberspace have been a reality for over a decade: elements of any political, economic, and military conflict take place in and around the internet. Furthermore, criminal and espionage activities aided by information and communication technologies take place every day. But in the entire history of computer networks, there have been very few examples of severe attacks that had the potential to disrupt or actually did disrupt the activities of a nation state in a major way. There are even fewer examples of cyber attacks that resulted in physical violence against persons or property. The huge majority of cyber attacks are low level and cause inconvenience rather than serious or long term disruptions. In fact, it has been convincingly shown that a "pure" (or strategic) cyberwar is very unlikely to ever occur, with attacks on computer systems more likely to be used in conjunction with other, physical forms of attack.[11]

Did this estimation change with Stuxnet? Classifying Stuxnet according to the escalation ladder is a challenge. Stories and speculations about the worm, its origins, and its intent exist by the thousands.[12] Well written or less so, they all contain bits and pieces of a puzzle that is inherently unsolvable. The pieces of the puzzle all seem to suggest that only one or several nation states – the usual "cui bono" logic pointing either to the US or Israel – would have the capability and interest to produce and release Stuxnet in order to sabotage the Iranian nuclear program. Though the world will probably never know for certain who is behind this piece of code, the majority of

strategic planners out there are willing to believe that a "digital first strike" has occurred and a virtual Pandora's Box has been opened.

However, even if the most extreme case is assumed – that the majority of states in this world have developed effective and powerful cyber weapons or will in the near future (which is very doubtful) – the mere existence and availability of such capabilities does not automatically mean that they will be used. The cyber realm seems to lead people to assume that because they have vulnerabilities they will be exploited. Still, in security and defense matters, careful threat assessments need to be made. Such assessment necessitates the careful deliberation of the following question: "Who has the interest and the capability to attack us, and why would they?" For many democratic states, the risk of war has moved far to the background. The risk of a cyber attack of the severest proportions should be treated the same if there is no natural enemy.

Unraveling the Stuxnet Effect

On the other hand, the publication of Stuxnet's code and many other details has already led to many piggyback attacks. SCADA systems – computer systems that monitor and control industrial, infrastructure, or facility-based processes – are therefore likely going to be the target of choice for any kind of hacker in the near to midterm future. This comes with an inherent danger of intended and unintended (side) effects, of course – but in fact, the critical infrastructure community has been talking about the threat to SCADA systems for over a decade. In addition, experts have been expecting a major occurrence in cyberspace for a long time. Seen this way, Stuxnet is less of a surprise and more of a confirmation of what has been discussed and feared for years. Though it has focused the minds of politicians on the upper two rungs of the ladder, at least temporarily, it does not change the probability of cyber terror or cyberwar occurring.

It also does not change the methods and tools available to counter cyber threats. This concerns information assurance measures, for example, or the many diverse activities, concepts, and processes subsumed under "critical infrastructure protection" (CIP). CIP is handled similarly in many states:[13] close partnerships with the corporate sector and international partners are sought, mostly in order to exchange information on threats and issues. In addition, more recently, a shift away from the concept of protection towards the concept of "resilience" can be observed.[14] Resilience is not

a new concept, of course, but its current rise indicates a significant and crucial shift in thinking. While protective (and defensive) measures aim to prevent disruptions from happening, resilience accepts that certain disruptions are inevitable.

Such thinking is absolutely necessary and needs to become rooted deeply in politicians' minds and subsequently in the minds of the population. Information networks can never be "secure" in the national security sense. In fact, the opposite is true: cyber incidents are fated to happen, because they simply cannot be avoided. In other words, even the most perfect defenses will not be able to guarantee that nothing severe will happen in a networked world.

States have the tendency to react forcefully to such a challenge and try to increase the level of security by all means. But cyberspace should not be mistaken for just another "realm" in which military action can be taken at will. To continue reaping the benefits of the cyber age, it is necessary to learn how to live with insecurity in pragmatic ways. Apart from legal and strategic restraints that will certainly be factored into any consideration of whether to use cyber attacks as weapons or not, the biggest impediment should be fears of uncontrollable blowback. First of all, repercussions could emerge directly through the interdependencies between various critical assets that characterize the environment. Second, blowback may be felt through the more intangible effect of undermined trust in cyberspace, with damaging repercussions for the global economy.[15]

By implicitly or explicitly moving an issue into the realm of national security and military actions, one tends to subject it to the rules of an antagonistic zero sum game, in which one party's gain is another party's loss. The logic of cyberspace, however, is a different one. Like the governance of space and the oceans, its governance requires globally accepted norms. The avenues currently available for arms control in this arena are primarily information exchange and norm building, whereas attempts to prohibit the means of cyberwar altogether or restricting the availability of cyber weapons are likely to fail. However, these difficulties should not prevent the international community from pushing all countries to adopt responsible limits and self-restraint in the use of cyber weapons and from thinking about new and innovative ways to enhance protection of vital computer networks without inhibiting the public's ability to live and work with confidence on the internet.

Notes

1 John Arquilla and David F. Ronfeldt, "Cyberwar is Coming!" *Comparative Strategy* 12, no. 2 (1993): 141-65.

2 Greg Rattray, *Strategic Warfare in Cyberspace* (Cambridge: MIT Press, 2001); Michael O'Hanlon, *Technological Change and the Future of Warfare* (Washington: Brookings Institution, 1999).

3 Myriam Dunn Cavelty, *Cyber-Security and Threat Politics: US Efforts to Secure the Information Age* (London: Routledge, 2008).

4 Elgin Brunner and Manuel Suter, *International CIIP Handbook 2008/2009* (Zurich: Center for Security Studies, 2009).

5 Myriam Dunn Cavelty, "Cyber-Security," in Peter Burgess, ed., *The Routledge Handbook of New Security Studies* (London: Routledge, 2010), pp. 154-62.

6 Chris Demchak, "Cybered Conflict as a New Frontier," *Atlantic Council*, October 28, 2010, http://www.acus.org/new_atlanticist/cybered-conflict-new-frontier.

7 Bruce Schneier, "Schneier on Security: A Blog Covering Security and Security Technology," Post: "Cyberwar," June 4, 2007, http://www.schneier.com/blog/archives/2007/06/cyberwar.html.

8 Cf. Clay Wilson, *Computer Attack and Cyber-terrorism: Vulnerabilities and Policy Issues for Congress*, Congressional Research Report for Congress (Washington: Congressional Research Service, 2003) and Dorothy Denning, "Activism, Hacktivism, and Cyberterrorism: The Internet as a Tool for Influencing Foreign Policy," in John Arquilla and David F. Ronfeldt, eds., *Networks and Netwars: The Future of Terror, Crime, and Militancy* (Santa Monica: RAND, 2001), pp. 239-88.

9 Martin Libicki, *Defending Cyberspace and Other Metaphors* (Washington: National Defense University, 1997), p. 38.

10 Schneier, http://www.schneier.com/blog/archives/2007/06/cyberwar.html.

11 Peter Sommer and Ian Sommer, *Reducing Systemic Cybersecurity Risk*, OECD/IFP Project on Future Global Shocks, 2011, www.oecd.org/dataoecd/3/42/46894657.pdf.

12 Two prominent examples are: Mark Clayton, "Stuxnet Malware is Weapon out to Destroy Iran's Bushehr Nuclear Plant," *Christian Science Monitor*, September 21, 2010, www.csmonitor.com/USA/2010/0921/Stuxnet-malware-is-weapon-out-to-destroy-Iran-s-Bushehr-nuclear-plant; and William J. Broad, John Markoff, and David E. Sanger, "Israeli Test on Worm Called Crucial in Iran Nuclear Delay," *New York Times*, January 15, 2011, http://www.nytimes.com/2011/01/16/world/middleeast/16stuxnet.html.

13 Myriam Dunn Cavelty and Manuel Suter, "Public-Private Partnerships are no Silver Bullet: An Expanded Governance Model For Critical Infrastructure Protection," *International Journal of Critical Infrastructure Protection* 2, no. 4 (2009): 179-87.

14 Christine Pommerening, "Resilience in Organizations and Systems: Background and Trajectories of an Emerging Paradigm," in *Critical Thinking: Moving from Infrastructure Protection to Infrastructure Resilience*, CIP Program Discussion Paper Series (Washington: George Mason University, 2007), pp. 9-22.

15 Andrew Rathmell, "Controlling Computer Network Operations," *Information & Security: An International Journal* 7 (2001): 121–44.

An Interdisciplinary Look at Security Challenges in the Information Age

Isaac Ben-Israel and Lior Tabansky

Introduction

Developments in electronics and computers since World War II have affected a broad range of fields and created the "information age." This article focuses on interrelationships among information technology, the information age, and security. More specifically, it aims to contribute to a discussion of the national security issues stemming from the development of information technology.

Much of the driving force behind computer development has been derived from military applications. Following new possibilities, thinking about the effect of technological change on defense issues has also progressed. In addition, the information age, which continues to develop rapidly, along with advances in computer communications and the penetration of computers into every area of life, has given rise to cyberspace. These developments challenge existing perceptions and force reconsideration of basic concepts. The need for an informed public debate and the design of a firm policy has likewise grown, given the fact that the cyberspace risk is already concrete – as dramatized by events in Estonia in the spring of 2007, as well as the Stuxnet affair.[1] In Estonia, daily life was disrupted following a technically simple but massive attack on internet-based services. With Stuxnet, it appears that a technically complex cyber weapon was used, designed to cause precise damage to the system controlling the industrial process at a protected nuclear fuel enrichment

Prof. Isaac Ben-Israel is head of the Yuval Ne'eman Workshop for Science, Technology and Security at Tel Aviv University. Lior Tabansky is a Neubauer research associate working on the Cyber Warfare Program at INSS, which is supported by the Philadelphia-based Joseph and Jeanette Neubauer Foundation.

facility in Iran. The weapon's design and method of operation included camouflage of its activity for a prolonged period. This cyber weapon apparently caused cumulative physical damage of strategic significance. The consensus is that in both incidents, states were behind the cyber attacks, though in both cases no definitive evidence exists.

A basic theoretical understanding of the information age is essential in order to consider cyber security issues. This article relies on ideas by philosopher Karl Popper, futurists Alvin and Heidi Toffler, and economist Paul Romer to illuminate the characteristics of the information age and to clarify the issues that emerge when technological development interfaces with national security. It analyzes the current characteristics of cyberspace, and discusses the implications for national security questions. It then reviews the field known as information warfare and focuses on the totally new phenomenon of computer warfare in cyberspace. The article then reviews cyber weapons and methods of warfare, discusses defense, attack, and deterrence, and presents key issues in the cyber defense realm. It appears that in order to maintain security and peace, a multidisciplinary assessment of the new issues and challenges is required.

Theoretical Background

Technological change occupies many thinkers who struggle to assess its social effects. Although the scope of this article does not permit a full review of the field, three thinkers relevant to an understanding of the dynamic reality must be mentioned.

The term "Third Wave," taken from the theories of the bestselling authors Alvin and Heidi Toffler, refers to a time period (table 1). According to the Tofflers, we are in the midst of a transition to the Third Wave, in which the economy is based on knowledge and control of information,[2] instead of on industrial mass production. Similarly, the form of warfare is changing as well. The name of the game has become obtaining information about the enemy and denying it information about yourself. The side that controls information technologies will win the war, even if it faces many weapons rolling off Second Wave assembly lines.

Table 1. The Waves According to the Tofflers

	Principal Resource	Who is Rich	Symbol	Weapons	Method of Waging War
The First Wave	Organized agriculture	Landowner	Sickle	Sword	Face-to-face battle at point blank range; land conquest
The Second Wave – from the mid-17th century until the end of the 20th century	Automated industry, mass production	Industrialist	Machinery of mass production assembly lines	Tank, airplane	Machines used at medium range, poor accuracy, attempt to damage production capacity
The Third Wave – from the end of the 20th century onwards	Knowledge	Bill Gates	Computer	Cyber warfare	Attempt to damage information through the use of computers. Remote damage to functional capacity, without physically reaching the target

Concepts developed by philosopher Karl Popper, who died in 1994, enhance the theoretical stage. Popper analyzed the world of knowledge as another existing concept, in addition to the material and spiritual worlds (table 2).[3] Popper insists that an entire "world" of human knowledge exists (World 3), populated by "beings" that are objective contents of thought, such as the Pythagorean Theorem and the laws of physics. These are neither "material" nor subjective "mental experiences." Once the Pythagorean Theorem was formulated, it became an objective truth independent of the spirit that created (or discovered) it. In other words, knowledge is objective, even though it is a product of the (subjective) human spirit.

Table 2. Popper's Three Worlds and Cyberspace

	Contents	Status	Examples	Example in Cyberspace
World 1	Material	Objective	Tables, airplanes	Hardware
World 2	Mental experiences	Subjective	Pain, happiness	Displays (the user experience)
World 3	Knowledge	Objective	Mathematics, physics	Software

Unlike material, knowledge can be used again and again and shared with many consumers without being diminished. Knowledge or information is a non-rival, partially excludable good. Paul Romer, a pioneer researcher in the new theory of economic growth, discusses the economic consequences of knowledge, and lays the foundations for a "different" knowledge-based economy.[4] He argues that growth in the economy, the basis of power and prosperity, is not solely a result of changes in capital and manpower. The development of knowledge is a new, potent source of endogenous growth. The character of this knowledge-based growth differs from what is familiar in the traditional economy.

If we combine Popper's metaphysical basis with Toffler's sociology and Romer's economic theory, we can suggest that the wars of the First and Second Wave were conducted mainly in World 1 ("material"). In these wars, the side with the largest and strongest army that was best able to mobilize troops and develop the mental factors (World 2) among its troops (e.g. the spirit of battle, motivation, and courage) would be victorious. According to this theory, future wars will also spread to World 3, the world of information. Without derogating the value of these elements in the future, while past wars relied on physical force (the First Wave) and present wars rely on the power of machinery (the Second Wave), future wars will rely more and more on brainpower.

Intellectual Approaches to National Security in the Information Age

The outstanding symbol of the information age – the electronic computer – was built at the end of WWII to help the US military in artillery ballistic calculations. In the decades following, especially after the invention of the transistor and the integrated circuit, computers have continually shrunk in size. Gordon Moore, co-founder of computer processors manufacturer Intel, stated in 1965 that the number of transistors that could be placed on an integrated circuit would double every 1-2 years, while the price would remain constant.[5] When this rule proved valid for semiconductors, the prediction was dubbed "Moore's Law." Futurist Ray Kurzweil presents persuasive arguments for extending Moore's Law to information technologies in general.[6]

With the development of the computer and its shrinking physical dimensions, defense institutions employ computing to improve the

performance of many systems. The chief benefit was a revolution in the accuracy of munitions, manifested first in airpower. Computers initially contributed to better operational planning. When it became possible to install a computer in warplanes, the power of computing was harnessed for the purpose of attack missions. An important strategic change occurred when the computer's dimensions and price were downsized enough that it could be embedded in ammunition itself. Thus was born the era of "smart weapons" – precision guided munitions that were initially adopted in aerial warfare. The operational results were stunning. In an attack on a specific individual target, such as a tank, one airplane armed with smart weapons can now do what 15 airplanes could do 30 years ago, or what 60 airplanes could do 40 years ago.[7] No wonder this technological revolution has had a decisive effect on the theory of warfare.

In order to adapt the art of war to information technology, a new theory of warfare dubbed "the Revolution in Military Affairs" (RMA) was developed in the early 1990s, based on four fundamental elements: precision strike, space power, dominant maneuver, and information warfare.[8] Information warfare involves several different aspects: computer warfare (computers are the main technological means of storing and transporting information), electronic warfare (mostly against sensors and communications systems), psychological warfare and managing the media (media briefings, embedding reporters in combat units, and manipulation of the information released to the public). These terms must be used accurately and the meaning of "information warfare" must be fully understood, particularly as these concepts have evolved with the advent and development of cyberspace.

The direct result of RMA is the absolute military superiority of the developed countries on the battlefield,[9] as reflected in the US wars in Iraq and Afghanistan, and in Israel's wars in Lebanon and against terrorist organizations. Indeed, a critical benefit of RMA is the unprecedented capability to conduct accurate and effective low intensity warfare, and the ability to defeat terrorism through military means, without causing widespread collateral damage.[10] As computer development continues, however, a change in approach is required. What follows is intended to provide a basis for an updated concept of national security in a reality that includes the new cyberspace.

Cyberspace

The ongoing growth of computers and communications networks generated a new situation at the beginning of the 21st century: an additional computerized layer above the existing older systems that effectively controls their function. The spread of computers, their integration in various devices, and their connectivity to communications networks have created a new space. Cyberspace is composed of all the computerized networks in the world, as well as of all computerized end points, including telecommunications networks, special purpose networks, the internet, computer systems, and computer-based systems. The concept also includes the information stored, processed, and transmitted on the devices and between these networks.[11] This picture enables us to understand what is happening in World 3[12] while focusing on the encounter with national security issues.

Unlike land, sea, air, outer space, and the electromagnetic spectrum, cyberspace is not a product of nature. Cyberspace is created by human beings, and would not exist without the information technologies developed in recent decades. Knowledge − which is perhaps the most important element in cyberspace − is a product of cumulative human endeavor.[13] The structure and design of cyberspace as it is today has significant consequences for national security (table 3).[14]

Table 3. Characteristics of Cyberspace and their Weak Points

Characteristic	Weak Point
Rapid change	Rapid obsolescence of means, including defense systems
TCP/IP protocol architecture	It is difficult to track the signal in the network and attribute it to a source.
High level of complexity	It is very difficult to connect an event to its cause, and difficult to distinguish a malfunction from an attack.
Extensive use of standard commercial off-the-shelf equipment	A narrowing gap between small and large players. The vulnerability of identical hardware and operating systems puts a broad range of systems at risk.
Entry-level cyber weapons are relatively cheap	The scope and price of defense is increasing.
An unclear legal environment	A gray area with a low probability of punishment encourages instability.

Cyberspace can be described as consisting of three layers.[15]

a. The most tangible layer, which currently provides the infrastructure of the computer world, is the physical layer. The physical components are the concrete building blocks of cyberspace – building blocks with natural characteristics: width, height, depth, weight, and volume.[16] In Popper's theory, the material layer corresponds to World 1.
b. The second layer is software logic, a variety of command systems programmed by people, intended to instruct a computing device. The physical components are controlled to a large extent by software, and the information stored on computers can be processed through software commands. The software layer is partly physical (World 1) and partly logical, meaning, again, World 3.
c. The third layer of cyberspace is the data layer that a machine contains and processes. The data and its processing generate information and knowledge. This layer is the least tangible of the three, mainly because the characteristics of information are very different from objective physical characteristics. This layer definitely belongs to Popper's World 3.

From Information Warfare to Cyber Warfare

In American and European professional literature,[17] information warfare is considered a significant feature of the information age. In American military terminology, information warfare is called "information operations," and its computerized component is called "computer network operations" (CNO).[18]

Table 4. Topics Included in Information Warfare

Topic	Relevant Systems and Technologies
Information collection	Various sensors in all parts of the electromagnetic spectrum
Transporting information for processing and the consumer	Broadband communications, compression, encoding, encryption
Storage and retrieval	Databases, de-duplication, compression
Processing and filtering information	Digital signal processing (DSP), automatic target recognition (ATR), data fusion, artificial intelligence (AI)
Making information accessible	Broadband communications, display systems, and a human-machine interface
Denial of information	Jamming, electronic warfare (EW), encryption, deception, obfuscation
Information protection	Denying unauthorized parties access to your information, encryption

Table 4 shows that the topics listed under information warfare are actually "classic" topics existing throughout the history of war. In the course of history, several classic methods of warfare have been developed for "information warfare," including intelligence gathering by human "sensors" (as in Joshua's use of spies in the conquest of the Promised Land) and the development of special gathering technologies (such as airborne intelligence sensors, satellites, etc.). Classic methods have also been developed in the prevention aspect of information warfare, such as camouflage, dummies and masks, jamming and blocking, deception and misdirection, propaganda, and so on.

Further analysis of table 4 indicates that the increasing dependence of information systems on computing is practically the only innovation in this field. In other words, while information warfare is not new, this is not true of computer-based information systems. Cyberspace makes it possible to define new targets, weapons, and methods of warfare. What is new about Third Wave warfare or war in the information age is not information warfare per se, but computer warfare. For this reason, it is best to limit the discussion by focusing on computer warfare in cyberspace. The change in cyberspace is so great that the basic concepts, such as "war," "weapon," "attack," and "defense," require a new explanation.

Computer warfare in cyberspace is unauthorized access to the adversary's computer systems for the purpose of intelligence gathering, disruption, deception, and prevention and delay of the use of information, while preventing the enemy from doing the same to one's own computer systems. A traditional attack (barrage, bombing, physical sabotage) on computer systems will also certainly cause disruption, prevention, and delay in the use of information. Such a physical attack, however, is not classified as cyberwar.

The characteristics of cyberspace[19] also define warfare in this sphere. The characteristics of cyberspace make it difficult to distinguish between a deliberate attack and malfunction, and complicate the effort to attribute action to a specific party, thereby also making it difficult to respond to an attack. The characteristics of cyberspace today empower marginal players, and give the attacker an advantage over the defender.

In recent years, a discussion has developed about the vulnerability created by the indispensability of cyberspace in all life processes in a developed society.[20] Computer warfare is not confined to military systems;

with the spread of computers and communications networks, it has become applicable to all areas of life. Most systems in the civilian economy and the entire critical infrastructure are now dependent on computers, and are part of cyberspace. This fact generates vulnerability and new possibilities for warfare, and also requires defensive preparation in developed countries.

Attack and Defense in Cyberspace

Cyber weapons[21] are malware and harmful hardware that damage the victim's computer resources and disrupt his data, deceive, and cause deprivation of service or the collection and transfer of intelligence. "Malware" is hostile software designed to disrupt orderly activity of a computer system and damage the process managed by that system. "Spyware" is hostile software designed for covert data collection and its potential transmission over a network. "Phishing" is a stratagem based on software and social engineering designed to fraudulently obtain personal data and details of user identities to gain unauthorized access to sensitive resources.

Hardware can be implanted through the addition of an electronic component to an existing unit, or an addition within an integrated circuit. The implant can take place during manufacture, transportation, operation and maintenance.[22] The use of software as a logical weapon, more common than the use of hardware, is what enables the most advanced methods of warfare. Knowledge and technology are non-rival, partially excludable goods; these inexhaustible characteristics make them hugely important in all matters pertaining to information warfare. Not all the consequences of this potential have been fully clarified.[23]

When there are good grounds to suspect that a cyber attack is underway, it is very difficult to identify the source and the attacker's identity. All parties operating in cyberspace use common tools and methods. Commercial cooperation, a kind of outsourcing, frequently takes place between the technical parties possessing attack capabilities (programmers, encoding hackers, owners of "captive networks") and those ordering the services (private investigators, organized crime, espionage organizations). In order to determine that a cyber attack is an act of war, several aspects must be examined:

a. The organizational and geographic source: whether a state is behind the action[24]

b. Motive: whether it is possible to identify an ideological, political, economic, or religious motive for the attack
c. Level of complexity: whether the attack required complex planning and coordinated resources that are available primarily to state agencies
d. Results: whether the attack caused damage and casualties, and whether it would have caused damage without the defensive actions taken.

The characteristics of cyberspace make it difficult to answer these questions, and answers sufficient for setting policy will undoubtedly be lacking.

For adequate defense, it is necessary to discern there is an attack, which is no simple matter in cyberspace. Early implantation of malicious hardware or software, especially before testing plans have been formulated, reduces the chances of detection. More accurate cyber weapons cause little collateral damage, which makes detection of the attack by the victim less likely. Defensive actions involve three aspects:[25]

a. Detection: the Achilles' heel – how to realize that a computer attack has taken place
b. Prevention: a means of stopping the attacker at the penetration stage
c. Response: recovery measures to limit the attacker's achievements, forensic means, and even retaliatory action.

Key Issues in Cyberwar

The technological change underlying the transition to the Third Wave, the rapid expansion of World 3, and the development of the information economy raise new questions. One of the most important is the debate on critical infrastructure protection. The feasibility of a cyber threat to the infrastructure of a modern society was presented through experiments, such as a power generator being put out of action and blown up by broadcasting commands to its command and control system.[26] It appears that this threat became a reality in the summer of 2010, when the Stuxnet worm virus that infected "Windows"-based computers was discovered. It searched for computers running Siemens-produced industrial command and control software of a certain type connected to an industrial controller of a specific model. Only if it located the relevant computers, the virus activated software code that disrupted the activity of the computerized controller, while concealing the change from the control software and equipment operators. Stuxnet allegedly damaged the proper operation

of the centrifuges for uranium enrichment in Iran. The source and duration of the attack are unknown.[27]

The US, the world's only superpower, is a pioneer and leader in the discussion of its cyber vulnerability.[28] A country's critical infrastructure is an obvious target in any conflict. Nonetheless, why has such concern been raised now, and in the strongest countries? The answer lies in the transition from the wars of Toffler's Second Wave to the wars of the Third Wave, the information wave. Discussion of critical infrastructure protection has been renewed because of the emergence of a new threat that could not have been carried out before. The development of cyberspace makes it possible, for the first time in history, to attack critical infrastructure systems in cyberspace, without physical access to the site and without exposure during or after the attack.

Critical infrastructure protection is one of the key issues of cyber security. The topic is outside the scope of this study, and deserves a specific discussion of its own.[29]

"Information warfare" immediately invites examination of the concept of war itself: is a cyber attack on computerized information involving no use of firepower an act of war? What constitutes a legitimate target in such a war? The extensive military use of civilian infrastructure (mainly communications) complicates the distinction between military and civilian targets. For example, the computer infrastructure of the US Department of Defense consists of 15,000 networks and seven million facilities dispersed all over the world. Most of the US Defense Department communications, however, are channeled through commercial civilian networks.[30] Civilians (even women and children) can be as effective as soldiers in computer warfare. Does this make them potential targets of a response? How should we act in a case of widespread economic damage? Moreover, the meaning of such an attack is unclear. Assume that one day the computer systems of the Israeli banks crash. Assume also that we manage to determine with certainty that the enormous damage was caused deliberately by a deliberate penetration, and assume that we succeed in tracing the attacker to the territory of a neighboring country. Now, is this an act of war? Ostensibly, the damage caused is "only" economic; there are no (direct) human casualties. Countries have frequently responded with restraint to traditional attacks that caused economic damage but did not take human life.[31] Economic damage, however, is liable to paralyze

an entire country. How do we estimate the indirect damage caused by an attack? Assume that a cyber attack caused prolonged disruption in the supply of electricity. Assume that one of its results is putting road lights and traffic lights out of commission, and the resulting darkness causes fatal traffic accidents. Should a victim of such an accident be considered a cyber warfare casualty? Should we respond with firepower and ground maneuver, or with a cyber counterattack? The problem is more complicated than the scenarios described, because a computer attack does not require a base in a country, and it can also be conducted by organizations and even by individuals.

Computer warfare is also conducted between friendly countries competing for diplomatic and economic intelligence. Is this "warfare?" Is it acceptable or advisable to use computer warfare in peacetime for such purposes?

A special problem in cyber warfare is detecting an attack; in contrast to a traditional attack occurring in World 1, the material world, the location of the strike and the attacker's identity are not necessarily exposed following the attack. There are no defined "front lines" in computer warfare, and geographic distance has almost no meaning in electronic networks. Given the characteristics of cyberspace, detecting an attack cannot be taken for granted: an attack and a malfunction have similar effects. While the computer world has become more sophisticated, as reflected in the multiplicity of software and applications and the growing number of transistors in each component, malfunctions are not less likely. The statistical probability of a software "bug" or programming error is constant, and its nominal value rises with increased complexity of software.[32]

The capability to detect that computers have been attacked and damaged, rather than malfunctioning "naturally," is inadequate. Without the ability to distinguish in real time between an attack and malfunction, large scale investment in constant cyber readiness is necessary. Defense against cyber threats must encompass all aspects of attack and be updated with new developments, and its cost is rising steadily. The argument on difficulty of defense is similar to the argument against an active anti-missile defense and the argument that defense against suicide terrorists is futile. Nevertheless, it is possible to devise a response to the new threats,[33] although the burden is substantial, since the characteristics of today's cyberspace give a clear advantage to attack over defense.[34] The field of

encryption is one of the few areas in cyberspace in which the defender still enjoys an advantage over the attacker.[35] Given the difficulty of identifying the fact of an attack, its geographic location, and the identity of the attacker, a state of uncertainty results that makes an escalating response difficult. Table 3 above summarizes the characteristics and many weak points that create the "attribution problem": it is hard to know the attacker's source and identity and on behalf of whom he operated, and it is certainly hard to prove guilt. In the traditional defense realm, great effort is expended on intelligence, advance warning, and deterrence in order to limit as much as possible the resources spent on a state of continual readiness. The problem of deterrence is particularly difficult in cyberspace, mainly because of the attribution problem.[36]

The characteristics of cyberspace give rise to problems for an attacker as well. How can one tell whether the cyber-attacked computers have really been damaged? In order to rely on a cyber attack, battle damage assessment is necessary. From this perspective, an open loop attack, i.e., one whose degree of success is unknown, is of limited utility. This problem is especially acute if the cyber attack was not intended to destroy data but to manipulate it.

In conventional warfare, rules have been developed that are anchored in international conventions. These conventions, which were written before the emergence of cyberspace, deal in "armed conflict," "physical confrontation," "territorial attack," and so on. These concepts are irrelevant to computer warfare, and the existing conventions require adaptation to cyber warfare – Third Wave warfare. Despite widespread research in this field, it is reasonable to assume that an assessment of the issues from a legal standpoint will take many years. The absence of rules makes it difficult to cope on a daily basis with the special problems of cyber warfare. The issues reviewed are not purely legal; they are essential issues for policymaking and taking decisions. In late 2011, NATO was in the midst of formulating a legal framework to enable it to respond to cyber attacks using methods currently of uncertain legality. An understanding of the theoretical foundations of the field is critical for improving the ability to cope with it.

Conclusion

Cyberspace is a fairly new product of the information age, and cyber security is part of the transition to the information age. In order to cope

with this challenging change, a multidisciplinary perspective should be adopted. Therefore some of the information age's important theoretical origins were presented, including ideas of the Tofflers, Karl Popper, and Paul Romer. Clearly there are other sources, and further multidisciplinary research on the information age is welcome.

The problems in dealing with security challenges are a function of the characteristics of cyberspace: rapid action, the rate of change, intricacy, and complexity. Cyber attack and defense take place in World 3, the world of knowledge. The significant consequences of the key issues of cyber warfare described in the last section of this study should be investigated in depth.

The key development is not "information warfare"; it is computer warfare in cyberspace. Discussion of solutions to "computer matters" tends to focus on the technical realm, far away from public debate and public policy. Clearly professional understanding of the field under discussion is essential, and it presents enormous challenges requiring a solution at the national public policy level. However, a review of the main issues of cyber security paints a complicated picture, beyond the technical computer professions. In order to provide national security in the dynamic environment of the information age, it is therefore correct to utilize inputs from every relevant field of knowledge, including the social sciences, psychology, biology, medicine, and philosophy. This study aims to encourage interdisciplinary research into the cyber security challenges, contribute to the development of an informed national security policy, and thereby contribute to security and prosperity in the information age.

Notes

1 "The Meaning of Stuxnet: A Sophisticated 'Cyber-Missile' Highlights the Potential – and Limitations – of Cyberwar," *Economist (GBR) Economist* 397, no. 8702 (2010), September 30, 2010, from the printed edition.

2 Information or data is distinguishable from knowledge, which also requires conceptualization and understanding of the raw information. This distinction is unimportant for the purposes of this article.

3 Karl Popper, *Objective Knowledge: An Evolutionary Approach* (Oxford: Oxford University Press, 1972), chapters 3-4.

4 Paul M. Romer, "Endogenous Technological Change," *Journal of Political Economy* 86, no. 5, pt. 2 (1990): S71-S102.

5 E. Mollick, "Establishing Moore's Law," *Annals of the History of Computing, IEEE* 28, no. 3 (2006): 62-75.

6 Ray Kurzweil, "The Law of Accelerating Returns," (2001).

7 Isaac Ben-Israel, "From Sword Blade to Computer Memory," *Odyssey 9*, October 2010.

8 For information on RMA, see: Michael E. O'Hanlon, "Technological Change and the Future of Warfare" (Washington, DC: Brookings Institute Press, 2000); Stuart E. Johnson and Martin C. Libicki, "Dominant Battlespace Knowledge: The Winning Edge" (Washington, DC: National Defense University Press, 1995).

9 This ascendancy has caused the enemy to retreat to a strategy of survival and asymmetric warfare.

10 This capability was demonstrated for the first time in Israel's victory in 2000-2005 over Palestinian suicide bombers during the intifada. See Lior Tabansky, *The Anti-Terrorism Struggle in the Information Age: Palestinian Suicide Bombers and the Implementation of High Technologies in Israel's Response, 2000-2005*, position paper published by Tel Aviv University, May 2007.

11 The great resemblance between the American and Israeli definitions is a result of shared values and a similar scientific and economic level. China, Russia, India, France, and other countries define cyberspace and cyber threats differently. Such a comparison, however, falls outside the bounds of this study.

12 See the discussion above of Karl Popper's theory.

13 A discussion of the status of knowledge appears in Karl Popper, and was mentioned in the preceding section.

14 For a discussion of cyberspace in the context of national security, see Lior Tabansky, "Basic Concepts in Cyber Warfare," *Military and Strategic Affairs* 3, no. 1 (2010): 75-92.

15 Martin C. Libicki, *Cyberdeterrence and Cyberwar* (Santa Monica, CA: RAND Corporation, 2009).

16 Today, the infrastructure of the computer world is electronics. Before electronics, there were mechanical calculators. And in the future? The practicality of utilizing biological infrastructure for computational purposes has already been demonstrated. DNA computing uses molecular biology and DNA instead of electronic components. Another possibility is peptide computing: bio-molecular computing based on amino acid compounds.

17 Compare the definitions of the US Defense Department, "Joint Publication Jp 3-13: Joint Doctrine for Information Operations," edited by United States Department of Defense, Washington, DC, 2006, with those of the European Union as defined in the tender of the European Defence Agency Study, "Computer Network Operations (CNO) for EU-led Military Operations," 10-CAP-OP-37 (EU Milops CNO Capability) – Annex, August 16, 2010.

18 This includes computer network defense (CND), computer network exploitation (CNE), and computer network attack (CNA). The technical basis for the three types of action is identical.

19 See table 2 above.

20 For example, see Richard A. Clarke and Robert K. Knake, *Cyber War: The Next Threat to National Security and What To Do About It* (New York: Ecco, 2010); Franklin D. Kramer, Stuart H. Starr, and Larry K. Wentz, *Cyberpower and National Security* (Washington, DC: Center for Technology and National Security Policy, National Defense University Press: Potomac Books, 2009); William Lynn III, "Defending a New Domain," *Foreign Affairs* 89, no. 5 (September-October 2010); Martin Coward, "Network-Centric Violence, Critical Infrastructure and the Urbanization of Security," *Security Dialogue* 40, no. 4-5 (2009): 4-5; Walter Gary Sharp, "The Past, Present, and Future of Cybersecurity," *Journal of National Security Law and Policy* 4, no. 1 (2010).

21 For a discussion of the technical issues, see Jeffrey Carr, *Inside Cyber Warfare: Mapping the Cyber Underworld* (O'Reilly Media, 2009); and Rick Lehtinen, Deborah Russell, and G. T. Gangemi, *Computer Security Basics* (Sebastopol, CA: O'Reilly & Associates, 2006).

22 Faulty hardware implanted by the CIA in a system for transporting gas purchased by the Soviet Union allegedly caused an enormous explosion in Siberia in 1982. See W. K. Clark and P. L. Levin, "Securing the Information Highway: How to Enhance the United States' Electronic Defenses," *Foreign Affairs* 88, no. 6 (2009).

23 For the economic consequences, see the discussion by Paul Romer mentioned above.

24 Following the September 11, 2001 terrorist attacks, the policy support threshold was lowered: sometimes circumstantial evidence, such as ideological support of an enemy or provision of logistic services to terrorists, is sufficient.

25 A detailed discussion of these matters is beyond the scope of this study.

26 "The Aurora Experiment," conducted in the national laboratories in Idaho, US; See James Andrew Lewis, "Thresholds for Cyberwar," Washington, DC: Center for Strategic and International Studies, 2010.

27 "The Meaning of Stuxnet," note 1.

28 United States, President's Commission on Critical Infrastructure Protection, "Critical Foundations: Protecting America's Infrastructures: The Report of the President's Commission on Critical Infrastructure Protection," Washington, DC: US GPO, 1997.

29 See Lior Tabansky, "Critical Infrastructure Protection against Cyber Threats," *Military and Strategic Affairs* 3, no. 2 (2011): 61-78; Myriam Dunn, "Securing the Digital Age: The Challenges of Complexity for Critical Infrastructure Protection and IR Theory," in Johan Eriksson and Giampiero Giacomello, eds., *International Relations and Security in the Digital Age* (Routledge, 2007).

30 Lynn, "Defending a New Domain."

31 Israeli governments behaved in this manner for years, when thousands of rockets "trickled" into Israel from Gaza and hit open areas in the western Negev.

32 One of the measures of software complexity is the number of source lines of code (SLOC). *Windows NT 3.1*, the Microsoft operating system, which was introduced in 1993, had 4.5 million SLOC. *Windows XP*, introduced in 2001, had 45 million SLOC. Linux Fedora 9 has 204 million SLOC.

33 See Tabansky, "The Struggle against Terrorism in the Information Age."

34 Ibid., and Lynn, "Defending a New Domain."

35 The dominant encryption method is based on a mathematical principle that it is difficult to factor a number whose factors are prime numbers. Quantum computing has features that will completely eliminate the advantage of the existing encoding methods. When a quantum computer is built, the security field will undergo an upheaval caused by the foundations of encryption being made obsolete.

36 Libicki, *Cyberdeterrence and Cyberwar*. See also Amir Lupovici, "Cyber Warfare and Deterrence: Trends and Challenges in Research," *Military and Strategic Affairs* 3, no. 3 (2011): 49-62.

Cyberspace and Terrorist Organizations

Yoram Schweitzer, Gabi Siboni, and Einav Yogev

In a scene in the 1990 movie *Die Hard 2*, terrorists take control of computer, traffic control, and aerial communications systems, impersonate flight inspectors, and feed in false data, thus leading the pilot and passengers to their death in the midst of a snowstorm with the plane crashing on the runway. Security personnel are helpless, incapable of providing a response; the movie's hero, John McClane (played by Bruce Willis), lacks the means to save the doomed flight and is left standing powerless in the fog on the landing strip, waving two improvised beacons at the approaching aircraft. At first it would seem that the movie is nothing but another Hollywood fantasy, dismissible as a wild exaggeration carried to yet further extremes in the sequel, *Die Hard 4*. However, the events of 9/11 and the changes in the nature of security threats over the last decade indicate that even the most far-fetched scenarios crafted in Hollywood studios are liable to find real-life expression in the public and security sphere in this day and age.

The use of cyberspace as a primary warfare arena between enemies or hostile nations has always been fertile ground for fantasy and lurid scenes on the silver screen. However, cyberspace is rapidly becoming a genuine central arena for future wars and hostile actions undertaken by various types of adversaries. These may include terrorist organizations, although until now they have relied primarily on physical violence to promote their own goals and those of their sponsors. In light of such threats, many nations in the West have in recent years established special authorities to

Yoram Schweitzer is head of the Terrorism and Low Intensity Conflict Program at INSS. Dr. Col. (ret.) Gabi Siboni is head of the Military and Strategic Affairs Program at INSS and head of the Cyber Warfare Program at INSS, which is supported by the Philadelphia-based Joseph and Jeanette Neubauer Foundation. Einav Yogev is a research assistant at the Terrorism and Low Intensity Conflict Program at INSS.

use innovative technological means to prepare for war-like actions against strategic infrastructure targets.

This essay focuses on an analysis of the factors that are likely to make terrorist organizations use cyber tools to perpetrate attacks on critical infrastructures of sovereign institutions and symbols, commercial and industrial infrastructures and systems, and public civilian targets. In addition, it examines the question of whether the threat is actual and imminent, or whether it is a far-fetched possibility that surfaces from time to time in the general discourse on the subject.[1]

The Cyber Threat from Terrorist Groups

Today there are five main groups that use or have the potential for future use of cyber attack tools: 1) states developing offensive and defensive capabilities as a growing part of their force capabilities; 2) criminal elements motivated primarily by illegal commercial interests; 3) commercial companies, primarily in the defensive mode (as the scope of cyber attacks in the commercial context is significantly growing), though some may resort to offensive moves against competitors; 4) terrorist organizations, out of cost-benefit considerations and other inherent advantages, are liable to try to carry out cyber attacks; and 5) anarchists opposed to the existing establishment who are interested in undermining it from within and without, and who endeavor to attack the entire system of computerization, which today is the basis for managing life as we know it, in order to disrupt or even destroy states' current social order and their fabric of life.

Cyber offense has the potential to change society's balance of power because it empowers those engaged in asymmetrical conflicts that operate from a position of inferiority, especially terrorist organizations. Capabilities in this sphere may enable them to attack installations, systemic processes, and sites while causing heavy physical damage and wielding a significant psychological impact on the society and public under attack. They thus acquire capabilities other than those familiar from conventional terrorist attacks, such as suicide bombings, booby traps, hostage situations, hijackings, and kidnappings.

Cyber offense affords several advantages. First, it removes the necessity of physical presence at the target. It is possible to damage communications networks and control systems of installations and processes from afar and thus avoid physical barriers and human systems. Second, it affords

a wider scope of damage. Cyber attacks occur not only in the physical space but also carry the potential for severe and sustained damage to control and infrastructure systems. Thus, while most conventional terrorist attacks are limited in time and space,[2] a cyber attack magnifies terrorism's psychological impact through fear and intimidation. Third, it is easier to conceal the identity and source of the attack; in cyberspace, identities and boundaries between states are more easily blurred. Terrorists attacking in cyberspace can not only conceal their identity but can also feed false information as to the source of the attack, for example, by attacking a site inside the target state using addresses of a friendly nation. Fourth, cyberspace attacks are cost effective. Using the cyber platform for attacks maximizes the cost-benefit ratio from the perspective of a terrorist organization, endowed with fewer resources and capabilities than the states it targets. Assuming that terrorist organizations would prefer less defended targets rather than well-protected ones, they presumably would be able to gain access and insert malicious code into target sites, or use technologies that are becoming ever more accessible to wider audiences. Fifth, cyber terrorism can be non-lethal. It can cause significant damage without direct fatalities or physical injury, granting terrorists success by means of intimidation and disruption of the routine. This gives the perpetrators the ability to devise a defense and logical explanations for their deeds, which after all did not spill blood but were only an indirect cause of lost lives. The innovativeness represented by such action would also garner terrorist organizations widespread media coverage and enable them to engage in non-lethal threats in which a price would be extorted in exchange for removing the threat of a cyber attack.

It has been claimed that terrorist organizations are not interested in cyberspace because they prefer showcase attacks with much higher visibility rather than the anonymity that supposedly is conferred by attacks in this domain.[3] However this claim does not take into account the basic rationale of terrorism strategy, which holds that terrorist activity should focus on minimizing the power differential in the struggle against a stronger enemy with more powerful means, carry out destructive actions while identifying the weaknesses in the enemy's defense, and achieve a position of superiority at tolerable costs given the relatively poor means at the disposal of the perpetrators. Already today global jihad terrorist organizations are making use of cyberspace, though still in limited and

relatively undeveloped fashion, to realize these advantages. A study examining the cyberspace warfare capabilities of jihadist organizations[4] identified a number of major features that serve to build and improve the organizational and operational infrastructures of terrorist organizations in the following fields:

a. Propaganda: using the web to disseminate ideas, decrees, directives, speeches, and opinion pieces by clergy and terrorist leaders.
b. Recruitment and training: using the web to identify and recruit potential members as well as to transmit instructional and training materials.
c. Fundraising and financing: using the web to fundraise under the guise of charities and aid organizations as well as to steal identities and credit cards.
d. Communications: using the web for operational communications while employing a range of tools, including accessible encryption tools.
e. Identifying targets and intelligence: using information available on the web to identify targets and gather intelligence.

It is thus clear that an essential upgrade of cyberspace tools available to terrorist organizations, from logistical and propaganda tools to actual operational tools, is liable to generate an innovative, dramatic, and relatively cheap type of attack with the power to effect severe damage, even if carried out with a low signature or in total anonymity. Therefore every terrorist organization, especially one seeking fame and wanting to affect the public psyche and morale in the targeted enemy, sees such an attack as an important and worthy challenge. Innovation would also guarantee the perpetrators international fame and transform them into role models. Thus, sub-state entities with more limited technological capabilities than the nations with which they are at war are liable to join the trend of using advanced technology needed for cyber warfare for their own benefit, either by receiving assistance from supportive nations or by acquiring such capabilities themselves in the future, by recruiting and operating individuals with the necessary skills in this field.

As for states supporting terrorism, cyberspace is very attractive for use of proxy organizations because of the anonymity afforded by the domain, the difficulty in proving the identity of the perpetrator, the high level of deniability by states about their involvement, and the satisfaction of causing severe damage to the enemy. Even if suspicions are aroused, it is still hard to prove guilt. Furthermore, the public under attack may

perceive a cyber attack to be less outrageous than a terrorist attack that employs firearms and causes direct death and destruction – even if the damage caused is greater, more destructive of property, and takes more lives than a violent terrorist act.

Despite these advantages of cyber attacks, to date no such attack has been traced to a terrorist organization. Development of significant capabilities in this field requires surmounting a considerable intelligence and technological threshold. At this stage one may assume that terrorist organizations find it hard to identify, harness, and maintain such high technological capabilities and access that would allow them to cross that bar. It is true that this limitation can be partially overcome through the assistance of state supporters of terrorism, but at least for now this is not enough to give terrorist organizations the significant, stable technological platform required for maintaining effective cyber attack capabilities. In addition, terrorist organizations face limitations posed by cyber surveillance and state intelligence and technological capabilities that enable them to identify suspicious conduct on the web, identify attempts at organization, and mount a defense against them and against threats to specific targets.

Weaknesses and Responses

Although to date terrorist organizations have not been able to overcome the difficulties in achieving offensive cyber capabilities, civilian systems and routine civilian life presumably remain their preferred targets, because these are much more difficult to protect than security systems. Strengthening defenses of critical national infrastructures such as electric, water, and communications supply networks would likely encourage terrorists to seek out less protected targets in the civilian and commercial sectors. Even though systems in these sectors are usually not included in the rubric of critical and protected infrastructures, from the terrorist perspective an attack against them could be effective, by breaching ordinary citizens' basic sense of security and enhancing the terrorists' image by instilling fear.

A significant part of constructing a defense against cyber attacks is general and independent of the source of the threat, whether terrorist, state or criminal. This is reflected organizationally – consider Israel's Information Security Authority and ministries specializing in cyber defense in various

nations – and also in certain components of defense from the fields of information systems and general security. In contrast, in fighting terrorist organizations it is also necessary to activate two designated components that require sustained development and improvement.

The first is intelligence. Effective gathering of accurate, high quality intelligence requires using a range of sources, including open sources and material from the terrorists' own computers and networks. To this end it is necessary to develop capabilities of infiltrating these systems covertly and inserting information effectively and continuously. The challenge that must be overcome is the widespread global deployment typical of terrorist organizations that use many chat rooms and transmit messages using unique code words. Intelligence agencies must be able to intercept these transmissions and decode them within the relevant timeframes and at the same time provide cyber defense systems with the tools needed to protect against and even disrupt the planned actions.

The second component is disruption. Unlike defense systems, which do not try to prevent an attack but rather obstruct its success once it has already been launched, the goal of disruption is to thwart the execution of the attack or to hamper its progress. Establishing an effective disruption structure against cyber attacks by terrorist organizations requires intelligence monitoring and control that can identify the organization of an attack before it takes place and operate effectively to foil it. This aspect relies primarily on tactical intelligence gathering capabilities, both from computers and from communications networks used by terrorist organizations.

Disruption attempts can also be directed towards damaging the organizational infrastructures of the organization. An example of this occurred in England when British intelligence hacked the online issue of the British al-Qaeda magazine *Inspire*. In addition, in recent years the various components of the electronic jihad have been targeted for occasional cyber attacks largely attributed to Western governments: the Taliban's website has been hacked time and again, as have exclusive jihadist forums and high profile fundamentalist websites. Meanwhile, American, Saudi Arabian, and Dutch authorities have extracted valuable information about potential Islamic terrorism from jihadist websites serving as honey traps for high quality intelligence.[5]

At the same time, it is necessary to deepen the defenses of civilian systems that represent the greatest weakness and therefore are also preferred terrorist targets. For example, the British government began taking legislative steps that include authorizing the use of invasive techniques such as telephone wiretaps, surveillance of emails in police files connected to crimes of terrorism, torpedoing internet radicalization processes, and specialized training of police units to confront cyber threats.[6] Nonetheless, in most states the defense of civilian systems is still in its infancy. Most states' cyber defense resources are allocated to security systems and to what are considered critical national infrastructures. Deepening the defense of civilian systems requires radical changes on a national scale that must be supported by appropriate regulation.[7]

Conclusion

In December 2001, at a meeting in New York shortly after the 9/11 attacks, the philosopher Jacques Derrida presented his understanding of the changes generated in the world as a result of those events. According to Derrida, the attacks were still part of the "archaic theater of violence," the real, visible world, in which events are still conducted in "clear and great order." However, according to him, cyberspace presents us with a more potent threat to our political and physical world; the dangers inherent in it change the relationship between terrorism, in the psychological and historical sense of a violent attack, and the concept of territory. Now, in the new techno-scientific world, the threat we knew in the past as real has become an invisible, quiet, and swift threat, devoid of bloodshed, which, according to Derrida, is worse than the 9/11 attacks, which at least were directed against a known location at a particular point in time. Now we are facing a challenge that threatens the social and economic fabric of life that connects all of us and upon which all of us depend in every place and at every moment.[8]

The rapid technological developments and innovations of recent years in the domain of cyberspace have indeed created a battlefield that simultaneously brings together many varied populations, local and international, representing a desirable target and fertile ground of activity by sub-state entities. Since thus far there has been no known cyber attack perpetrated by a terrorist organization, the threat does not seem acute. The challenge facing those who would try to use cyberspace for malicious

purposes is three-pronged: attaining high level intelligence, the ability to crack computerized systems protected with advanced technology (or accessibility to such ability), and very high levels of calculation and computerization skills.

However, the advantages afforded by attaining cyberspace capabilities as described in this essay are liable to serve as an incentive for terrorists to develop, acquire, or harness such capabilities in the future. Gaining control of the advanced technological and intelligence capabilities required in cyberspace is likely to give these elements who seek to seriously damage their enemies by causing massive destruction and sowing terror and intimidation in the public at large the ability to disrupt the normal routine of civilian life, undermine civilian trust in their governments, and of course gain valuable prestige and media stature.

Therefore, Western nations must work diligently to meet this threat and improve the effective intelligence and defensive capabilities of civilian systems, while at the same time construct accurate intelligence gathering capabilities and the ability to disrupt cyberspace organization and attack by terrorists. Neglecting the civilian cyberspace domain, which is an attractive target for terrorists, is liable to prove disastrous in the future and place security personnel, when the time comes, in the same position as that fictional Hollywood hero of *Die Hard 2* trying to save airplanes from crashing using nothing other than improvised beacons.

Notes

1 The use of the term cyber terrorism in this essay refers to the use of cyber tools liable to be used by terrorist organizations to attack economic infrastructures and civilian systems in targeted nations.

2 There are of course important exceptions: the 9/11 attacks in the United States had a global effect on flight security systems.

3 Shmuel Even and David Siman-Tov, *Cyber Warfare: Concepts, Trends, and Implications for Israel*, Memorandum No. 109 (Tel Aviv: Institute for National Security Studies, 2011, p. 42).

4 *Examining the Cyber Capabilities of Islamic Terrorist Groups*, Institute for Security Technology Studies at Dartmouth College, Technical Analysis Group, March 2004.

5 Adam Rawnsley, "Stop the Presses! Spooks Hacked al-Qaida Online Mag," *Wired*, June 3, 2011, http://www.wired.com/dangerroom/2011/06/stop-the-presses-spooks-hacked-al-qaida-online-mag/June 4, 2011.

6 "Warning of Rise in Cyber-terrorism," *The Independent*, July 12, 2011, http://www.independent.co.uk/news/uk/crime/warning-of-rise-in-cyberterrorism-2312434.html.

7 Gabi Siboni, "Protecting Critical Assets and Infrastructures from Cyber Attacks," *Military and Strategic Affairs* 3, no. 1 (2011): 93-101, http://www.inss.org.il/upload/(FILE)1308129638.pdf.

8 Jacques Derrida, in Giovanna Borradori, *Philosophy in a Time of Terror: Dialogues with Jürgen Habermas and Derrida* (Hebrew translation, United Kibbutz Press, 2004), pp. 173-74; also available (in English) at http://www.press.uchicago.edu/Misc/Chicago/066649.html: "One will be able to do even worse tomorrow, invisibly, in silence, more quickly and without any bloodshed, by attacking the computer and informational networks on which the entire life (social, economic, military, and so on) of a 'great nation,' of the greatest power on earth, depends. One day it might be said: 'September 11' – those were the ('good') old days of the last war. Things were still of the order of the gigantic: visible and enormous! What size, what height! There has been worse since. Nanotechnologies of all sorts are so much more powerful and invisible, uncontrollable, capable of creeping in everywhere. They are the micrological rivals of microbes and bacteria. Yet our unconscious is already aware of this; it already knows it, and that's what's scary."

Cyber Warfare and Deterrence: Trends and Challenges in Research

Amir Lupovici

In recent years a growing number of researchers have expanded the discussion of deterrence strategy to a host of new threats. Unlike the Cold War era in which the study of deterrence focused primarily on deterrence among nations and superpowers and on nuclear deterrence, recent years – particularly since 9/11 – have seen much research on deterrence strategy in relation to other threats, such as terrorism, rogue states, and ethnic conflicts. These studies share several elements: they are based primarily on an effort to examine the relevance of conditions necessary for successful deterrence, formulated in the context of the Cold War, and to a large degree are policy oriented, particularly regarding the challenges confronting the United States.[1] These same elements dominate the evolving debate on the connection between deterrence and cyber warfare.[2] Much of the research on deterrence strategy and cyber warfare is based on an American perspective. It examines the possibility of successfully implementing the strategy of deterrence in order to prevent cyber attacks, or analyzes the way the US can use cyber warfare in order to deter other threats it faces.[3]

These studies make it clear that the possibility of successful deterrence against cyber attacks is limited with regard to each of the dimensions required for its success: the existence of capability (weapons), the credibility of the threat, and the ability to convey the threatening message to the potential challenger.[4] Nonetheless, there are several elements to consider that under certain circumstances are likely to serve as the basis for successful deterrence even in the realm of cyberspace. This essay surveys the literature and proposes directions for continued research on the topic.

Dr. Amir Lupovici is a lecturer in the Department of Political Science at Tel Aviv University.

The essay begins by presenting the necessary conditions for a successful strategy of deterrence. It then reviews the central claims regarding the difficulties in applying successful deterrence in cyber warfare vis-à-vis each of these conditions. The third part discusses some benefits and shortcomings of certain factors that may strengthen deterrence against cyber warfare. Finally, it highlights the importance of continuing the discussion of deterrence and cyber warfare, indicating a number of directions for future research.

The Conditions for Successful Deterrence

There are different ways in which actors can try to prevent their enemies from taking undesirable action. The strategy of deterrence by punishment is one of the most studied. This type of deterrence has several definitions,[5] with the definition by George and Smoke, whereby deterrence is the ability to persuade a potential enemy that the price it will pay as the result of carrying out the undesirable action will outweigh any possible profit, is among the most commonly used.[6] This type of deterrence differs from deterrence by denial,[7] which is based on the attempt to persuade potential aggressors that they must avoid taking action because they will fail to attain their goals.[8] The concept of deterrence also differs from the concept of compellence, which is based on the use of threats in order to make an enemy undertake an action, whereas the aim of deterrence is to make the enemy avoid taking undesirable action.[9]

A central question regarding the strategy of deterrence by punishment concerns the conditions under which it is likely to be successful, i.e., cause a potential enemy to avoid challenging the defender. The research, developed mostly during the Cold War and dealing with deterrence between the superpowers, focuses on three central conditions: the defender's capabilities, the credibility of the threat, and relaying the threat message to the challenger.

The first essential condition for successful deterrence by punishment is that the defender be able to exact a price from the challenger. It is therefore not surprising that studies in deterrence arose in particular during the nuclear era, as this weapon allowed both sides to make the cost of a future war very clear. Nuclear weapons gave leaders a crystal ball of sorts, allowing them to see the effects of the next big war and thus encourage them to exert caution in their conduct.[10] At the same time, capabilities are

not limited to the non-conventional, as conventional means too may be used to take a toll on the challenger.[11] Moreover, an important part of the capabilities dimension is the means of delivery available to the defender, such as aircraft, missiles, and even roads and vehicles that may play a role in the element of capabilities within the context of deterrence.

A second condition for successful deterrence is the credibility of the threat. In order for the deterrence threat to be effective, the defender must be ready to use the capabilities at its disposal. Various researchers have presented a range of factors that may limit this willingness, e.g., internal or international public opinion, or even the deterrence capabilities of the enemy (the challenger).[12] Common to all these elements is that each in its own way raises the cost of taking action, thereby reducing the actor's credibility in terms of carrying out the threat, if necessary.[13]

The third condition is effective delivery of the messages to the challenger concerning the two previous conditions – capabilities and intentions. In other words, the challenger must be aware of the defender's capabilities and its willingness to use them. Researchers who have developed psychological approaches to deterrence claim that this condition is the most important of all, whereby the perceptions and misperceptions of decision makers directly affect the success of deterrence.[14] In this sense, what matters are neither the capabilities nor the intentions of the defender, rather how they are perceived by the potential challenger.

Finally, because the strategy of deterrence may prevent different types of threats, it is difficult to discuss the conditions for successful deterrence uniformly, as they must be adapted not only to the challenger but also to the type of action the defender is trying to prevent. So, for example, while nuclear weapons may be effective in deterrence against an all-out attack ("general deterrence"), its effectiveness would be lower against more limited types of threats.[15]

Difficulties of Deterrence in Cyber Warfare

Many of the studies analyzing the strategy of deterrence against cyber warfare are based on Cold War theories. Researchers analyzed the central conditions for successful deterrence discussed in the literature: defensive capabilities, the credibility of the threat, and communication, or the ability to transmit the message of capabilities and the credibility of the threat to the challenger. Most researchers believe that an analysis of these conditions

shows that the strategy of deterrence may be expected to fail when applied to threats created by cyber warfare.[16]

Capabilities

Cyber warfare allows weak players to move the confrontation into a sphere in which they can maximize profits while risking little – which makes deterrence harder to establish. In effect, an actor that is more technologically developed is also more susceptible to cyber warfare.[17] In fact, the possibility of retaliation against a weaker player is reduced, and thus the ability to establish a credible threat of deterrence is also lessened. For example, it is very difficult to deter players, especially individuals, who do not own information systems that can be threatened with damage.[18] This challenge also exists in the confrontation with nations with less developed information systems infrastructures, where the possibility of creating an effective threat by means of cyber warfare alone is limited.

Credibility

A second challenge to deterrence against cyber threats relates to the defender's credibility. The defender's vulnerability may limit its willingness to tap its capabilities out of concern that retaliation could lead to escalation. The problem for the defender is that such escalation is liable to be much more dangerous to itself than to the challenger, which in turn is likely to strengthen the challenger's belief that the defender's willingness to act is low.[19] This challenge is further amplified by the fact that cyber warfare entry costs are usually lower for the weaker side.[20] In other words, the cost to the challenger of engaging in cyber warfare is often limited, which further increases the difficulties in presenting and executing the deterrent threat required in order to prevent such action.

Internal as well as international public opinion may limit the credibility of the threat of retaliation because of the nature of cyber warfare. In situations in which it is difficult to establish the identity of the source of the attack,[21] the ability to employ a retaliatory measure likely to cause damage is constrained.[22] A potential challenger may view these constraints as undermining deterrence credibility. In this way a potential aggressor, assessing that the chances of the defender making good on its threats are low because of the damage it is likely to incur as a result, will be more willing to take risks and challenge the defender.

Conveying the Threat

A third problem stems from the defender's difficulty in conveying the message about its capabilities and about the credibility of its response to the challenger. Beyond the fundamental problems regarding each of the dimensions described above, challengers may be not only anonymous but even individuals who often have no identifiable physical address.[23] Libicki, for example, claims that to this day the source of the 2007 attack on the Estonian servers is in question: it is not at all certain that the attack was directed from above by the Russian government, as claimed by many who have analyzed the case.[24] The source of an attack can be another state entity, organizations or individuals operating from within the borders of another state, or organizations or individuals operating from within the targeted state. This situation reflects the frequent blurring between crime, terrorism, and warfare.

Moreover, when speaking of deterrence, it is necessary to identify the challenger in advance, before any challenge takes place, in order to target the deterrent threat. This is a key issue, because deterrence is based on the fact that the potential challenger is aware of the defender's capabilities and its willingness to use them ahead of time. However, if the defender is hard pressed to identify the source of the damage even after the attack, it will certainly find it difficult to do so prior to it. While intelligence capabilities may provide a partial solution, the threat that the defender can envision in most situations is general only, and is meant to cover a relatively broad range of potential challengers that the defender thinks would be likely to attack. However, deterrence is more effective when the threat – even if not completely explicit – is aimed at specific actors rather than at anonymous and undifferentiated sets of actors or types of actors liable to issue a challenge.[25]

Another difficulty directly related to the transmission of messages to the challenger involves the specific platform used.[26] This difficulty is amplified in light of the multiplicity of actors capable of creating threats. Unlike the Cold War era, when enemies were a limited number of known state entities with relatively clear capabilities, the number of possible aggressors has multiplied in the information age, lowering the possibility of presenting stable and credible deterrence.[27] The large number and variety of threats possible in cyber warfare creates an arena in which it is more complex to operate and in which it is not completely clear how or to whom to transmit the deterrent message.

Opportunities for Deterrence in Cyber Warfare

Despite these difficulties, the possibility of successful deterrence in cyber warfare exists, at least in part and under specific circumstances. For example, a number of researchers have stressed that retaliation need not be limited to cyberspace but may be effected by more traditional means. Thus, in the case of a state threatening to act by means of cyber warfare, the deterrent threat towards it may be based on the broadest range of capabilities the defending nation has at its disposal. Different threats, whether economic or military, may be effective in deterring a state enemy using cyber warfare against another state entity. Similarly, against threats posed by individuals or terrorist organizations seeking to use cyber warfare, states may, as proposed by a number of researchers (and also several decision makers), choose means of deterrence that do not require use of cyber capabilities. For example, they can employ threats through the judicial system (internal or international) and through internal security services, as well as use of traditional military threats.[28] As such, if actors assess that they will profit by diverting the confrontation into cyberspace, where they enjoy superiority, the actors under attack that might be attacked are under no obligation to limit the theater to cyberspace and may instead move the confrontation into theaters more convenient to them.

Another measure is deterrence by denial. The benefit inherent in this sort of strategy is that it may be based on defensive measures and thus not only be a means of preventing the enemy from acting but also providing a solution in case the challenger decides to act. Moreover, according to Morgan, making extensive use of various defensive measures may help identify the aggressor and strengthen the ability to take retaliatory action, which in turn strengthens deterrence by punishment.[29] Nonetheless, the challenges of using this strategy lie in overcoming problems similar to those linked to the successful use of deterrence by punishment. In both cases, the low entry cost required of challengers when they engage in cyber warfare remains a central difficulty.

Morgan also suggests that serial deterrence[30] may be useful in confronting cyber warfare threats: "Cyber attacks are very likely to turn out to be manageable primarily through applications of serial deterrence, repeated harmful responses over an extended period, to induce either temporary or eventually permanent suspensions of the most bothersome

attacks or of attacks by the most obnoxious opponents."[31] While this is an original way to confront threats in cyberspace and represents an interesting attempt to use existing concepts in an innovative way, it is not without difficulty. For example, it is unclear whether the enemy can be affected over time by repeated attempts, as these are liable to teach the challenger that the deterrence of the defender is not working (and that therefore the defender needs to engage in the same repetitive actions).[32]

Another problem regarding a strategy based on serial deterrence is exposing the capabilities of the defender. Although this problem is inherent in every form of deterrence in cyberspace (deterrence by punishment or denial), it is particularly acute when what is at issue is deterrence over time, as with the strategy of serial deterrence.[33] In such situations, exposing the offensive capabilities as the consequence of repeated attacks may serve as the basis for knowledge or inspiration for the challenger.[34] Morgan himself has referred to this issue and argues that revealing capabilities is liable not only to provide inspiration to enemies and motivation to attain similar capabilities but is also likely to allow enemies to prepare for a future threat, thereby damaging its measure of effectiveness.[35]

Directions for Further Research

While indeed some scholars have started to suggest new directions for research on deterrence in cyberspace, I would like to point to two main avenues through which cyber deterrence thinking can be further developed. First, research dealing with threats in cyberspace should be sharpened. It seems that there is a growing gap between practice and types of threats in the international arena, and the way in which research in this field examines the strategy of deterrence. This gap exists in other research dealing with deterrence, but it is particularly prominent in the realm of cyberspace, which includes many types of interaction between many different sorts of actors representing various kinds of threats. Therefore it is necessary to expand the discussion about the types of actors, the threats they create, and the ways and challenges of deterring each one. In addition, similar to the broader research relating to the strategy of deterrence, there is a tendency to focus on the deterrence of states against various types of players (e.g., terrorist organizations, rogue states),[36] while an important aspect not given sufficient attention is the deterrence of these actors against the states they seek to challenge. This aspect exists also in cyber warfare

and intensifies the problems of states that must now deal with a much more complex setting than in the past.

Moreover, research on cyber warfare tends to deal with more classical aspects of security, whereas the arena of threats is complex and varied.[37] For example, states are worried about the growing strength of economic players (such as Google) or ideological ones (e.g., individuals seeking to promote government reforms) using cyberspace. Irrespective of whether or not the existing definitions of cyber warfare include interactions with these actors, a considerable contribution could be made by analyzing these relations using theories of deterrence. The concept of the strategy of deterrence might be used, for instance, to study the interactions between Google and China with regard to the implied or direct threats presented by these players to one another in the context of search engine censorship. In this sense, dividing research on deterrence and cyber warfare according to different types of threats (e.g., internet war, cyber terror, cybercrime, cyberwar) and the actors operating them (states, individuals, economic institutions) may be not only more accurate and productive but may also identify the conditions for raising the chances of success of each actor's strategy of deterrence against its enemy.

The second theme that should be expanded is analysis of the traditional literature on the strategy of deterrence in critical and original ways. This has already been done in some of the essays published on the topic. However, it remains to analyze further concepts regarding deterrence strategy already discussed in the literature, such as immediate deterrence,[38] general deterrence, and extended deterrence,[39] and to try to understand the significance and relevance of applying these practices to cyberspace.

Similarly, the concept of ambiguity should be studied. This concept may serve as a framework for practical thinking in confronting the dilemma inherent in the need for revealing capabilities on the one hand,[40] balanced against the concern that the enemy will be able to exploit this exposure to increase its own strength and immunity to attack. Using insights developed in different contexts may provide an interesting foundation for developing ideas on cyberspace ambiguity, not only with regard to intention and willingness to make good on threats but generally with regard to the existence of capabilities. In this respect, it is possible, for example, to analyze the different efforts made by several nations in recent years in the field of cyber warfare. Not only are the means developed by nations

likely to strengthen their strategy of deterrence against these threats, but the very prominence of these efforts may also serve as a deterrent tool. The same is true of the American establishment of a strategic command to manage cyber warfare:[41] it has a range of objectives and functions, but its very reference and prominence allow not just improvements in capabilities but also demonstrate US willingness to invest resources in reducing threats and damage. It may be that stressing the desire to invest in measures of this sort and revealing the scope of the budgets, resources, and manpower dedicated to the subject – even absent a detailed breakdown of the measures acquired and their capabilities – can help increase the credibility of the deterrent message against threats in cyberspace, especially with regard to threats involving high levels of violence on the part of other nations. In other words, a partial revelation of capabilities while maintaining ambiguity about their essence allows for a reduction of the harmful effects described above but also transmits a forceful message. At the same time, one may expect that the low entry threshold for operating in cyberspace, especially in cases of asymmetrical confrontations, will continue to present a challenge to establishment of a strategy of deterrence seeking to prevent threats in this realm.

Conclusion

The research that deals with cyber warfare deterrence discusses primarily the difficulties inherent in deterring enemies from using this strategy. Although deterrence may work under certain circumstances, the problems associated with the defender's capabilities, the defender's willingness to use them, and the defender's ability to convey a message of deterrence to its potential enemy greatly limit the possibility of successful deterrence. Nonetheless, in light of the benefits inherent in the strategy of deterrence in reducing the scope of violence of conflicts, it is important to try to further the research dealing with the connections between deterrence and cyber warfare. This essay has indicated some directions for further thought and development of these ideas. However, as claimed by Morgan, these insights should be applied carefully, because additional empirical knowledge about the essence of cyber warfare is required, in terms of both the damage it can generate and the way in which it may be used.

Notes

1 Amir Lupovici, "The Emerging Fourth Wave of Deterrence Theory: Toward a New Research Agenda," *International Studies Quarterly* 54, no. 1 (2010): 705-32.

2 "Cyber warfare" refers here to a certain type of information warfare, though at times the concept of "information warfare" serves as a synonym for cyber warfare. This type of warfare is based on various attempts to prevent, disrupt, or destroy the enemy's information systems, while protecting the information systems of the defender against similar threats. See Richard J. Harknett, "Information Warfare and Deterrence," *Parameters* 26, no. 3 (1996): 93-107; Gary F. Wheatley and Richard E. Hayes, Infor*mation Warfare and Deterrence* (Washington, DC: National Defense University Press, 1996), pp. v-vi, 5-6; Roger C. Molander, Andrew S. Riddile, and Peter A. Wilson, "Strategic Information Warfare: A New Face of War," *Parameters* 26, no. 3 (1996): 83, 86-90. For a review of central concepts in cyber warfare, see Lior Tabansky, "Basic Concepts in Cyber Warfare," *Military and Strategic Affairs* 3, no. 1 (2011): 75-92.

3 On the general tendency of research dealing with cyber warfare and security to analyze policy oriented issues and to minimize the incorporation of broader theoretical dimensions, see Johan Eriksson and Giampiero Giacomello, "The Information Revolution, Security, and International Relations: (IR)relevant Theory?" *International Political Science Review* 27, no. 3 (2006): 221-44.

4 This essays use the common terms to describe the actors involved in deterrence strategy: the *defender* – the actor seeking to use the strategy of deterrence in order to prevent undesirable action against it, and the *challenger* – the actor seeking to act against the defender. The sometime usage of the alternative terms – the deterring actor or the deterred actor – is problematic because it assumes the success of the strategy.

5 For an excellent survey of definitions of the concept of deterrence by punishment, see Patrick M. Morgan, *Deterrence Now* (New York: Cambridge University Press, 2003), pp. 1-2.

6 Alexander George and Richard Smoke, *Deterrence in American Foreign Policy: Theory and Practice* (New York: Columbia University Press, 1974), p. 11.

7 Deterrence by denial also differs from the strategy of defense. While there is an overlap, defense seeks to provide a solution to a situation in which the strategy of deterrence has failed, while deterrence by denial seeks to prevent the action by making the challenger understand that it lacks the capacity to execute the action because of the defender's capabilities.

8 Glenn Snyder, *Deterrence and Defense* (Princeton: Princeton University Press, 1961). Nevertheless, deterrence by punishment and deterrence by denial may in theory support one another. If a potential challenger is made to realize that not only are its chances for success low but it will also be

required to pay a steep price for aggression, there is a higher chance it will refrain from action.

9 Thomas Schelling, *Arms and Influence* (New Haven: Yale University Press, 1966).

10 Albert Carnesale, Paul Doty, Stanley Hoffmann, Samuel P. Huntington, Joseph S. Nye, Jr., and Scott D. Sagan, *Living with Nuclear Weapons* (Cambridge: Harvard University Press, 1983).

11 For a discussion of conventional deterrence, see., e.g., John J. Mearsheimer, *Conventional Deterrence* (Ithaca: Cornell University Press, 1983) and Jonathan Shimshoni, *Israel and Conventional Deterrence: Border Warfare from 1953 to 1970* (Ithaca: Cornell University Press, 1988).

12 For example, it has been claimed that the development of international norms calling for the ban on nuclear weapons and international public opinion in support of this call have weakened the strategy of deterrence because they have raised the cost of their use of them. See T. V. Paul, "Nuclear Taboo and War Initiation in Regional Conflicts," *Journal of Conflict Resolution* 39, no. 4 (1995): 696-717.

13 Various researchers have debated the question of how to increase the credibility of the threat and have even proposed measures to attain this goal, e.g., by means of costly signals. See James Fearon, "Domestic Political Audiences and the Escalation of International Disputes," *American Political Science Review* 88, no. 3 (1994): 577–92. Still, some researchers have cast doubt on the effectiveness of some of these measures. For a discussion of the topic, see, for example, Paul Huth, "Reputations and Deterrence: A Theoretical and Empirical Assessment," *Security Studies* 7, no. 1 (1997): 72-99.

14 Morgan, *Deterrence Now*, pp. 15-16.

15 For an excellent survey demonstrating the different types of Israeli deterrence, see Uri Bar-Joseph, "Variations on a Theme: The Conceptualization of Deterrence in Israeli Strategic Thinking," *Security Studies* 7, no. 3 (1998): 12-29.

16 Harknett, "Information Warfare and Deterrence"; Br*uce* D. Berkowitz, "Warfare in the Information Age," in John Arquilla and David F. Ronfeldt, eds., *Athena's Camp: Preparing for Conflict in the Information Ag*e (Santa Monica: RAND, 1997), pp. 183-84; Emily O. Goldman, "Introduction: Security in the Information Technology Age," in Emily O. Goldman, ed., *National Security in the Information Age* (London: Taylor & Francis, 2004), p. 3; John Arquilla. "Thinking about New Security Paradigms," in Emily O. Goldman, ed., *National Security in the Information Age* (New York: Routledge, 2004), pp. 210-13. Morgan reaches similar conclusions, claiming that the different elements affecting the practices of deterrence of the Cold War, based both on this strategy and on supportive measures such as arms control, are less relevant to deterrence in cyberspace, though he does not entirely rule out the possibility of using different types of deterrent strategies in confronting these threats. See Patrick M. Morgan, "Applicability

of Traditional Deterrence Concepts and Theory to the Cyber Realm," in John D. Steinbruner et al., eds., *Proceedings of a Workshop on Deterring Cyberspace* (Washington: National Academies Press, 2010), pp. 55-76. In light of the various limitations regarding the ability to establish deterrence against cyber warfare, it has been proposed – especially for the United States, which is the primary subject of the research – to take alternative measures, such as using defensive means. See Wheatley and Hayes, Infor*mation Warfare and Deterrence*, p. 9, and James Adams, "Virtual Defense," *Foreign Affairs* 80 (2001): 107-12.

17 Harknett, "Information Warfare and Deterrence"; Wheatley and Hayes, *Information Warfare and Deterrence*, p. 9; Berkowitz, "Warfare in the Information Age," pp. 183-84; Martin C. Libicki, *Conquest in Cyberspace* (Cambridge, Cambridge University Press, 2007), p. 272. On societies' vulnerability to electronic attacks, see Ron Deibert, "Circuits of Power: Security in the Internet Environment," in J. P. Singh and James N. Rosenau, eds., *Information Technologies and Global Politics: The Changing Scope of Power and Governance* (NY: SUNY Press, 2002), p. 115. For sensitivity to threats – both external and internal – to information systems, see Martin C. Libicki, *Cyber Deterrence and Cyberwar* (Santa Monica: RAND, 2009), www.rand.org/pubs/monographs/2009/RAND_MG877.pdf. At the same time, for Libicki the scope of threat created by cyber warfare in the present age is neither clear nor certain. According to Libicki, the issue of the scope of damage liable to be created by cyber warfare is a central question at the heart of the debate about the importance of the strategy of deterrence against this type of warfare (*Cyber Deterrence and Cyberwar*, p. 36). For similar reasons having to do with the paucity of available information and the newness of the subject, Morgan cautions against drawing hasty conclusions about the possibilities of deterrence against cyberspace threats, "Applicability of Traditional Deterrence Concepts and Theory to the Cyber Realm," pp. 61-62.

18 Libicki, *Cyber Deterrence and Cyberwar*, p. 26.

19 Harknett, "Information Warfare and Deterrence," p. 104.

20 Molander, Riddile, and Wilson, "Strategic Information Warfare," p. 87.

21 For more on the difficulties in identifying the source of cyber warfare attacks, see also Libicki, *Cyber Deterrence and Cyberwar*, pp. 44-45.

22 For more on internal and international public opinion limiting the possibility of using force, thereby affecting the defender's deterrence, see., e.g., Robert Jervis, "Deterrence, Rogue States, and the Bush Administration," in T. V. Paul, Patrick Morgan, and James Wirtz, eds., *Complex Deterrence*: *Strategy in the Global Age* (Chicago: University of Chicago Press, 2009), p. 153.

23 Wheatley and Hayes, Infor*mation Warfare and Deterrence*, p. 9; Harknett, "Information Warfare and Deterrence," p. 104; Berkowitz, "Warfare in the Information Age," pp. 183-84; Anthony Cordesman and Justin Cordesman, *Cyberthreats, Information Warfare, and Critical Infrastructure Protection:*

Defending the US Homeland (Westport: Praeger, 2001), p. 7; and Arquilla, "Thinking about New Security Paradigms," pp. 210-11.

24 Libicki, *Cyber Deterrence and Cyberwar*, pp. 1-3.

25 The reason is that a deterring threat must be adapted to the type of threat and the type of element posing it. Therefore it is important to establish the deterrence in the context of the threat for the specific aggressor. For example, deterrence against a state actor enjoying sovereignty in a particular territory and possessing valuable target differs from a non-state actor and therefore requires the presentation of different types of threats. This issue has in recent years been at the center of an extensive debate in the context of tailored deterrence, particularly in the context of deterring terrorism. For a discussion of the concept, see Jeffrey S. Lantis, "Strategic Culture and Tailored Deterrence: Bridging the Gap between Theory and Practice," *Contemporary Security Policy* 30, no. 3 (2009): 469-71. For a discussion of the concept vis-à-vis cyber warfare, see Richard L. Kugler, "Deterrence of Cyber Attacks," in Franklin D. Kramer, Stuart H. Starr, and Larry K. Wentz, *Cyberpower and National Security* (Washington, DC: National Defense University Press, 2009), pp. 331-33, and Lantis, pp. 469-71.

26 Harknett, "Information Warfare and Deterrence," pp. 98-100.

27 Libicki, *Conquest in Cyberspace*, p. 272. For more on the effect of the Revolution in Military Affairs on deterrence and the ability to deter, see Morgan, *Deterrence Now*, pp. 219-24.

28 Hayes and Wheatley, Infor*mation Warfare and Deterrence*, pp. 13, 19-20; Kugler, "Deterrence of Cyber Attacks," p. 328; and in Cordesman and Cordesman, *Cyberthreats, Information Warfare, and Critical Infrastructure Protection* , p. 7.

29 Morgan, "Applicability of Traditional Deterrence Concepts and Theory to the Cyber Realm," p. 59.

30 Doron Almog uses the similar concept of "cumulative deterrence" with regard to the way to deter terrorist threats not in the cyber arena. See Doron Almog, "Cumulative Deterrence and the War on Terrorism," *Parameters* 34, no. 4 (2004-2005): 4-19.

31 Morgan, "Applicability of Traditional Deterrence Concepts and Theory to the Cyber Realm," p. 59.

32 Lupovici, "The Emerging Fourth Wave of Deterrence Theory: Toward a New Research Agenda," p. 722.

33 Thus, for example, a challenger is likely to learn about the defensive measures (or be inspired to attain such measures) on the basis of the means used by the actor trying to use deterrence by denial, thereby limiting the ability to deter effectively with this strategy.

34 Similar criticism was raised after the reports about the Stuxnet virus, which reportedly disrupted the systems of the Iranian reactor in Bushehr. The concern presented by a number of information security specialists was that this cyber attack would serve as inspiration not only for what can be

done using such warfare but also that some of the codes of the virus itself were revealed and could conceivably serve various actors in their attempts to damage sensitive infrastructures. See., e.g., "Experts Fear Hackers Can Launch Stuxnet-Like Attacks on Power Plants, Prison Gates," *The Globe and Mail,* October 24, 2011.

35 Morgan, "Applicability of Traditional Deterrence Concepts and Theory to the Cyber Realm," p. 63.

36 For reference to this issue in the context of information warfare, see, e.g., Goldman, "Introduction: Security in the Information Technology Age," p. 3.

37 For a discussion of the range of these threats, see Tabansky, "Basic Concepts in Cyber Warfare," especially pp. 80, 86-88.

38 A basic distinction existing in the study of deterrence deals with the difference between general deterrence, based on the attempt to prevent the enemy from thinking at all about the possibility of attacking (e.g., as with nuclear deterrence), and immediate deterrence, touching on a situation in which an actor would like to take an action (e.g., move troops) and by using threats the defender dissuades the enemy from taking such action. An important discussion in this context could deal with the meaning of each of these types of deterrence in cyberspace.

39 Libicki, for example, has started to analyze extended deterrence in cyberspace. See Libicki, *Cyber Deterrence and Cyberwar*, pp. 104-6), and it is possible to develop the discussion of theoretical issues discussed in the literature with regard to extended deterrence. For a discussion of the concept of extended deterrence see., e.g., Paul Huth,, *Extended Deterrence and the Prevention of War* (New Haven: Yale University Press, 1988).

40 The literature about deterrence stresses that it is necessary to transmit the threat message to the enemy, including the price it will have to pay. Therefore messages about defensive capabilities or revealing capabilities have been noted as important elements in this context.

41 "U.S. Cyber Command Fact Sheet," *US Department of Defense*, May 25, 2010, http://www.defense.gov/home/features/2010/0410_cybersec/docs/cyberfactsheet%20updated%20replaces%20may%2021%20fact%20sheet.pdf.

The Decline of the Reservist Army

Yagil Levy

Every spring, with great pomp and circumstance, Israel celebrates the contribution of the reservists to the country's security, and political and military leaders laud the contribution of reserve soldiers to national security. In 2011, however, discordant notes marred the festive event, namely the ongoing protest by organizations of reservists about the gap between the nation's commitments and their fulfillment. The protest by reservists was heard while the President of Israel and the IDF Chief of Staff visited the Ze'elim training base and during a stormy debate in the Knesset.

The IDF reserves, formerly the backbone of the military's force, is now at a crossroads, and it appears that even the IDF command and the political echelon are not sure how to reshape it. This essay argues that a combination of political and economic costs involved in operating the reserves is accelerating the decline of this force, and is part of the general move towards the transformation of the IDF from conscript to professional army.

The Rising Costs of the Reserves Model

The IDF's reserves model is expensive, both politically and economically. Initially the opposite was the case: the ethos of "nation in arms" ensured that reservists would serve in their capacities with full political obedience. At the same time, reserve duty was also economical, as either the employers or the reservists themselves bore the brunt of compensating the reservists for loss of income. This was the case before the full compensation system was implemented by the National Insurance Institute of Israel (NIII), particularly after 1967.

The political costs rose after 1967. Starting with the three-week waiting period before the Six Day War, while the mobilized reservists were disquiet

Prof. Yagil Levy is a member of the faculty at the Open University.

in the face of the government's hesitation in going to war, the political cost of mobilizing the reserves slowly started to dawn on the decision makers. Reservists have political bargaining chips, both because they are enlisted civilians living simultaneously in both worlds[1] and because of their natural position in the middle class – whether originally (because they reflect the standing army of yesterday in which there was a much higher representation of the middle class than today) or because of social mobility. This potential cost figured among the leading considerations in the decision to avoid a mass mobilization of the reserves on the eve of the Yom Kippur War in light of the approaching elections and after the pointless but expensive and much-criticized mobilization of some of the reserve units just a few months previously, given the concern about a possible Egyptian attack.[2] Refraining from this mobilization in no small way shaped the outcome of that war.

The various groups organized by reservists that arose after 1973, from Motti Ashkenazi (who led the anti-government protests of army reservists at the end of the war with the demand for resignations of the government for its misconduct of the war) to Peace Now, contributed to the breakup of the military decision making monopoly among the political elites and expansion of the political discourse in a way that gradually eroded the government's autonomy in making military and political decisions. This process grew stronger after the 1982 Lebanon War. The length of that war, the expansion of its objectives, and its entanglement in the quagmire of a war of attrition encouraged new reservist movements that for the first time included selective, organized disobedience. Foremost among these were Yesh Gvul ("There is a Limit") and Soldiers against Silence, alongside the older Peace Now organization. The protests they generated made a decisive contribution to the unilateral redeployment in Lebanon in 1985, two years after the government directed a partial withdrawal from Beirut and the Shuf Mountains to the Awali River. "We left Lebanon because of the reservists," said Minister of Defense Moshe Arens, referring to their protests.[3]

From this point onwards, decision makers grasped the idea that the deployment of reservists comes with a significant political price tag that narrows the scope of autonomy of the political decision making process. Thus when the first intifada erupted in 1987 and the right wing parties in government wanted to pressure the army to put down the civilian uprising

with force – an approach opposed by the left – Chief of Staff Dan Shomron told the government that the uprising had a political solution but not a military one. In thus seeking to mitigate the army's role in putting down the uprising, Shomron sought to prevent the dissolution of the army, which comprised essentially an even number of soldiers from the left and the right, especially at a time when the deployment in the territories to a very large extent depended on reservists, "alumni" of Lebanon. His statement to the government almost certainly tempered the potential opposition of left-leaning soldiers by lending their activity the sense of a necessary temporary measure not meant to decide the confrontation; this would be achieved diplomatically through negotiations. The restraint of the army paved the way for a partial withdrawal from the West Bank and the Gaza Strip in the form of the Oslo Accords. Moreover, Prime Minister Yitzhak Rabin testified that the concern lest the government fail to fully implement a general reserves mobilization in a controversial war played a role in his decision to embark on the Oslo process.[4]

The reservists' protest embodied some of the growing sensitivity among Israeli society to military casualties. This was expressed well by Soldiers against Silence, a group of released reservists who demonstrated opposite Prime Minister Menahem Begin's house against the war of attrition in Lebanon and carried signs with regularly updated tallies of the dead. This sensitivity made the army formulate a policy of casualty aversion that tried to minimize putting soldiers at risk and even avoided undertaking risk-laden operations, similar to comparable processes that have occurred in Western armies particularly since the war in Vietnam.

One component of this policy was to try to keep reservists away from sensitive hotspots. Indeed, the IDF's guerilla war in Lebanon in 1985-2000 relied on regular army personnel. According to the testimony of Moshe (Chico) Tamir, one of the commanders in Lebanon, this dependence on regular conscripts also minimized news coverage of the front.[5]

This was likewise the case with the al-Aqsa intifada of 2000-2005. Reservists were deployed in the heart of the combat primarily in Operation Defensive Shield (2002), in which the military reoccupied part of the West Bank cities, only after the legitimacy of the fighting was established on the basis of the activity of the regular army forces for close to two years. Despite the impressive response by reservists to the mobilization, not surprisingly, the protests about the nature of the missions and the division of labor and

compensation resurfaced once the operation was over, even as the "war for our homes," as the fighting against the Palestinians was described in the public discourse, continued. For example, the government's decision to extend reserve duty from 30 to 37 days after Operation Defensive Shield passed, but over much opposition, and a subsequent proposal to lengthen annual reserve duty was rejected.[6]

An additional element of casualty aversion was manifested in the formulation of a new military doctrine. Since the 1990s, the army has hinged its new doctrine on technology based on standoff fire: departing from the traditional approach to combat until the 1980s, the main principle involves moving the fire – not the forces – into enemy territory. The doctrine was built on the acquisition of high capabilities of destroying targets by aerial and artillery fire, with emphasis on precision armaments and without ground troops in enemy territory. The new doctrine was implemented for the first time in the 1990s in Operations Accountability and Grapes of Wrath against Hizbollah. Similar to the Revolution in Military Affairs promoted by the American army in the 1980s, the IDF's new doctrine was in part meant to reduce the number of casualties by intensive use of technology (the "shock and awe" technique), so that in context of the new political constraints it would be possible to shorten the duration of fighting and generate a rapid decision.

This approach was manifested in the Second Lebanon War, which represented a significant break in the relations between the army and the reservists. The fighting relied on aerial bombardments serving as standoff fire. The reservists were called up only after 16 days of fighting.[7] The hesitation in mobilizing the reserves expressed the dual political price: the concern that a high casualty toll, especially of reservists, would erode public support for the military action and damage the legitimacy of the government and the army,[8] and the understanding that from the moment the reservists were called up the government's freedom of movement would be constrained. This is how Chief of Staff Dan Halutz put it when trying to persuade the government to embark on a comprehensive rather than a graduated ground operation at the end of the war: "There are no middle courses here, of doing half, a quarter, or a third in order to satisfy some of our desires...It's all or nothing, because there are also people behind this willingness and there are reservist ORBATs ready to go and we can't just keep their hands tied saying, 'wait, wait.'"[9] In other words,

there was an echo to the legacy of 1967, meaning that the government could not allow itself extended waiting that could conceivably be accompanied by unrest on the part of the reservists.

This political cost did in fact reveal itself as relevant in light of the protest by the reservists after the war. Reservists joined bereaved parents and other groups that protested the army's flawed performance during the war. The protest intensified the constraints of the army by solidifying the expectation that the government would avoid risking soldiers' lives for nothing. Such a risk is present when the government has no political ability to complete the military operation and under circumstances in which the government cannot carry out the operation because of the army's lack of preparedness, even if the justification for the operation is not in doubt. Not coincidentally, a high estimate of casualties, including among reservists, played an important role is the government's postponing a ground operation in the Gaza Strip for a long time. The government authorized it only once the conditions were ripe for implementing a firepower approach that would reduce the exposure of IDF soldiers to danger in exchange for increasing the danger to the residents of Gaza.[10] Against the background of similar political consideration, reservists were barely called up to participate in the 2005 disengagement from Gaza, as this politically controversial task was assigned to the regular army.

A reservist battalion serving on the Israeli-Egyptian border in 2011 demonstrated anew the political cost of deploying reservists. At the start of their reserve duty, soldiers and officers in the battalion made it clear to the sector commander that they would not participate in "hot returns," the procedure authorizing IDF soldiers and border police to return asylum seekers (such as Sudanese refugees) to Egypt and turn them over to Egyptian police after brief questioning to make sure they were not seeking political refuge but were in fact trespassers. Reservists took this stance when it became clear to them that those returned are liable to encounter violence at the hands of the Egyptian police. The regional brigade commander acquiesced to the soldiers' request and instructed that while on duty the battalion in that sector would not use the controversial procedure, which is carried out routinely when the regular army Caracal Battalion is stationed there.[11] This is a demonstration of how reservists can limit, even if only temporarily, the army's autonomy.

Even before the Second Lebanon War, protests by reservists sharpened the sense that there was a "reserves crisis," as this phenomenon has been dubbed since the 1990s. The sense of crisis impelled the government, under pressure by reservist organizations and the reserves lobby in the Knesset, to approve reforms in the mobilization model. The formulation of these reforms started in a committee headed by Chief Reserves Commander Brig. Gen. Ariel Hyman and continued in the Braverman Committee, appointed by the government to propose a reform of the reserve system. The reforms were finally formalized with the passage of the 5768/2008 Reserve Duty Law. The law limits the state's authority to call up reservists and subordinates it to more explicitly defined rules than in the past. For example, it was determined that no reservist would be called up for the purpose of operational employment more than once during a period of three consecutive years (Paragraph 7, C, 2). This directive represented the political cost of deploying reservists by the very statement that operational employment would ordinarily be based on regular army forces whereas reservist deployment would be the exception.

The political cost of the reservist structure rose together with the economic cost. The first withdrawal from Lebanon in 1985 allowed the government to make the reservist structure selective in part, though not necessarily as the result of an explicit or even conscious decision. As part of the 1985 extensive cuts to the defense budget, it was decided to transition gradually the budgeting of reservist days from the NIII to the IDF. Previously the cost of reservist days, particularly compensation for the reservist's loss of income, was not borne by the defense budget and therefore the reservist structure was managed with no regard to economic considerations. As a result of this decision the army had a new incentive to reduce reserves days and use the savings for other purposes. In addition to the budgetary significance, the reserves were thus subordinated to the principles of a market economy, and for the first time an economic price tag was attached to reserve duty. The result was a significant cut in reserve duty days and a relief of the burden of service. For example, the budgetary basis for 1985, prior to the change, was 10 million reserve duty days annually. As a result of the cuts, reserve duty days fell to 3 million for 2006.[12] The downward trend was even felt in years when reserve units were deployed, especially during the two intifadas.

However, the reduction in reserve duty increased the inequality in the division of the service burden insofar as the army identified alternatives to the administrative but not the combat roles. About one-third of reservists bore some 80 percent of the burden as of the early 2000s, and only 10 percent of those obligated to perform army service (the total number of people eligible for reserve duty is about half of all males in Israel in the reservist age bracket) did annual reserve duty in excess of 10 days.[13] In other words, a very low percentage participates in what was once considered the institution that defined Israeli manhood.[14] The ethos of "the people's army" has ceded to the ethos of the marketplace. This growing inequality was the background for the organizing by reservists through various organizations (the Battalion Commanders Forum was the pioneer, followed by the Hapashim Forum and BALTAM) with the demand for a more equitable distribution of the burden and compensation for those serving, pressures that resulted in the Reserve Duty Law.

The economic cost limited the training of reservists to fulfill their missions in emergencies, and reservist training was cut significantly starting in 1989. At the same time, the training of regular army units was slashed starting in 2002 as the result of a difficulty in recruiting reserve units to replace the regular army units deployed in the Palestinian arena. This was yet another blow to the fitness of the reservist structure fed by the regular army units, in addition to the damage to the fitness of the regular army units themselves. From 2003 until 2005 the IDF reported to the political echelon that continuous damage was being done to the training of the reserve ground forces because of budget cuts.[15] In light of this, further development of the standoff fire approach became an entrenched fact of life thanks to budgetary limitations that resulted in damage to the fitness of the ground forces. In turn, adopting the doctrine and reshaping operational plans on the basis of standoff fire further eroded the investment of resources in the ground forces, as investing in it became redundant given the new alternative. Therefore, once the Israeli government decided to respond with force to the abduction of two reservists in June 2006, a response that became the Second Lebanon War, an aerial assault was the primary and preferred – practically exclusive – response. The reservists were called up late, and the circumstances of the mobilization and the execution gave rise to protests.

The Reserve Duty Law, a result of the "reserves crisis," raised the economic costs of deploying reserves by determining special compensation for reservists because of compensation for loss of income. As early as 1998 another law first determined that reservists would be compensated not only for loss of income but also for the service itself ("special compensation"); the new law formalized the practice and added the "special compensation" by means of the tax returns. The idea of a professional army began to become institutionalized.

In short, the political leadership and the army command internalized the political and economic costs of calling up the reserves. This understanding lay the groundwork for the process that followed, which gradually led to a reduction in the function of the reserves force. This is not to claim that the decision makers were fully conscious of the process, rather that this internalization of constraints shaped the strategic culture of the army, thereby delimiting the sphere of available decisions. Thus, gradually, the hands of the army became tied in terms of using the reserves. Its use became entrenched in economic bargaining and increased political bargaining, making the performance of some operations conditional on the values of the reservists called up for duty.

Looking Ahead

The Reserve Duty Law and the massive investment in training the reserves seemingly marked a change in the approach of the army and the political echelon regarding the importance of the reserves. However, this force is destined to decline as its costs continue to rise.

The political cost will rise as long as the missions of the army are politically controversial, be it regarding what is targeted or the cost of achieving a controversial objective. This controversy is intensified given the growing public sensitivity to casualties. This sensitivity enhances public criticism about the army's performance and makes the army and the political echelon behave with utmost care before calling up the reserves.

This cost is joined by the economic price tag. The pressure by reservists to improve the compensation package to make sure it covers the full cost of service did not cease even after the law was passed. A survey by the army's Behavioral Science Department showed that only one-third of those serving in combat units feel that the benefits and compensation package are significant.[16] Moreover, in the dialogue between reservists with the

army and government one hears over and over again of discrimination against reservists, especially commanders, in the workplace. To many employers, reserve duty is no longer social capital that the reservist brings to his place of civilian employment, rather the cause of negative yield. Over time the difficulty in confronting this phenomenon will increase the pressure to compensate and reward reservists.

Another source for pressure of this type is the expansion of the rate of inequality in bearing the burden. The Reserve Duty Law institutionalized the transition to a selective service model. The meaning of this is that young people who insist on being exempted from reserve duty will in most cases be discharged from service even if the formal obligation remains in place. Moreover, not only has the army's agreement to discharge those who are no longer in the regular army from military service been institutionalized, but the law even lays the groundwork for encouraging such discharges by means of two mechanisms. First, limiting reserve duty to training for the purpose of fulfilling the soldier's function during emergencies and operational employment winnows out those in various administrative positions. Second, obligating the army to ensure the level of fitness of reservists encourages the removal of those in whom the army will not invest to keep fit. From a different direction, the Brodet Commission, appointed by the government after the Second Lebanon War to formulate the desired size and composition of the defense budget in the short and long terms, recommended civilianizing many auxiliary roles, including those staffed by reservists.[17] This supports the incentive to minimize the employment of reservists.

On the other hand, while the Braverman Committee and the Reserve Duty Law sought to temper the inequality by reducing the scope of mobilizations, the army has shown consistent opposition. Recently, Deputy Chief of Staff Yair Naveh made this explicit: "Our need for missions along the border requires us to enlarge our ORBAT of employment... Personally I posit – and the Chief of Staff agrees – that it is preferable to harm the Reserve Duty Law rather than to harm training."[18] This is an example of the contradiction-riddled pattern of the state's handling of the reservist structure: on the declarative level, it recognizes its importance in emergencies and makes an effort to compensate those who serve, but it also exacerbates the inequity, making symmetrical compensation impossible.

The upshot is that the army will have to increase the monetary compensation. However, as it does so, it will tend to choose fair compensation while reducing the scope of the reserves to keep costs down, to the point that there will be an inevitable transition, as part of a spiral of compensation and selectivity, towards a professional army. The need to professionalize the reserves force will support this trend. Furthermore, making the reserve forces professional will keep down the political costs required for its deployment as long as the pattern of relations is established that turns the contractual relationship between the army and its people from a republican contract at the level of state-group, which grants reservists the right to express their political voice in the name of their contribution to the army, into a employment contract at the level of army-individual. Such a contract would weaken the infrastructure for political protest coming from within the ranks of the reserves. Governments will always prefer the economic to the political cost, particularly if the economic cost balances out the political one.

The Second Lebanon War offered a demonstration of this. The government decided to compensate the reservists who took part in the war (for at least 8 days) with a special compensation called "expense reimbursement" of NIS 400, plus another NIS 50 per day from the ninth day of service onwards.[19] This compensation was beyond that set by law for loss of income. In past wars, in which reservists were called up for much longer periods and for much more difficult service, no compensation beyond the formal compensation mechanisms that always existed was ever offered. This special compensation may be read as a mechanism to dampen protest by the reservists (which, however, was not needed after Operation Cast Lead, a situation viewed as an achievement). Moreover, the compensation was approved in August 2006 after the reserve units were released and the reservists' protest about the war began, initially focusing on the low level of fitness in the units. The more the mobilization is based on hiring rather than calling up, i.e., the greater the extent to which the monetary compensation plays a central role, so the state and army bypass the need to confront demands, expectations, and protests of a political nature or that may spill over into the political arena. This compensation, and later the monetary compensations enacted by the Reserve Duty Law put into place a system that will become more entrenched the more selective the service becomes, and bolsters the hiring profile over calling up

the reservists and lowering the potential for a political voice. The civilian political consciousness will play a secondary role, and the economic cost thus balances out the political one.

In the long run, the reserves structure will grow smaller and be based on a professional model founded on the service of the relatively few, gradually on a volunteer basis for relatively long periods of time, which will ensure their fitness, in exchange for adequate monetary compensation, similar to a model that several Western nations have adopted on top of the ruins of mandatory service. Israel is marching in that direction. The professional autonomy of the army and its political operators will be better off for it, but democracy, in which the voice of the reserves represented a critical cornerstone by its very ability to restrain the use of the army, will not.

Notes

My thanks to Aleh Mikanowski from Hapashim Forum (Forum for Soldiers who Serve in the Reserves) for his useful comments on the draft of this essay.

1 Nir Gazit, Edna Lomsky-Feder, and Eyal Ben-Ari, "Reservists between Worlds," *Maarachot* 394 (May 2004), pp. 87-94.
2 See Benny Morris, *Casualties: The History of the Arab-Zionist Conflict 1881-2001* (Tel Aviv: Am Oved, 2003), p. 357; and Uri Bar-Joseph, The *Watchman Fell Asleep: The Surprise of Yom Kippur and its Sources* (Tel Aviv: Zmora-Bitan, 2001), pp. 225-26.
3 Ofer Shelah and Yoav Limor, *Captives in Lebanon: The Truth about the Second Lebanon War* (Tel Aviv: Yediot Books, Tel Aviv, 2007), p. 319.
4 Yoram Perry, "The Relationship between Society and Army in Israel in Crisis," *Migamot* 39, no. 4 (1999): 394.
5 Moshe Tamir, *A War without a Sign* (Tel Aviv: Maarachot Press, 2005), pp. 10-11, 274.
6 These moves are documented in the Hapashim Forum at http://miluim.ipaper.co.il/1411.
7 *The Commission of Inquiry into the Events of the Campaign in Lebanon – Final Report*, The Prime Minister's Office, Jerusalem, 2008, p. 250.
8 Ibid., pp. 411, 526.
9 Ibid., p. 180.
10 Yagil Levy, *Who Governs the Army: Between Supervising the Army and Controlling Militarism* (Jerusalem: Magnes Press, 2010), pp. 168-70.
11 Anshil Pfeffer, "Reservists Opposed to Forced Return of Infiltrators to Egypt and Procedure Is Stopped," *Haaretz*, April 22, 2011, http://www.haaretz.co.il/hasite/spages/1225775.html.
12 As may be concluded from "Damage to Employees as a Result of Serving in the Reserves," Knesset Research and Information Center, Jerusalem,

2003, and "Assessment of the Budgetary Cost of Implementing the National Insurance Institute Bill," Knesset Research and Information Center, Jerusalem, 2007.

13 Ariel Hyman, "The Reservist Structure, the IDF and Israeli Society: Past, Present and Future," *Maarachot* 394 (May 2004), p. 5.

14 As demonstrated by Sara Helman, "Militarism and the Construction of Community," *Journal of Political and Military Sociology* 25, no. 2 (1997): 305-32.

15 The State Comptroller, *Annual Report 58A* , The Office of the State Comptroller and the Public Ombudsman, Jerusalem, 2007, pp. 87-97.

16 IDF Behavioral Sciences Department, *Stances of Commanders and Class A Reservists 2011*, 2011, at http://portal.knesset.gov.il/Com4bitachon/he-IL/CommitteeHistory/24052011.htm.

17 *The Report of the Committee Examining the Defense Budget*, The Prime Minister's Office, Jerusalem, 2007, p. 105.

18 Yoni Shoenfeld and Noa Horowitz, "In Coming Years, We'll Call up Reservists More, Not Less," *Bamahaneh*, May 25, 2011, at http://www.idf.il/1137-11284-he/Dover.aspx.

19 The IDF, *The Reservist's Handbook*, 2006, p. 7, http://www.aka.idf.il/SIP_STORAGE/files/4/59004.pdf.

Think Before You Act:
On the IDF Withdrawal from Lebanon in 2000

Giora Eiland

This article presents several facts and conclusions stemming from the Israeli withdrawal from Lebanon in 2000. It also discusses two other events that occurred after the withdrawal: the Second Lebanon War in 2006, and a relatively small yet important event, the departure of Syrian forces from Lebanon in 2005. All these events are connected to one another.

Before the decision by then-Prime Minister and Defense Minister Ehud Barak to withdraw from Lebanon – as he said at the time, with or without an agreement – Israel's military campaign in Lebanon appeared as follows: tactical fighting with Hizbollah, almost entirely in or on the edges of the security zone; attempts by the IDF to improve its capabilities while the enemy was also improving its capabilities – with the improvements on one side more or less offsetting the improvements of the other; and a fairly steady number of Israeli casualties each year – 20-25 killed every year, essentially irrespective of particular events (other than the collision of two Israeli helicopters in 1997).

The main questions discussed in the IDF were whether the fighting with Hizbollah could be conducted differently, and whether the existing situation was tolerable. The IDF was of the opinion that the situation was tolerable, and that it would be possible to continue in the same manner for a long term. In fact, there was no genuine, thorough discussion of what alternatives the State of Israel had at its disposal.

Maj. Gen. (ret.) Giora Eiland is a senior research associate at INSS. This article is based on a lecture delivered at the INSS conference "Ten Years Since the Withdrawal from Lebanon," June 28, 2010.

Moshe Arens, who was Defense Minister for a short period before 1999, attempted to present another approach: that Israel's effectiveness, as long it was confronting Hizbollah, was limited, and therefore, Hizbollah was not necessarily the correct opponent to confront. Moreover, the other opponent was not necessarily Syria. A direct confrontation with Syria would pose greater risks for Israel. According to Arens, the correct approach was to place responsibility on the Lebanese government, which despite its weakness, had the obligation of full state responsibility. Arens believed that it was correct to persist in attacking infrastructures in Lebanon.

This approach did not run its course apparently because of political changes in Israel. In 1999, when Ehud Barak was elected Prime Minister and also assumed the role of Defense Minister, he made his famous declaration: by July 1, 2000, the IDF would leave Lebanon, with or without an agreement.

IDF leaders did not like this statement. The army's preparations for departure from Lebanon were delayed a great deal, mainly because there was an intensive diplomatic process with Syria in late 1999 and early 2000. There was a feeling that it would be possible to reach an agreement with Syria. The IDF was directly involved in the negotiations, and the sense was that the moment Israel agreed to withdraw from the Golan Heights, it would be possible to reach an agreement, and if it was possible to reach an agreement, it was clear that it would include Lebanon, and therefore, there was no hurry to take any action.

The dramatic change took place in early March 2000, when after the Clinton-Assad meeting it was obvious that there was no agreement. Barak was firm in his commitment, and it was clear that the withdrawal from Lebanon would be unilateral and undertaken without an agreement. There were at most four months remaining for preparations, and there was an awareness of the need for some tactical surprises.

When the preparations for the withdrawal from Lebanon began, another heated argument on the meaning of unilateral withdrawal arose between the IDF and Chief of Staff Shaul Mofaz, and PM and DM Barak. The army's view of the unilateral withdrawal was very different from that of the Prime Minister. The IDF understood, or wished to understand, that the unilateral withdrawal was a tactical withdrawal whose meaning was: it is difficult for us to hold on to the security zone in terms of the effectiveness of the fighting, we have many casualties, remaining in Lebanon in the current format means that we are mainly serving as targets, and we should

therefore withdraw to another tactical line one kilometer from the fence. The IDF made statements to the effect that "it is inconceivable that we would leave regions like Jabal Hamamis, above Metulla, or other places that overlook our towns." In its view, there would be a withdrawal to such a tactical line (also called "shortened defense lines"), but from a strategic point of view, everything would be the same. The IDF would continue to provide as much support as possible to the South Lebanese Army (SLA), and it was clear that the IDF would continue to attack in Lebanon and operate beyond the new line. The change would be manifested in a different tactical deployment.

The Prime Minister–Defense Minister realized that this tactical withdrawal would not change the situation at all, and in a certain way would resemble the previous tactical withdrawal to the security zone, in 1985. As he put it, the crux is a strategic idea of withdrawal to an international line to receive legitimacy from important countries, though not from Lebanon or Syria, because Israel will not receive it from them. This legitimacy will ultimately lead to a better security situation.

The diplomatic proceedings were in fact conducted in a short period of time. One purpose was to reach a border that Israel and Lebanon agreed on. There was a reference line for the border in the area between the sea and the Hatzbani River that was based on the Sykes-Picot Agreement of 1916, a border line that was made official and recognized in 1923. In other words, there was a border, and it had to be restored. East of the Hatzbani there had never been a border between Israel and Lebanon. In the time of the French and British mandates, both Syria and Lebanon were part of the French mandate and there was no border between them. As such, there is no prior international basis for a border line between Israel and Lebanon in the eastern section (the Golan Heights-Mount Dov area).

There was no possibility of dialogue with the government of Lebanon, and therefore there was a need to withdraw to a line that the UN would recognize as the international border. The question was how to create this line so that it would earn international support. Colonel Haim Srebro, then head of the IDF's mapping department, found a UN map from 1974 that defined the mandate of the UN force in the Golan Heights. (After the Yom Kippur War, Israel and Syria agreed to a separation of forces, and a mandate was given to a UN force to oversee it. This mandate defined the boundaries of the force's responsibility, which included the Golan

Heights.) Therefore, the Golan Heights is the territory included in the map, and for purposes of this issue, any territory that is not found in the map is not included in the Golan Heights. In other words, if such a territory is in the north, then it is in Lebanese territory.

Since this line, which delimits the Golan Heights, left the important mountain range of Mount Dov (known to the Lebanese as Shab'a Farms) in Israeli territory, along with the important intelligence bases there, from Israel's point of view, this was a good line. Since the line was based on a UN map, Israel claimed that the map actually defined the border line all along the Mount Dov region, and the UN accepted this line. From Israel's point of view, the conspicuous price of this recognition is the village of Ghadjar, which is cut in two by this border line. The IDF then prepared for July 1, 2000, intending to take advantage of the time to conduct a withdrawal that was as orderly as possible.

Other speakers in the conference "Ten Years Since the Withdrawal from Lebanon" have described a process considered to be the start of the collapse of the SLA, mainly in the Western sector. The IDF did not anticipate the civilian processions that took place near the time of the withdrawal or their significance. On May 21, the chief of staff was visiting the Northern Command for a corps exercise, when information began to arrive about the processions. This was a surprise. One day later, when the civilian processions were gaining strength, a tense meeting took place at the Zarit outpost attended by the Prime Minister–Defense Minister, the chief of staff, the OC Northern Command, and other generals from the General Staff. The question was what to do now, once the SLA had begun to collapse. There were two possibilities: the first was to send one or two IDF divisions to replace the SLA in order to fight to retain that line. The question was whether it would be correct to fight for a line that the IDF was planning to withdraw from one month later. The other possibility was to adjust to the situation that was taking shape and speed up the withdrawal from Lebanon.

The decision was not at all simple. What helped the Prime Minister decide that the IDF would withdraw immediately was related not only to the military situation as described, but also to the coincidence of a diplomatic event. That same morning, a session at the UN gave approval to the UN secretary general to finalize with Israel the issue of the departure from Lebanon and the international border line. This meant that the Prime

Minister could rely on the UN resolution, which allowed Israel to do what it actually intended to do. The diplomatic conditions thus interfaced with the military situation. Accordingly, the withdrawal from Lebanon was indeed carried out in great haste, but Hizbollah too was surprised. The fact that the withdrawal took place in one night made it difficult for Hizbollah to attack the departing forces.

In 2005, for reasons not directly connected to Israel, a situation developed in which the entire world was angry with Syria, especially following the murder of Rafiq Hariri. A broad coalition came together – which included Saudi Arabia, the UN, France, and the United States – that accused the Syrian regime in Lebanon of the murder. As a result, Syria was forced to withdraw its soldiers from Lebanon.

Israel was a silent partner to the process of Syria's withdrawal from Lebanon and after an internal debate in government circles enthusiastically supported it. The main question was, is Syria's withdrawal from Lebanon good for Israel? There were those (I among them) who thought that the move was not favorable to Israel. Strangely, Israel and Syria had an interest in a Syrian presence in Lebanon for three reasons. First, the Syrians in Lebanon served as a restraining force, and they were the condition for ensuring that if Israel reached an agreement with Syria, it would also include Lebanon, and Syria would not be able to evade responsibility. Second, it was not at all clear that moderate democratic forces would be strengthened by a Syrian withdrawal from Lebanon, and it was possible that other forces would grow stronger. This is what actually happened, and Hizbollah and Iran have filled the void left by Syria. Third, the moment that Syria ostensibly gave up Lebanon, its focus would be on the Golan Heights. Those who opposed negotiations with Syria over the Golan Heights believed that it was preferable for Syria to fight over its hold in Lebanon and not to focus on the battle for the Golan Heights.

Syria's departure from Lebanon caused people in Israel to think that perhaps Israel would reap the benefits of the withdrawal from Lebanon in 2000, at least indirectly, five years after the fact: the IDF was no longer in Lebanon, the Syrians were hated, the Syrians were leaving Lebanon, and Lebanon would be a pro-Western, democratic country. This hope was quickly extinguished. Hizbollah actually took the place of Syria, and from this point of view, the situation in Lebanon is certainly worse than it was expected to be.

From 2000 to 2006, as with many other issues, the government of Israel did not hold a single discussion on the question of the correct policy toward Lebanon. The policy was determined de facto, mainly by the IDF. This meant that even if once every two-three months a tactical event occurred, attempts were made to get out of it in the best way possible, without escalating in the wake of the event beyond the minimum necessary. What emerged was a "policy of containment," and even restraint.

The lack of a strategic discussion on Lebanon is typical of Israel. In general, when a situation is ostensibly calm and does not require decisions, no diplomatic-strategic alternatives are created because it is not necessary; when there is an incident, the situation requires a quick response.

In 2006, Israel was given all the tools to place the conflict on the diplomatic level. Hizbollah was not just another organization, but part of the parliament. It was part of the regime in Lebanon, part of the government, and therefore Israel saw the Lebanese government as responsible for the fire directed from its country and could respond accordingly. This was the great missed opportunity of the Second Lebanon War: Israel defined the enemy in a manner that was too restrictive. If from the outset Israel had seen Lebanon and not only Hizbollah as the enemy, it would have been possible for this war to be much shorter, and the deterrence achieved at war's end would also have been more successful than it is today.

If we return to the decision made in 2000, I believe that Israel failed in the 1990s in that it did not attempt to develop a real alternative at any point, other than the two alternatives that were ostensibly available: 1) to preserve the existing situation; 2) to withdraw unilaterally. With the disengagement from Gaza as well, the public discussion was limited to two possibilities: Are you in favor of the disengagement or against it? Was that the entire range of possibilities? The answers are connected to the time at which the question was raised.

In the 1990s, Israel did not exhaust the full range of possibilities, and it is certainly possible that Israel had the ability to create a different situation. We have alluded here to steps in a certain direction that Moshe Arens attempted to promote, but in his case, the time was very limited because of the elections.

If we agree that there were only two possibilities, to maintain the existing situation or to withdraw from Lebanon, it would appear that Prime Minister and Defense Minister Ehud Barak made a courageous and

correct decision. I say this in context of my having been a general in the General Staff at the time, who, like most of my colleagues, was opposed to a unilateral withdrawal. Perhaps this discussion was conducted too loudly and with too much unanimity – and indeed, it is worth examining how such a dynamic is created within the General Staff.

In spite of all the problems, from the perspective of ten years, it appears that the decision to withdraw from Lebanon was correct. The great missed opportunity was that in 2006, Israel did not know how to leverage more successfully the legitimacy it was given six years earlier.

INSS Memoranda 2008 – Present

No. 115, March 2012, Emily B. Landau, *Decade of Diplomacy: Negotiations with Iran and North Korea and the Future of Nuclear Nonproliferation*.

No. 114, March 2012, Yoel Guzansky and Mark A. Heller, eds., *One Year of the Arab Spring: Global and Regional Implications* [Hebrew].

No. 113, March 2012, Yoel Guzansky and Mark A. Heller, eds., *One Year of the Arab Spring: Global and Regional Implications*.

No. 112, Uzi Rabi and Yoel Guzansky, eds., *The Gulf States: Between Iran and the West* [Hebrew].

No. 111, December 2011, Benedetta Berti, *The Ongoing Battle for Beirut: Old Dynamics and New Trends*.

No. 110, November 2011, Meir Elran, Owen Alterman, and Johannah Cornblatt, eds., *The Making of National Security Policy: Security Challenges of the 21st Century – Conference Proceedings*.

No. 109, June 2011, Shmuel Even and David Siman-Tov, *Cyber Warfare: Concepts, Trends, and Implications for Israel* [Hebrew].

No. 108, May 2011, Emily B. Landau and Tamar Malz-Ginzburg, eds., *The Obama Vision and Nuclear Disarmament* [Hebrew].

No. 107, March 2011, Emily B. Landau and Tamar Malz-Ginzburg, eds., *The Obama Vision and Nuclear Disarmament*.

No. 106, November 2010, Yehuda Ben Meir and Olena Bagno-Moldavsky, *Vox Populi: Trends in Israeli Public Opinion on National Security 2004-2009*.

No. 105, August 2010, Meir Elran and Yoel Guzansky, eds. *Vision and Reality in the Middle East: Security Challenges of the 21st Century – Conference Proceedings*.

No. 104, June 2010, Gallia Lindenstrauss, *Mediation and Engagement: A New Paradigm for Turkish Foreign Policy and its Implications for Israel* [Hebrew].

No. 103, May 2010, Tamar Malz-Ginzburg and Moty Cristal, eds., *A Nuclear Iran: Confronting the Challenge on the International Arena* [Hebrew].

No. 102, December 2009, Michael Milstein, *Muqawama: The Challenge of Resistance to Israel's National Security Concept* [Hebrew].

No. 101, November 2009, Meir Elran and Judith Rosen, eds. *The US and Israel under Changing Political Circumstances: Challenges of the 21st Century – Conference Proceedings*.

No. 100, September 2009, Aron Shai, *Sino-Israeli Relations: Current Reality and Future Prospects*.

No. 99, June 2009, Meir Elran, ed., *The Civil Front* [Hebrew].

No. 98, April 2009, Anat N. Kurz, *The Palestinian Uprisings: War with Israel, War at Home*.

No. 97, March 2009, Shmuel Even and Amos Granit, *The Israeli Intelligence Community: Where To?* [Hebrew].

No. 96, September 2008, Ron Tira, *The Struggle over the Nature of War* [Hebrew].

No. 95, August 2008, Anat N. Kurz, *The Palestinian Uprisings: Struggle on Two Fronts* [Hebrew].

No. 94, July 2008, Ephraim Kam, ed., *A Nuclear Iran: Implications for Arms Control, Deterrence, and Defense*.

No. 93, April 2008, Shmuel Even and Zvia Gross, *Proposed Legislation on the IDF: Regulating Civil-Military Relations in the Wake of the Second Lebanon War* [Hebrew].