THE OPTIMAL DISTRIBUTION OF IAEA INSPECTION EFFORT

by

D. Marc Kilgour

and

Rudolf Avenhaus

Research Report for the

Non-Proliferation, Arms Control and Disarmament Division

Department of Foreign Affairs and International Trade

WILFRID LAURIER UNIVERSITY
VERITAS OMNIA VINCIT

The Laurier Centre for
Military Strategic and
Disarmament Studies

UNIVERSITÄT DER BUNDESWEHR MÜNCHEN

THE OPTIMAL DISTRIBUTION OF IAEA INSPECTION EFFORT

FINAL REPORT

BY
DR. D. MARC KILGOUR
WILFRID LAURIER UNIVERSITY

AND

PROF. RUDOLF AVENHAUS
FEDERAL ARMED FORCES UNIVERSITY, MUNICH

OCTOBER 1994

COMPLETED FOR THE
NON-PROLIFERATION, ARMS CONTROL AND DISARMAMENT DIVISION,
FOREIGN AFFAIRS AND INTERNATIONAL TRADE CANADA

PREFACE

In general, inspections under an arms control regime are intended:

- to deter violations,
- to detect violations, and
- to minimize the damage due to undetected violations.

Inspections are usually subject to quotas, however, and inspecting agencies' budgets are limited, so there is always pressure to accomplish these objectives with minimal inspection effort. One way to make fewer inspections go further is by making inspection choices unpredictable, relying on the risk, rather than the certitude, of detection to deter. The only reliable way to achieve unpredictability is randomness; this is the justification for introducing random elements into the scheduling of inspections.

There are many different ways to incorporate randomness in the choice of which objects of verification to inspect. Uniformity (making each possible inspection pattern equally probable) is always feasible but, where there are significant disparities in value, uniformity is far from optimal in ability to deter and to limit damage.

After an agency has decided where to inspect, it must then decide how much. The level of effort applied to a verification task often affects the results. Intensive inspections are more likely to result in unambiguous reports, and less likely to miss evidence about current or recent anomalies. On the other hand, an extensive inspection program is more likely to give timely warning of problems, especially large scale problems, although there is a greater tendency for evidence to be incomplete and misleading. In summary, inspection locations, types and intensities, must be selected very carefully in order to cost-effectiveness.

The specific objective of this research project was to study how the probability and intensity of Non-Proliferation Treaty (NPT) inspections can best be spread over the objects of verification to achieve the general goals listed above. In particular, how should the IAEA's inspection patterns reflect the variable characteristics of objects of verification within or among states, and how much of the IAEA's resources should be expended on the detection of undeclared sites and activities? Finally, what is the relation between the cost and the maximum effectiveness of NPT inspections?

Other arms-control regimes, in addition to the NPT, might benefit indirectly from this research. These other arms-control contexts include most existing or proposed accords on conventional forces, chemical weapons, and space-based weapons.

In summary, the aim of this project is to contribute to an assessment of current and potential levels of cost-effectiveness of inspections of nuclear materials and activities. This assessment is timely, in view of recent events and of the impending (1995) NPT Extension Conference. A further purpose is to raise ideas about how, despite budgetary constraints, the IAEA can fulfil its mandate to deter, or detect in a timely fashion, not only diversion of nuclear materials, but also undeclared facilities and activities. It is hoped that the results will also help to improve cost-effectiveness of other forms of arms-control inspection, thereby making arms control better and cheaper

This report represents the results of a research project conducted under the Department of Foreign Affairs' Verification Research Program. It is being shared with interested parties as part of a long-standing Canadian policy to make such research findings available to assist in negotiations and to promote a dialogue on these important issues. The views expressed herein are those of the authors and do not necessarily represent those of the Canadian Government.

ABSTRACT

Under the Treaty on the Non-Proliferation of Nuclear Weapons (NPT), the International Atomic Energy Agency (IAEA) has responsibility for implementing safeguards. The mandate of the IAEA includes the inspection of declared facilities in non-nuclear-weapon states to confirm equipment and procedures and to verify self-reported data. As well, the IAEA's charge includes the detection of undeclared facilities and activities within these states. Recently, both the effectiveness and the costs of safeguards have received much increased attention: Experience in Iraq has demonstrated that safeguards against undeclared nuclear weapons development programs need to be strengthened, yet at the same time concern has increased about the rising costs of safeguards programs.

These fundamental problems are addressed in this report. Its objective is to present an assessment of current and potential levels of cost-effectiveness of inspections of nuclear materials and activities, and to suggest avenues for improvement. A specific framework demonstrating what determines required levels of inspection effectiveness is provided. With the help of the mathematical tools of Decision Theory and Game Theory, models are analyzed representing states' decisions to comply with or violate the NPT, and, if violation is chosen, where to violate. The models also include the IAEA's decisions of where and how much to inspect.

We hope that our analysis will focus the attention of policy makers on the crucial determinants of cost-effectiveness for NPT safeguards programs. Our study is timely in view of the impending NPT Extension Conference. But its ultimate goal is to contribute toward increases in effectiveness for all forms of arms control.

The Optimal Distribution of IAEA Inspection Effort

D. Marc KILGOUR & Rudolf AVENHAUS

1. Introduction

For many, the Treaty on the Non-Proliferation of Nuclear Weapons (NPT) [11] is preeminent among the world's arms control treaties because of its central role in reducing the threat to peace posed by the most powerful and destructive of all weapons. Since the NPT came into force in 1970, there have been no instances of the use of nuclear weapons in hostilities — yet the degree of success of NPT safeguards against nuclear proliferation is quite controversial. The possible renewal of the NPT, and possible changes to its provisions, will be the main issues at the NPT Extension Conference to take place in 1995.

Under the NPT, the International Atomic Energy Agency (IAEA) has responsibility for implementing safeguards, which are measures carried out "with a view to preventing the diversion of nuclear material from peaceful purposes to nuclear weapons." ([11], Article III.1) Furthermore, each non-nuclear-weapon state is required to accept safeguards "on all source or special nuclear fissionable material ... for the exclusive purpose of verifying that such material is not diverted to nuclear weapons." [7] The mandate of the IAEA thus includes the inspection of *declared* facilities in non-nuclear-weapon states to confirm equipment and procedures and to verify self-reported data. As well, the IAEA's charge includes the detection of *undeclared* facilities and activities within these states.

Recently, both the effectiveness and the costs of safeguards have received much increased attention within the IAEA. Experience in Iraq has demonstrated that safeguards against undeclared nuclear weapons development programs need to be strengthened, yet at the same time concern has increased about the rising costs of safeguards programs, and the IAEA's diminishing reserves. [3]

The IAEA faces very difficult problems in organizing its safeguards operations. (For details, see [5].) It must carry out its NPT inspection activities in well over 100 non-nuclear-weapon states around the globe, using total resources that are clearly far short of what would be necessary to ensure immediate and certain detection of non-compliance. Moreover, it must divide its efforts between two quite different problems — the verification of non-diversion for nuclear material in declared facilities, and the timely detection of undeclared facilities and activities.

These fundamental problems relating to both the effectiveness and the costs of IAEA's NPT safeguards operations are addressed in this report. The approach adopted here is based on an identification of the determinants of safeguards effectiveness and costs. With a more complete understanding of what levels of cost-effectiveness can potentially be achieved, the IAEA's safeguards programmes can be reassessed, and new routes to the solution of certain problems can be mapped out.

In fact, the problems faced by the IAEA in implementing an NPT safeguards programme are not qualitatively different from those faced by inspecting agencies under most arms-control regimes. In general, arms-control inspections are intended to deter and detect violations. Moreover, inspections are usually subject to quotas and inspecting agencies' budgets are almost always severely limited, so there is pressure to deter and detect with minimal inspection effort, i.e. minimal costs. Thus, although this study is specifically aimed at increasing the cost-effectiveness of IAEA safeguards programs for the NPT, many of the conclusions are also applicable in other arms-control contexts.

The objective of this report is thus to present an assessment of current and potential levels of cost-effectiveness of inspections of nuclear materials and activities, and to suggest avenues for improvement. A specific framework demonstrating what determines required levels of enforcement effectiveness is provided, and used to answer the following questions:

- *How should inspection resources for NPT safeguards be allocated over non-nuclear-weapon states to fulfill the IAEA's mandate?*
- *How should the IAEA's inspection effort be divided between the task of verification at declared sites and the task of detecting undeclared sites?*
- *How should the variable characteristics of states, and of declared and undeclared sites, affect the answers to the previous questions?*

The organization of this report is now summarized. In Section 2, the fundamental assumptions underlying the modelling and analysis to follow are discussed, and a framework is set out for determining the required level of effectiveness of an inspection program. Section 3 is divided into three subsections, each featuring a model (or group of models) that provides important information about safeguards operations:

3.1 Determination of the total inspection effort required.

3.2 Allocation of inspection effort among states.

3.3 Allocation of inspection effort within a state (against declared or undeclared sites).

(The technical analysis of these models is confined to the Appendix.) Section 4 contains some general conclusions about whether and how much the IAEA can improve its NPT safeguards programs through careful allocation of its inspection effort.

It is hoped that the answers to questions raised above will focus the attention of policy makers on the crucial determinants of cost-effectiveness for NPT safeguards programs, and in other arms-control arenas. A study like this one is timely, in view of recent events and the impending NPT Extension Conference. But its ultimate goal is to contribute toward increases in effectiveness, and reductions in costs, for all forms of arms control.

2. Basic Modelling Assumptions

In this section, the assumptions that form a basis for the subsequent analysis are introduced, explained, and justified. Some terminology used throughout will also be introduced.

All of the modelling of inspection effectiveness below is based on the analysis of decisions. It is assumed that decision makers use their full knowledge of the situation they face, and make choices that are in their own best interests. The formal methodology is called Decision Theory; when the outcome depends on the choices of two or more concerned decision makers, the branch of Decision Theory called (Non-Cooperative) Game Theory is invoked. Good background references are [4], [6], and [10].

As an illustration, consider Figure 1a, which provides a very simple model of the situation facing a state when it considers violating a treaty. In this vastly simplified model, the state's only choice is whether to violate or not — all details, such as how, where, or how much to violate, are suppressed. If the choice is "Violate," then the eventual outcome depends on a further event, whether the violation is "Undetected" or "Detected." In this model, the state sees this latter bifurcation as uncertain, and out of its control. On the other hand, if the state chooses "Comply," the outcome is completely determined; there are no intervening random events.

The state makes its choice based on its assessment of the values it could receive contingent on each of the three outcomes that could arise in this model. Here, and below, a state's

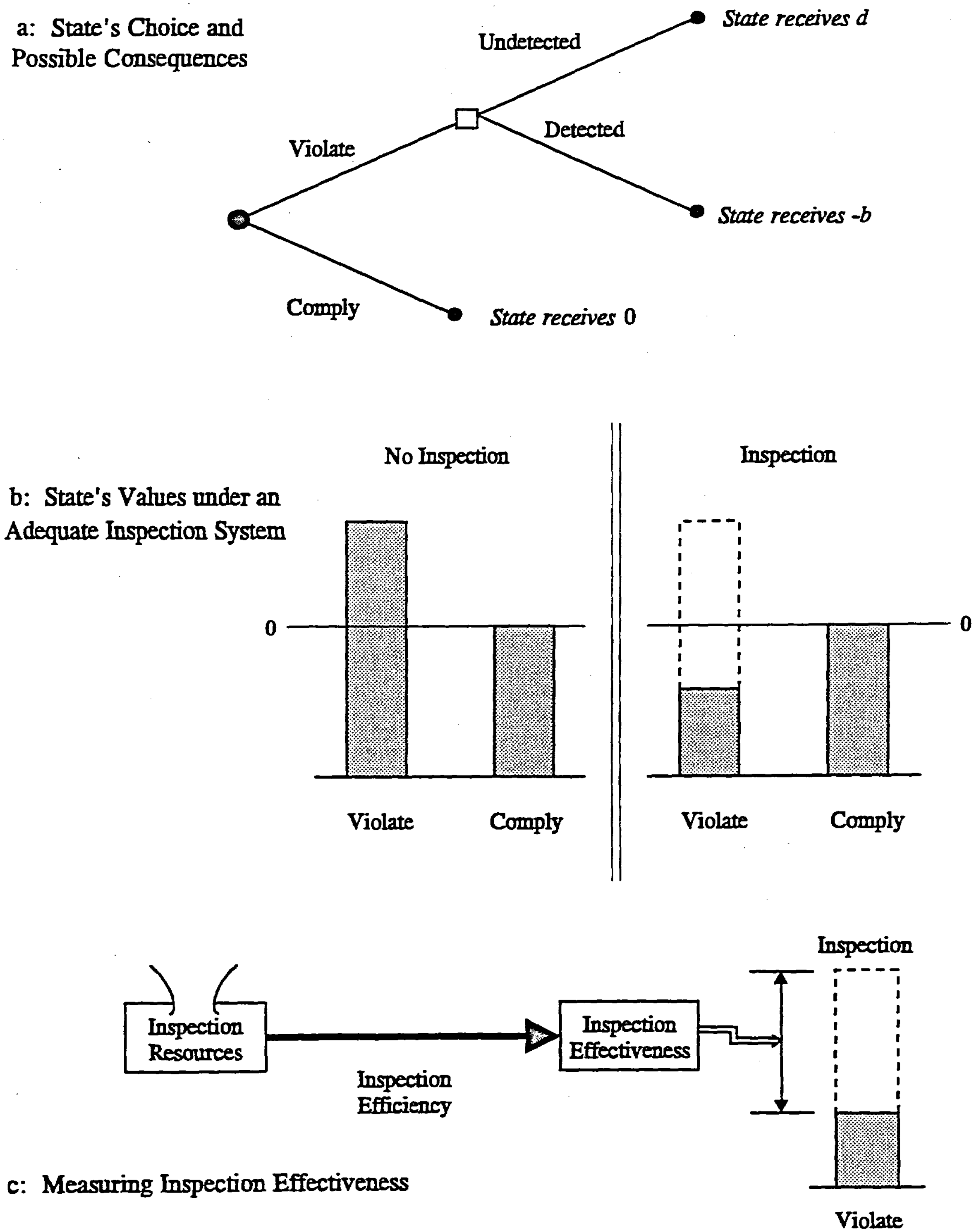
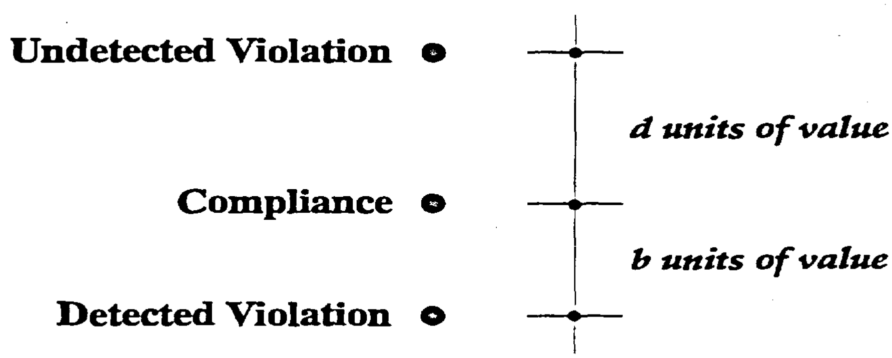


Figure 1: Inspection Effectiveness

values will always be measured relative to its value for compliance. Thus, the indication *"State receives d"* beside the Undetected Violation outcome in Figure 1a means that the state expects that for an undetected violation it would gain D units of value more than what it would gain for Compliance. Similarly, *"State receives $-b$"* at the Detected Violation outcome means that the state anticipates that it would lose b units of value if a violation were detected, relative to what its position would be if it chose Compliance.

In summary, the state's values are as described in the following diagram:



In other words, compliance represents the status quo level, against which all gains and losses are measured. (Technically, value differences are measured in von Neumann-Morgenstern utilities — see [4] for details.)

Clearly, the notation assumes a positive gain for an undetected violation. Under the NPT, this is almost surely not the case for most states most of the time. But note that if an undetected violation is worth less than compliance, then nothing is ever gained by violating, so there is no compliance problem, and safeguards are actually unnecessary. The assumption made here — that an undetected violation improves the state's position — means that the model of Figure 1a, and all of the models below, address situations in which there is a potential compliance problem. Likewise, the notation implies that a detected violation is worth less than compliance. Without this assumption, there is never any reason for the state not to violate. Thus the relative values implied by the notation and described schematically above ensure that the model does not address situations in which a safeguards program is either unnecessary or infeasible.

How does a state make its choice when faced with a decision problem such as the one shown in Figure 1a? The state must evaluate the consequences of each of its alternatives, Violate or Comply. To do this, it must include in its assessment not only the values of the possible consequences of choosing Violate, namely Undetected and Detected, but also it must include the relative likelihood of each possible consequence. Here the safeguards program has an important role. The state's value for violating will be a blend, or weighted average, of the values of the two possible consequences of violating — Detected or Undetected — with the weights reflecting the relative likelihoods of these outcomes.

A typical situation is shown in Figure 1b. If there is no inspection (left side of Figure 1b), the likelihood of detection is small, so the net value of the Violate alternative is large, as it reflects mainly the $+d$ units that the state gains for an undetected violation. But if there is inspection (right side of Figure 1b), the detected violation component, $-b$, has much greater weight, and the net value of the Violate alternative declines.

A role of a safeguards program is quite apparent from Figure 1b. If the state's understanding of the inspection program is sufficient to drive its evaluation of the Violate option below 0 (on the scale adopted here), the state will choose Comply. If, however, the detection of a violation is unlikely enough that the net value of the Violate option remains positive, the state will choose Violate. Note that the dependence on the state's assessment of the inspection program is applicable to the NPT, particularly insofar as IAEA inspection procedures are known in advance.

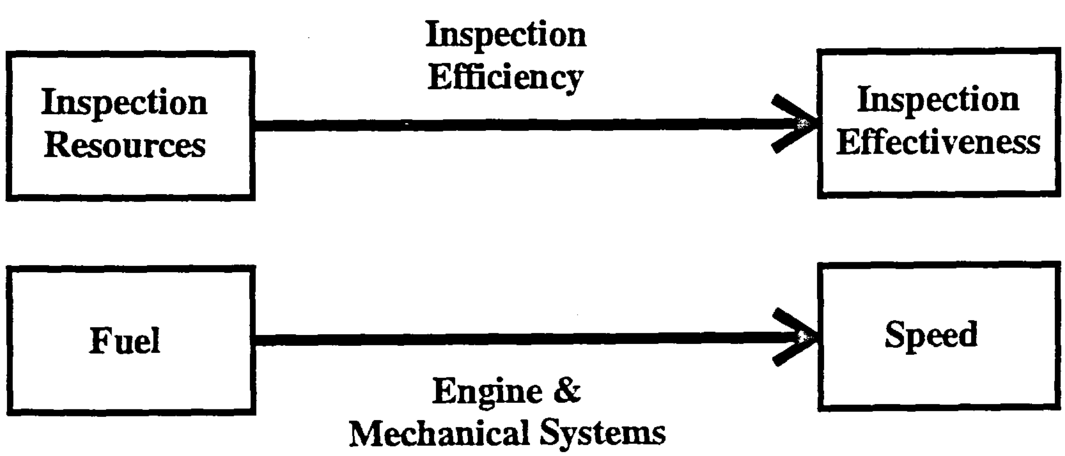
What determines whether the net value of the Violate option exceeds the net value of the Comply option? For now, note that the net value of Violate (relative to Comply) reflects the values for detected and undetected violations. It also reflects the state's assessment of the inspection programme, because the likelihood of detection determines which specific mixture of these two values will be used. Further information is given in Section 3.1, where the question is formulated in a more specific context.

There is, however, another important lesson to be drawn from Figure 1b. An inspection program is effective if it results in a large enough drop in the state's net value for the Violate option. This drop in value (shown on the right side of Figure 1b) is therefore a natural measure of the effectiveness of the programme. In this sense, ***an inspection program deters violations if, and only if, it is sufficiently effective.***

Thus, *inspection effectiveness* refers to the reduction in the state's net evaluation of its option to Violate. This is shown schematically in Figure 1c, where other terminology relating to inspections is also shown. *Inspection resources* are those factors that increase an inspecting agency's capacity to inspect. In the case of the IAEA, for example, inspection resources refer not only to the IAEA's inspection budget, but also to its trained personnel, specialized equipment, etc. Note that inspection resources can be increased in the short term, reallocated, etc. The level of inspection resources allocated to a particular task will be called the *inspection effort* for that task.

The conversion of inspection resources to inspection effectiveness is represented by a quantity called *inspection efficiency* in Figure 1c. The rate of conversion, which reflects procedures, equipment, personnel, etc. is generally not constant. Sometimes (see below, Section 3.3) doubling inspection resources may more than double inspection effectiveness; when this happens, efficiency increases with resources. But most often, increases in inspection resources result in less than proportionate increases in inspection effectiveness; in other words, inspection efficiency usually decreases when resources increase. In general, these "diminishing marginal returns" to inspection resources reflect that, after detection has become fairly certain (or as its likelihood approaches some inevitable ceiling), additional inspection resources must result in smaller and smaller increases in inspection effectiveness. A more formal representation of inspection efficiency is given in the Appendix.

With reference to Figure 1c, a natural analogy to the inspection resources-efficiency-effectiveness relationship can be found in the speed of a vehicle, as shown below.



Just as doubling the fuel consumption does not (usually) double the speed, so doubling the inspection resources does not usually double the inspection effectiveness.

3. Findings

In the Appendix some formal models are developed and analysed, using appropriate tools from Decision Theory and Game Theory. The models were developed to emphasize three different problems relating to the IAEA's allocation of inspection resources, problems which were introduced above. Each of these problems is addressed in one of the subsections to follow. The content of each subsection is based on the analysis in the corresponding section of the Appendix.

3.1 How much inspection effort should be applied in a state?

The model appearing in the Appendix as Problem 1 is in fact a more complete specification of the violate/comply decision problem given in Figure 1a. It provides a simple but useful picture of the mechanism by which inspection programs deter violations. The model admittedly lacks in verisimilitude, but it does demonstrate, at a very fundamental level, what determines whether enforcement is effective enough to deter violations.

Inspection effectiveness, defined in Section 2, is proportional to the probability that a violation would be detected in an inspection. As the Appendix indicates, if a threshold level of inspection effectiveness can be identified, there may be a minimum level of inspection resources that would achieve it, depending on the particular relationship between inspection resources and inspection effectiveness.

The analysis in the Appendix shows that the level of inspection effectiveness necessary to deter violation by a state is related inversely to that state's *value ratio,* defined as

$$\frac{\textit{Loss} \text{ for } \textit{detected violation}}{\textit{Gain} \text{ for } \textit{undetected violation}} = \frac{b}{d}$$

where b and d are the quantities defined in Section 2. To illustrate, the value ratio increases as the loss for a detected violation increases, or as the gain for an undetected violation decreases. The value ratio can thus be interpreted as a "political" parameter, measuring the state's motivation not to violate.

Increasing the state's value ratio, either by increasing the penalty following detection or reducing the benefits of undetected illegal behaviour, decreases the threshold level of inspection effectiveness necessary to deter violation. Conversely, if the state discovers it has less to fear from a detected violation, or more to gain by violating, then the level of inspection effectiveness sufficient for deterrence increases.

It is worth emphasizing that this model shows that the success of a safeguards program depends essentially on a comparison between value ratio, a political parameter, and inspection effectiveness, a "technical" parameter. The value ratio depends strongly on the political preferences, objectives, and interests of the particular state. Inspection effectiveness, on the other hand, is a technical parameter because it is determined by the IAEA's allocation of inspection effort, plus the inspection efficiency, which reflects inspection equipment, procedures, etc. Thus, the success of safeguards does not depend exclusively on the IAEA's technical capability and effort, nor on the state's temptation to violate, but on the interplay of these two factors.

In general, therefore, the level of enforcement effort sufficient to deter a state from violating is determined by the state's value ratio as well as the technical relationship between inspection effort and inspection effectiveness. What determines the value ratio, and how can the threshold effectiveness be decreased? The detection loss component (numerator) of the value ratio reflects the penalties the state would face were it detected in a violation. These might include negative publicity, trade and other economic sanctions, embargoes, and the possibility of military action. But these discouragements to violation are themselves reduced by the actual gain achieved by the illegal behaviour. In the extreme case, when the state is better off even after its violation has been detected (i.e. when the net detection loss, b, is negative), the value ratio falls below zero, and no safeguards program, no matter how effective, can deter the state from illegal behaviour.

The determinants of the violation gain component (denominator) of the value ratio are even more difficult to specify. This quantity reflects the state's own view of how much its position would be improved by successful completion of illegal actions. This might depend on the perceived threat from other states, the history of disputes involving the state, the size of its military establishment, its political objectives and intentions, and geographical factors. The perceived threat from other states depends on the size and power of their military establishments, their objectives and intentions, etc. In short, estimating the denominator of the value ratio is difficult, although the problem is somewhat eased by the need to estimate only the relative sizes of the numerator and denominator.

The minimum necessary level of inspection effectiveness depends on the value ratio, but the actual or attained level of inspection effectiveness depends on factors relating to the IAEA's operations. The primary consideration is the level of resources available for inspection and related activities, such as data analysis. But there are other relevant variables as well — those determining inspection efficiency. For instance, different types of violation may be easier or harder to detect, and some violation technologies may be more transparent than others. Also, the types and amount of information that the IAEA may draw from other sources can vary considerably.

The contribution of the analysis in this section is to establish a link between the technical variables that determine inspection effectiveness and the political considerations that determine the value ratio. This fundamental relationship will now be explored more fully, in connection with the IAEA's problem of how best to allocate inspections within a state, and across states.

3.2 How should inspection effort be allocated among states?

In Section 3.1, some guidance is given in the determination of the level of inspection effectiveness, and therefore of the level of inspection effort, that should be directed against a particular state. Here, in Section 3.2, the focus is on how a fixed total amount of inspection effort should be allocated among many states. The difference in approach is important, because the sum of the inspection resource levels necessary to deter violation in each of the individual states may exceed the total available. As is demonstrated in the Appendix, this shortfall does not necessarily mean that all states cannot be deterred. And, even if it does, it is nonetheless possible to ascertain when some allocations are better than others.

In the first model analysed under Problem 2 in the Appendix, there are two states, each of which chooses to violate or not, based solely on its own interests. The model permits the states to be similar, or different, with respect to the political parameters measuring their propensities to choose violation over compliance. In other words, the model allows for variations between the states in all aspects of their value ratios.

In this model, the IAEA is also a decisionmaker. It possesses only enough resources for a single inspection, and it must choose which state to apply that inspection against. Note that the IAEA is not even allowed the flexibility to spread its inspection resources over both states — it must allocate them all to one state or the other. A further complication is that, because of differences in inspection efficiency, there may be differences in inspection effectiveness between the two states.

Thus, in this simple model there are three decisionmakers — the IAEA, which must decide which of the two states to inspect, and the states themselves, which must decide whether to violate prior to learning the IAEA's inspection plans. A complete game-theoretic solution of this model is given in Theorem 2.1 of the Appendix. The main point, however, is illustrated by Figure 2.

As noted above, the IAEA's single inspection would have one effectiveness level if used against state 1, and another, usually different, level against state 2. Effectiveness against state 1 is measured horizontally in Figure 2, and effectiveness against state 2 is measured vertically. Note that a maximum effectiveness level is shown on both dimensions. The potential of the inspection is thus described by a point representing these two effectiveness levels; this point must lie within the rectangle defined by the minimum and maximum levels in each dimension.

Figure 2 shows the threshold levels of effectiveness for deterring each of the two states for violating, on an individual basis. In other words, if the point decribing the inspection lies to the right of the vertical line labelled "Threshold," then the IAEA can deter state 1 by committing to inspect it. This threshold is simply the maximum effectiveness level determined in Section 3.1. Likewise, the horizontal "Threshold" line in Figure 2 shows the minimum level of effectiveness necessary to deter state 2 by committing to inspect it. Thus, an inspection represented by a point in the upper right-hand quadrant could be used to deter either state.

Unfortunately, there is a catch. The IAEA can deter only one state in this way — by committing to use its one inspection on state 1, say, it is also committing to leave state 2 uninspected and free to violate without threat of detection. In other words, if the IAEA is obligated to identify inspectees in advance, it needs not one, but two sufficiently effective inspections to deter both states.

There may, however, be a way to deter both states using only one inspection. As shown in the Appendix, if the states' value ratios are high enough and the inspection is sufficiently effective against both states (condition (2.9)), then by adjusting the likelihoods of inspection appropriately (condition (2.10)), the IAEA can deter them both. The levels of inspection effectiveness necessary to achieve this joint deterrence are shown in the upper right-hand corner of Figure 2, above and to the right of the curved line.

It is important to note that the IAEA achieves this joint deterrence using only one inspection — but that inspection is sufficiently effective against either state. In summary, the phenomenon captured here is that states are deterred from violating not so much by inspection by the threat of detection that would result from inspection — when the latter threat can be made great enough for both states, then both are deterred.

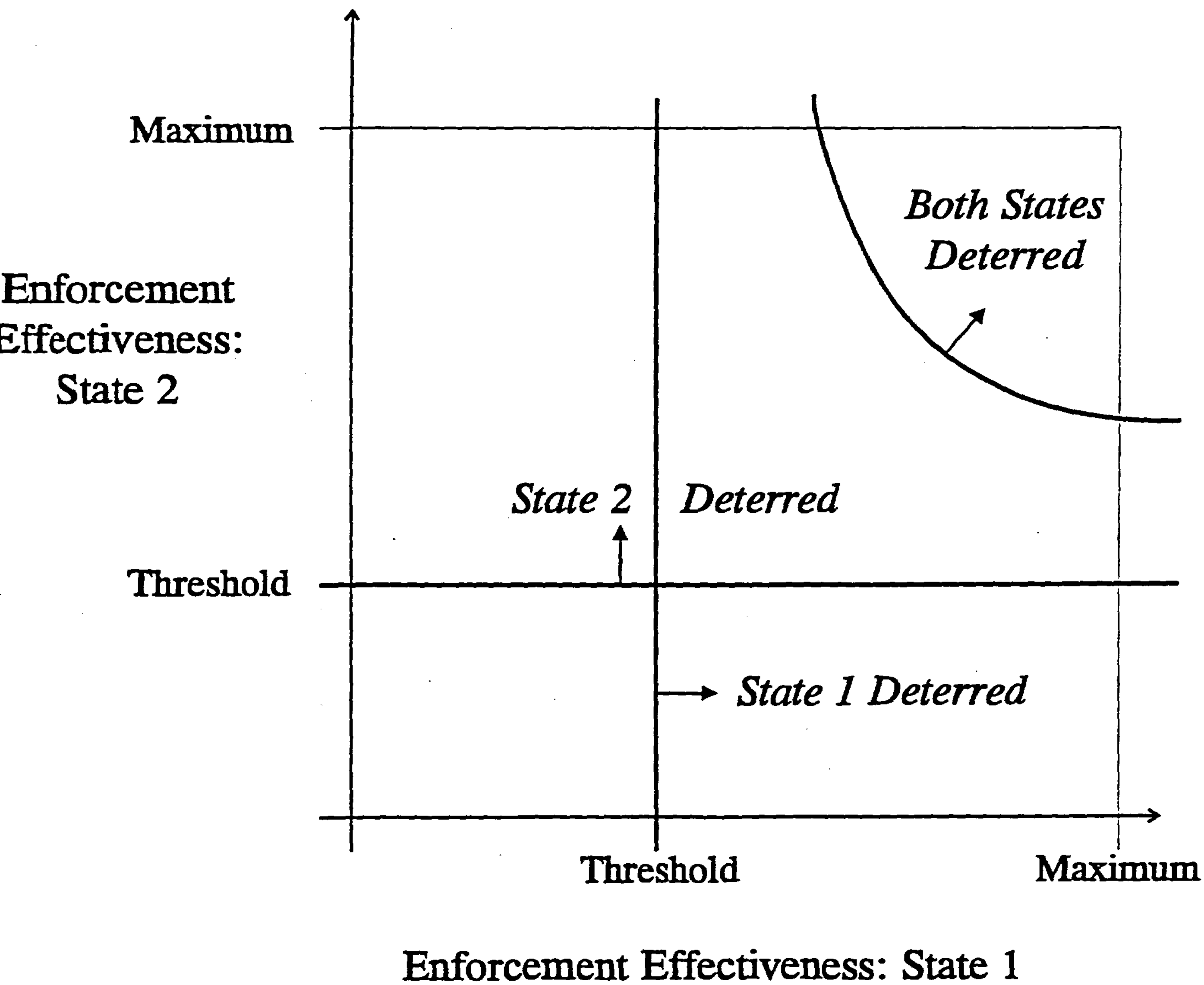


Figure 2: Flexible Use of Inspection Against Two States

Just like the thresholds for deterring a single state, the exact position of the threshold (curved) line for deterring both states depends on the political parameters (value ratios) for both states. Details are found in the Appendix. Likewise, in order to make multiple use of a single inspection in this way, the IAEA must adjust appropriately the likelihoods that the inspection will actually be applied against each state. The IAEA may have considerable latitude in selecting these likelihoods, as discussed in the Appendix and illustrated in Figure A2.

There are good reasons to believe that the phenomenon illustrated in Figure 2 is quite general. In the second model discussed under Problem 2 of the Appendix, a similar situation occurs when there is one inspection to be used on any fixed set of states. The analysis demonstrates when a single, sufficiently effective inspection can deter all states from violating, provided the likelihoods for its use are selected appropriately.

As discussed in the Appendix, our models for Problem 2 incorporate the assumption that inspections cannot be split between or among states. Clearly this is arguable — a more general and realistic view is that, rather than scheduling inspections, what the IAEA does is to allocate inspection effort among states. However, as pointed out in the Appendix, this can only increase the feasibility of using relatively low levels of inspection effort to deter violations in relatively large numbers of states.

The contribution of the analysis in this section is to establish and illustrate the principle that, if inspection effectiveness is relatively high, then smaller levels of inspection effort are sufficient. The key is that predetermined allocation of inspection effort must be avoided. No matter what the size of the safeguards program, unpredictability in its application increases its capability of ensuring that no violations occur.

3.3 How should inspection effort be divided between declared and undeclared sites?

Sections 3.1 and 3.2 provide some guidelines about how much total inspection effort should be directed against a particular state. Now the question of how inspection effort should be allocated within a state is addressed. In summary, the problem involves a state which has declared facilities, and may also have an undeclared (clandestine) program of nuclear weapons development. With respect to this state, IAEA safeguards must include not only inspection of the nuclear fuel cycle associated with declared facilities, but also efforts to detect and identify undeclared facilities and operations. These two aspects of a safeguards program are essentially different, so a natural question is how the IAEA should divide its resources between them.

The model presented and analysed in the Appendix as Problem 3 bears directly on this allocation problem. Though the model represents simple decisions about violating and inspecting at declared vs. undeclared sites, it explicitly includes variable levels of inspection effort — which adds considerably to its complexity. Nevertheless, two important special cases are solved completely, and a partial solution is given for the general case. Important conclusions result concerning the IAEA's relative level of inspection effort against undeclared nuclear weapons development programs.

In the model, there is one state which possesses a declared site and an undeclared site. The state must choose whether to violate or comply, and, if it violates, at which site. (For technical reasons, violation at both sites is not permitted.) The model allows the state's value ratios at its two sites to differ. Such variation would reflect not so much the losses for detected violation (numerators of the value ratios), which are likely to be roughly equal at the two sites because they reflect mainly sanctions and penalties. Rather, the gains for undetected violations (denominators of the value ratios) may differ substantially between declared and undeclared sites, because of differences in timing, scale of operations, availability of equipment, etc. Possible differences in value ratios are an important feature of the model, allowing it to represent the influence of political considerations on choice of violation location.

The model has two decision makers — the state and the IAEA. The IAEA must decide how to allocate a fixed quantity of inspection resources between the inspection of the declared facility and the search for the undeclared facility. It is assumed that neither type of inspection ever yields evidence about the other type of violation, so the IAEA must somehow arrange that their is at least the threat of detection against either type of violation in order to deter it. Thus, the IAEA has flexibility in its decision of where to inspect, and must use it. A further complication is that large differences in inspection effectiveness must be taken into account by the IAEA when it makes its allocation.

Theorems 3.1 and 3.2 of the Appendix contain complete solutions to two special cases of this model — when the relation between inspection resources and effectiveness shows increasing and decreasing returns to scale. As well, the general problem of guaranteeing legal behaviour is addressed, and conditions guaranteeing its solvability are determined. The main point is illustrated in Figure 3.

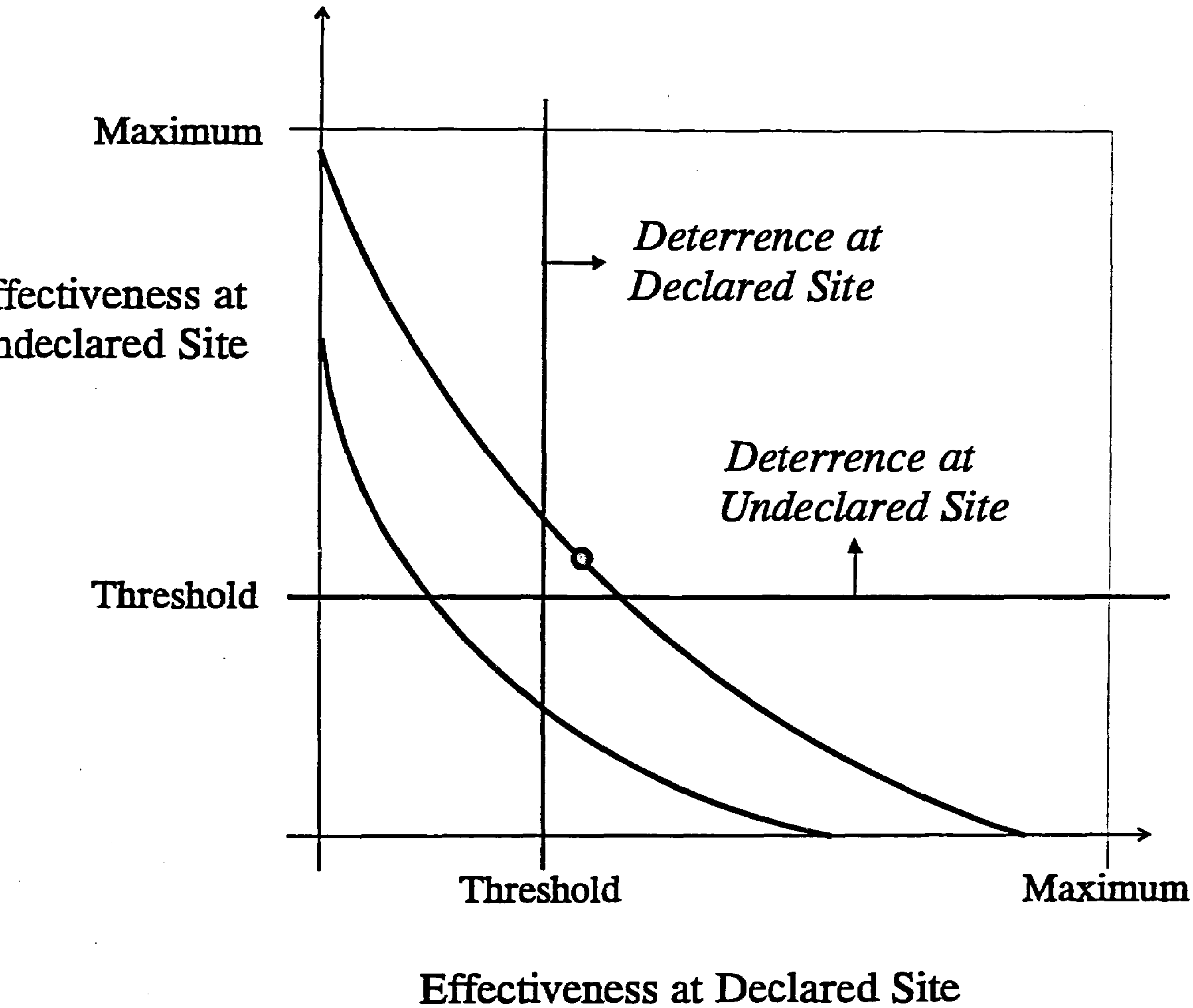


Figure 3: Inspection Effectiveness
Tradeoff For Declared vs. Undeclared Site

Any inspection program chosen by the IAEA (i.e. any allocation of inspection resources over the two tasks) results in a pair of inspection effectiveness ratings — against declared and undeclared sites. These two ratings represent a point in the large rectangle of Figure 3, just as Figure 2 shows a pair of effectiveness ratings against two states.

Many different rating combinations are possible, but, in general, to approach maximum effectiveness at one site requires such a large commitment of resources that effectiveness at the other site falls to the minimum. Figure 3 also shows the thresholds of inspection effectiveness at each site that guarantee deter violation at that site. Again, note that it is the threshold locations that reflect political parameters; technical parameters are reflected in the location of the point representing the pair of effectiveness ratings.

Once again, it is clear that the possibilities for success of the IAEA's safeguards programs depend on the interplay of political and technical parameters. If inspection resources are high enough, the IAEA can allocate them in such a way that both types of violation are deterred. If inspection resources fall short, then no allocation by the IAEA can deter violations at both sites.

The Appendix contains two additional important points that help in understanding the problem of allocating inspection effort between declared and undeclared sites. First, if inspection efficiency is known, Theorems 3.1 and 3.2 may show how the IAEA can achieve its most effective allocation. Roughly, if there are increasing returns to scale, then it is best to concentrate resources; if there are decreasing returns to scale, then it is best to spread resources. Thus, if doubling the inspection resources at a site does not increase effectiveness by very much, then it is best to use lower levels of resources at most sites. In contrast, if doubling the inspection resources at a site more than doubles the effectiveness at the site, then it is best to concentrate resources in a few sites, chosen unpredictably.

The second contribution of these theorems is to determine some formulae for the best possible allocation of resources in cases where resources are insufficient to deter violations at both sites. As illustrated in Figure A3, following a formula that is optimal in a situation of illegal behaviour is never disadvantegeous in situations where legal behaviour is to be expected. Thus, because detailed knowledge may be uncertain in practical cases, efforts to adhere to optimal allocation formulae are always prudent.

As in Section 3.2, this seems to be an illustration of a fairly general phenomenon. As the political parameters shift, the thresholds move — sometimes in concert, and sometimes independently. An allocation which exhibits good balance and is well inside the upper right-hand quadrant of Figure 3 is likely to remain adequate, and an allocation which is barely sufficient on one or both dimensions of Figure 2 is vulnerable. In particular, the political parameters of states at different states of clandestine weapons development programs may differ, suggesting that inspection effort against undeclared sites may need to be increased whenever circumstantial evidence suggests the existence of such a weapons program.

As discussed in the Appendix, our model for Problem 3 incorporates several arguable assumptions. The assumption that violations do not take place simultaneously at declared and undeclared sites is one of these — and it means that conclusions drawn here are suspect whenever that state's motivation to violate is extremely high. But conclusions regarding the deterrence of violations are likely firm. Likewise, the assumption that the two types of inspection are completely unrelated is also incorrect, but this means that levels of resources necessary to deter have probably been overestimated here. The general patterns of optimal allocation would probably remain the same under more realistic models. The simplifications made in developing the model of Problem 3 of the Appendix have been made to increase tractability, and do not seem to alter the conclusions about successful safeguards programs too much.

The contribution of the analysis in this section is to establish and illustrate the principle that the specific allocation of inspection resources within a state is important, and should be subject to the same kinds of consideration of political values and inspection efficiency as the allocation of inspection resources among states. What is important is that all possible types of violation be evaluated both politically and technically, and inspection effort allocated accordingly. Furthermore, change in political factors over the short term may need to be reflected in changes in a safeguards program. For achieving the goals of detecting and deterring violations, the optimal strategy is dynamic.

4. Conclusions

The objective of this study is to develop on which principles to base the IAEA's safeguards programs under the NPT. New ideas for improving effectiveness or decreasing costs of safeguards operations may be especially valuable at this time.

In this study the mathematical tools of Decision Theory and and Game Theory have been applied to models representing

- states' decisions to comply with or violate the NPT, and, if violation is chosen, where to violate;
- the IAEA's decision of where and how much to inspect.

Important determinants of success in scheduling safeguards operations have been identified. The most important conclusions are outlined below.

4.1 The required size of the safeguards program depends on the interplay of political and technical parameters.

The *political* parameter that plays a major role in determining the necessary level of safeguards with respect to a particular state is termed *value ratio.* For states that are motivated to violate, but not at any cost, the value ratio measures the incentive to comply. It reflects the magnitude of the state's perceived decrease in value for a detected violation (in comparison to its value for compliance), relative to the magnitude of its perceived increase in value for an undetected violation (again in comparison to its value for compliance). For example, the value ratio is increased when the sanctions following detection of a violation are made more severe; it is decreased when the state comes to place a higher value on the prohibited weapons.

The *technical* parameter that plays an equal role in determining the success of safeguards is *inspection effectiveness.* Inspection effectiveness is measured by the reduction in the state's anticipated (expected) value were it to choose violation, in consequence of the possibility of detection in an inspection. It was noted that additional *inspection resources* increase inspection effectiveness, but generally at a decreasing rate. In other words, the conversion rate, called *inspection efficiency* is usually lower at higher levels of inspection resources.

The required levels of inspection resources vis-à-vis particular states, and thus the size of safeguards programs, are determined by some very simple comparisons. If inspection effectiveness exceeds a particular threshold depending on the value ratio, a state will be deterred from violating. If inspection effectiveness falls below this threshold, the state will choose to violate. The threshold increases as the value ratio decreases, i.e. as the incentive to comply decreases. The required level of inspection effectiveness thus determines the required level of inspection resources, for each particular state.

The interplay of political considerations (value ratio) and technical considerations (inspection effectiveness) is an important theme running through all of the conclusions of this study.

4.2 If the quality of inspections is high, then fewer of them are needed — unless each state's inspection quota is prespecified.

This conclusion refers to how the IAEA spreads its inspection resources over different states. If there are high quality inspections (i.e. if inspection effectiveness is high), then fewer inspections are needed, provided the IAEA may choose which state to inspect just prior to the actual the inspection. This conclusion reflects the fact that it is not inspection that deters violation, it is the threat of detection. Thus, the possibility of a very effective inspection is as much of a deterrent as the certainty of a less effective inspection. The precise levels of inspection effectiveness at which this phenomenon arises are determined not by technical (effectiveness) considerations, but by the political considerations embodied in the value ratio.

The budgetary implications of this conclusion are important. If, using high quality personnel and equipment, the IAEA can ensure that all of its safeguards operations are highly effective, then fewer of them are needed. However, to profit from this principle, the IAEA must be able to schedule its inspections not according to any predetermined and known pattern, but in an unpredicatable way governed by likelihoods reflecting political considerations (states' value ratios).

4.3 Technical and political considerations at both declared and undeclared sites should determine the allocation of inspection resources between them.

This conclusion bears on how the IAEA should allocate its inspection effort within a state. Because a program designed to detect and identify undeclared facilities and activities is essentially independent of a safeguards program designed for declared facilities, the IAEA must establish both capabilities wherever appropriate. Sufficient levels of inspection effectiveness of both types must be maintained to deter all violations.

In principle, this problem is similar to the simultaneous inspection of two states. The political factors determining a state's incentive to engage in these two forms of violation may differ, so the threshold effectiveness levels may be quite different. Likewise, the two types of inspection may have different efficiency characteristics, complicating the question of how best to split inspection resources between them. The optimal division of inspection resources between the two tasks appears to be difficult in general, though one important idea has been developed here: Resources should be concentrated when they yield increasing returns to scale, and spread when they yield decreasing returns to scale.

References

[1] Avenhaus, Rudolf. "On the Distribution of On-Site Inspections for the Verification of Conventional Forces Agreements." Universität der Bundeswehr München. 1991.

[2] Avenhaus, Rudolf. *Safeguards Systems Analysis.* Plenum, New York, USA. 1986.

[3] von Baechmann, Adolf. "The Treaty on the Non-Proliferation of Nuclear Weapons (1968)." Chapter 6 in *Verification of Current Disarmament and Arms Limitation Agreements: Ways, Means, and Practices,* Serge Sur, editor, UNIDIR, Dartmouth, Brookfield VT, USA. 1991.

[4] Binmore, Ken. *Fun and Games: A Text on Game Theory,* Heath, Toronto, Canada. 1992.

[5] Fischer, David. *Towards 1995: The Prospects for Ending the Proliferation of Nuclear Weapons,* UNIDIR, Dartmouth, Brookfield VT, USA. 1991.

[6] Fudenberg, Drew and Jean Tirole. *Game Theory,* MIT Press, Cambrdige, MA, USA. 1991.

[7] International Atomic Energy Agency, "The Structure and Content of Agreements between the Agency and States Required in Connection with the Treaty on the Non-Proliferation of Nuclear Weapons," INFCIRC/153 (Corrected), Vienna, Austria. 1972.

[8] Kilgour, D. Marc. "New Research in Arms-Control Verification Using Decision Theory," Verification Research Unit, External Affairs and International Trade Canada, Ottawa, Canada. 1991.

[9] Kilgour, D. Marc. "Site Selection for On-Site Inspection in Arms Control," *Arms Control: Contemporary Security Policy,* 13(3), 439-6. 1992.

[10] Schlaifer, Robert. *Analysis of Decisions under Uncertainty,* McGraw-Hill, New York, USA. 1969.

[11] *Treaty on the Non-Proliferation on Nuclear Weapons,* deposited at London, Moscow, and Washington. Opened for signature 1 July 1968.

60984 81800

APPENDIX

This Appendix presents formal models and calculations in support of the accompanying text. Problem 1 is a simple model analysed using Decision Theory; Problems 2 and 3 include several models analysed using Non-cooperative Game Theory.

All payoffs are measured (in von Neumann-Morgenstern utilities) relative to the situation of legal behaviour by the state. In general, a state's relative value for undetected illegal action is denoted $+d$, and its relative value for detected illegal action is $-b$. Detection is always by attribute sampling — if inspected, illegal behaviour is detected with probability $1-\beta$, and missed with probability β. In the case of legal behaviour, there is no possibility of apparent detection. In the models that include IAEA as a player (Problem 2 and 3), IAEA's relative utility for illegal behaviour by the state is $-a$ if detected and $-c$ if not. It is always assumed that

$$0 < a < c, \quad 0 < b, \quad 0 < d, \quad 0 < \beta < 1.$$

Problem 1

This problem refers to a state with one site, which is to be inspected for certain. IAEA is not modelled explicitly. The state's expected value is 0 for legal behaviour and

$$-b(1-\beta) + d\beta$$

for illegal behaviour. The State is deterred from illegal behaviour if and only if

$$1-\beta > \frac{d}{d+b} = \frac{1}{1+(b/d)}. \tag{1.1}$$

A form equivalent to (1.1) and similar to other conditions for guaranteeing legal behaviour that will be obtained below is

$$\frac{d}{b+d} \cdot \frac{1}{1-\beta} < 1. \tag{1.2}$$

Condition (1.1) prescribes the conditions under which inspection is sufficiently effective to guarantee compliance. As noted in the text, condition (1.1) relates a "technical" parameter — the detection probability, $1-\beta$ — to a "political" parameter — the value ratio, b/d.

Detection probability can be understood as proportional to Inspection Effectiveness, as discussed in the text. The relationship of level of inspection resources, ε, to inspection effectiveness is then expressed in a function, such as the one shown (heavy line) in Figure A1, relating inspection resources and detection probability. In this context, condition (1.1) shows that there is a level of inspection resources that is just sufficient to deter; this level is ε_0, defined by

$$1-\beta(\varepsilon_0) = \frac{1}{1+b/d}.$$

In general, when the state knows that the level of inspection resources exceeds this threshold, it behaves legally; when it knows that the level of inspection resources falls below this level, it behaves illegally.

Problem 2

This problem concerns the efficient use of inspection resources to induce legal behaviour. It illustrates that states are deterred not by inspection, but by the threat of inspection.

First, suppose that there are two states, called state 1 and state 2, that may or may not behave legally under the NPT. The IAEA spends its inspection effort either entirely within the first state, or entirely within the second. Let $1-\beta_i$ be the probability of detecting an illegal action occurring in state i if state i is inspected (i = 1, 2), given that an illegal action is in fact occurring there.

The payoffs to the three players are: for the IAEA, the sum over both states of

$$\begin{aligned} &0 \text{ for legal behaviour of a state} \\ &-a \text{ for detected illegal behaviour of a state} \\ &-c \text{ for undetected illegal behaviour of a state} \end{aligned} \tag{2.1}$$

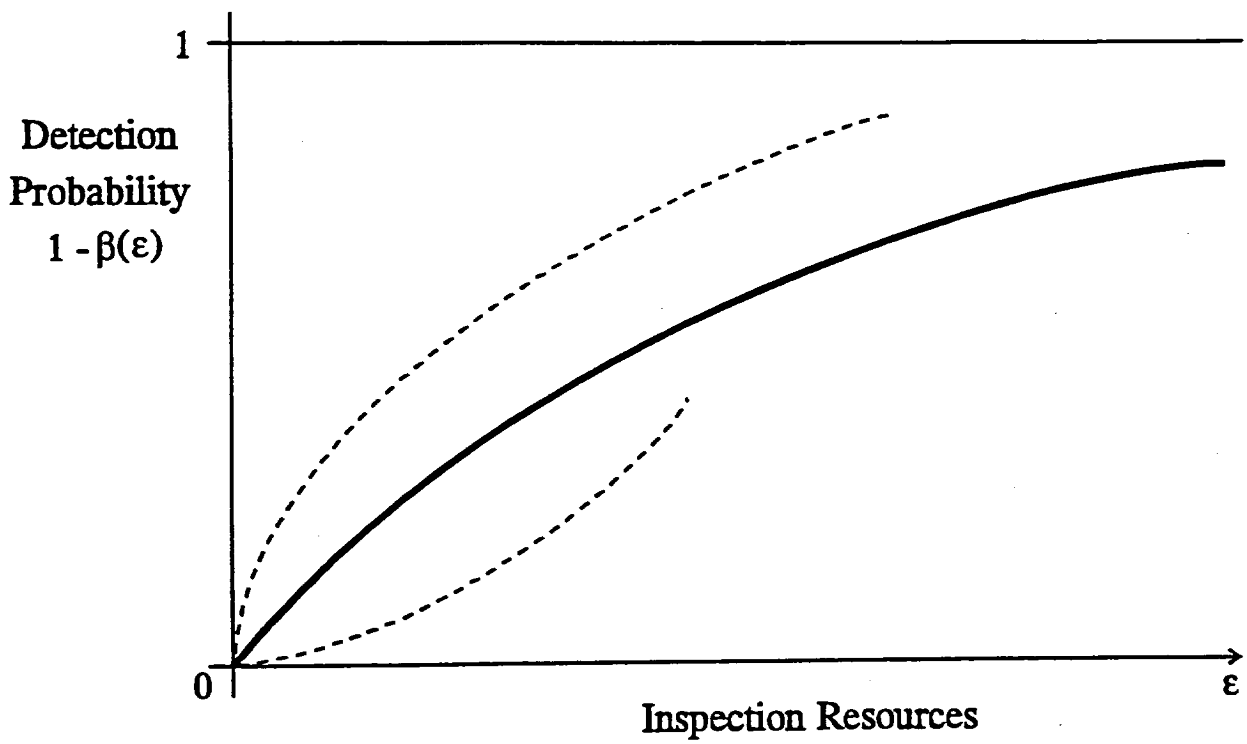
and for state i (i = 1, 2),

$$\begin{aligned} &0 \text{ for legal behaviour} \\ &-b_i \text{ for detected illegal behaviour} \\ &d_i \text{ for undetected illegal behaviour.} \end{aligned} \tag{2.2}$$

Here, we assume

$$0<a<c, \quad 0<b_i, \quad 0<d_i, \quad i = 1,\ 2. \tag{2.3}$$

Figure A1: Violation Detection Probability vs. Inspection Resources



The IAEA inspects state 1 with probability p and state 2 with probability $1-p$. Let state i behave illegally with probability q_i, and legally with probability $1-q_i$, $i = 1, 2$.

The unconditional expected payoffs are therefore

$$\{[-a-(c-a)\beta_1]q_1(1-q_2)-c(1-q_1)q_2+[-a(1-\beta_1)-c\beta_1-c]q_1q_2\}p$$
$$+\{[-a-(c-a)\beta_2]q_2(1-q_1)-c(1-q_2)q_1+[-a(1-\beta_2)-c\beta_2-c]q_1q_2\}(1-p)$$
$$=(c-a)[(1-\beta_1)q_1-(1-\beta_2)q_2]\cdot p+(c-a)(1-\beta_2)q_2-c(q_1+q_2)$$
$$\equiv E_0(p, q_1, q_2) \tag{2.4}$$

to the IAEA;

$$\{[-b_1+(b_1+d_1)\beta_1]q_1\}p+d_1q_1(1-p)$$
$$=[-(b_1+d_1)(1-\beta_1)p+d_1]q_1 \equiv E_1(p, q_1, q_2) \tag{2.5}$$

to state 1; and

$$d_2q_2p+\{[-b_2+(b_2+d_2)\beta_2]q_2\}(1-p)$$
$$=[-(b_2+d_2)(1-\beta_2)(1-p)+d_2]q_2 \equiv E_2(p, q_1, q_2) \tag{2.6}$$

to state 2.

We assume that the three players do not cooperate. Thus, we model our problem as a non-cooperative three-person game $(\{p\}, \{q_1\}, \{q_2\}, F_0, F_1, F_2)$ with strategy sets and payoffs as given above. The equilibria (p^*, q_1^*, q_2^*) of this game are determined by the Nash conditions

$$E_0(p^*, q_1^*, q_2^*) \geq E_0(p, q_1^*, q_2^*)\ \forall p$$
$$E_1(p^*, q_1^*, q_2^*) \geq E_1(p^*, q_1, q_2^*)\ \forall q_1$$
$$E_2(p^*, q_1^*, q_2^*) \geq E_2(p^*, q_1^*, q_2)\ \forall q_2. \tag{2.7}$$

Using $c-a>0$ and (3.4), (3.5), and (3.6), these inequalities are equivalent to

$$[(1-\beta_1)q_1^*-(1-\beta_2)q_2^*]\cdot p^* \geq [(1-\beta_1)q_1^*-(1-\beta_2)q_2^*]\cdot p\ \ \forall p \tag{2.8a}$$

$$[d_1-(b_1+d_1)(1-\beta_1)p^*]\cdot q_1^* \geq [d_1-(b_1+d_1)(1-\beta_1)p^*]\cdot q_1\ \ \forall q_1 \tag{2.8b}$$

$$[d_2-(b_2+d_2)(1-\beta_2)(1-p^*)]\cdot q_2^*$$
$$\geq [d_2-(b_2+d_2)(1-\beta_2)(1-p^*)]\cdot q_2\ \ \forall q_2. \tag{2.8c}$$

The theorem to follow generalizes results in [9].

Theorem 2.1

The equilibria (p^*, q_1^*, q_2^*) and the equilibrium payoffs $E_i^* = E_i(p^*, q_1^*, q_2^*)$, $i = 0, 1, 2$, of the game described above are as follows:

(i) If $$\frac{d_1}{b_1+d_1} \cdot \frac{1}{1-\beta_1} + \frac{d_2}{b_2+d_2} \cdot \frac{1}{1-\beta_2} \le 1, \tag{2.9}$$

$$\frac{d_1}{b_1+d_1} \cdot \frac{1}{1-\beta_1} \le p^* \le 1 - \frac{d_2}{b_2+d_2} \cdot \frac{1}{1-\beta_2}, \quad q_1^* = q_2^* = 0. \tag{2.10}$$

$$E_0^* = E_1^* = E_2^* = 0 \tag{2.11}$$

(ii) If $$\frac{d_1}{b_1+d_1} \cdot \frac{1}{1-\beta_1} + \frac{d_2}{b_2+d_2} \cdot \frac{1}{1-\beta_2} > 1, \text{ there are four cases:} \tag{2.12}$$

If $\beta_1 < \beta_2$ and $$\frac{d_1}{b_1+d_1} \cdot \frac{1}{1-\beta_1} \le 1, \tag{2.13}$$

$$p^* = \frac{d_1}{b_1+d_1} \cdot \frac{1}{1-\beta_1}, \quad q_1^* = \frac{1-\beta_2}{1-\beta_1}, \quad q_2^* = 1; \tag{2.14}$$

If $\beta_1 < \beta_2$ and $$\frac{d_1}{b_1+d_1} \cdot \frac{1}{1-\beta_1} > 1, \tag{2.15}$$

$$p^* = 1, \quad q_1^* = q_2^* = 1; \tag{2.16}$$

If $\beta_1 > \beta_2$ and $$\frac{d_2}{b_2+d_2} \cdot \frac{1}{1-\beta_2} \le 1, \tag{2.17}$$

$$p^* = 1 - \frac{d_2}{b_2+d_2} \cdot \frac{1}{1-\beta_2}, \quad q_1^* = 1, \quad q_2^* = \frac{1-\beta_1}{1-\beta_2}; \tag{2.18}$$

If $\beta_1 > \beta_2$ and $$\frac{d_2}{b_2+d_2} \cdot \frac{1}{1-\beta_2} > 1, \tag{2.19}$$

$$p^* = 0, \quad q_1^* = q_2^* = 1. \tag{2.20}$$

The equilibrium payoffs are obtained by inserting the equilibria into (2.4), (2.5) and (2.6).

Proof

(i) For $q_1^* = q_2^* = 0$ (2.8b) and (2.8c) are equivalent to

$$d_1 - (b_1+d_1)(1-\beta_1)p^* \le 0 \quad \text{and} \quad d_2 - (b_2+d_2)(1-\beta_2)(1-p^*) \le 0.$$

These lead immediately to the assertion.

(ii) If $\beta_1 < \beta_2$ and $\frac{d_1}{b_1+d_1} \cdot \frac{1}{1-\beta_1} \le 1$, (2.8a) is satisfied iff

$$(1-\beta_1)q_1^* = (1-\beta_2)q_2^*$$

which holds when

$$q_2^* = 1, \quad q_1^* = \frac{1-\beta_2}{1-\beta_1}.$$

In this case, (2.8b) is satisfied when

$$d_1-(b_1+d_1)(1-\beta_1)p^* = 0, \text{ i.e. } p^* = \frac{d_1}{b_1+d_1}\cdot\frac{1}{1-\beta_1}$$

and (2.8c) is satisfied provided

$$d_2-(b_2+d_2)(1-\beta_2)(1-p^*) > 0, \text{ i.e. } p^* > 1-\frac{d_2}{b_2+d_2}\cdot\frac{1}{1-\beta_2},$$

which is fulfilled because of (2.12) and (2.13). The remaining cases are treated similarly.

■

Now set $b_1=b_2=b$, and let b^* be defined by

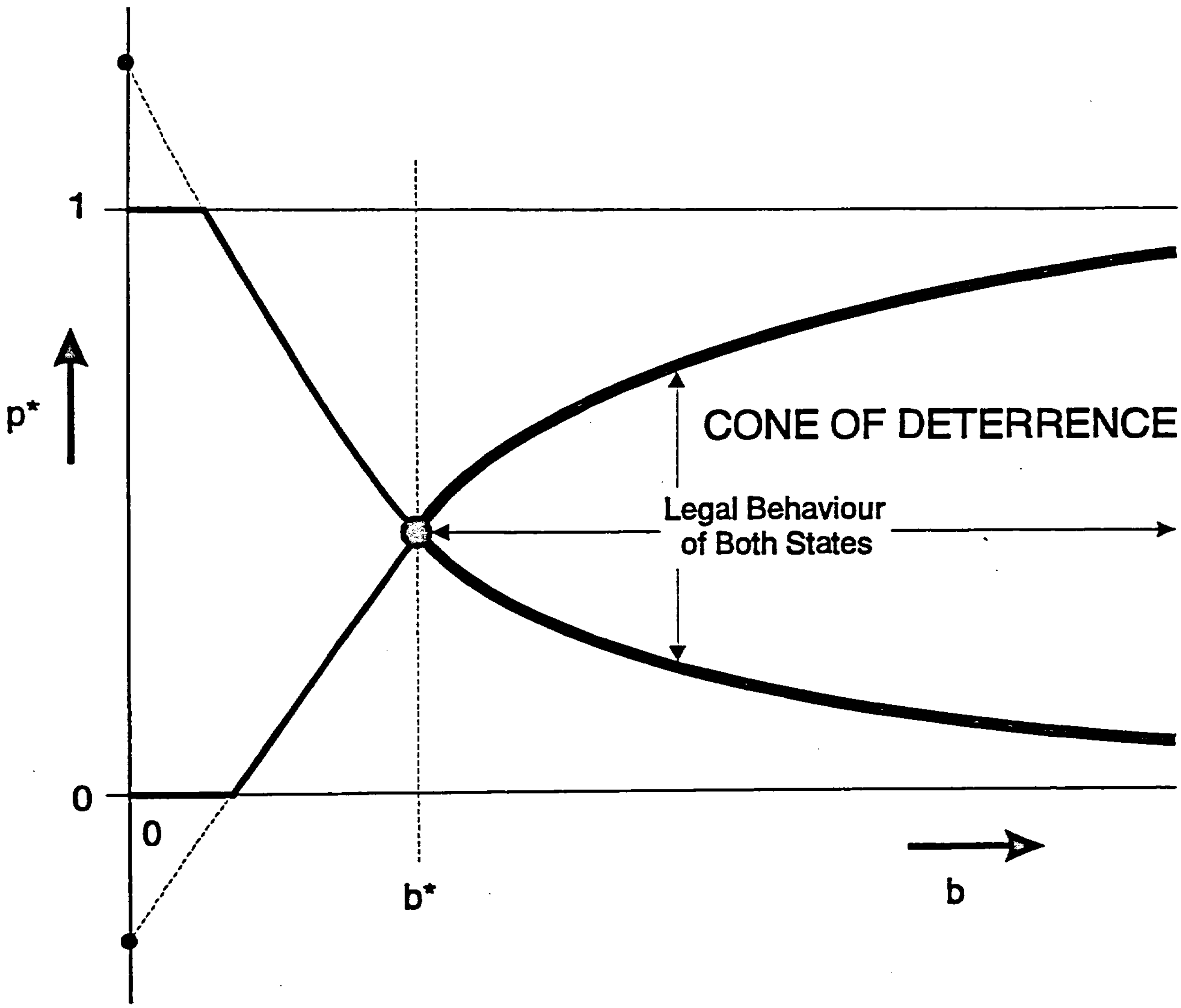
$$\frac{d_1}{b^*+d_1}\cdot\frac{1}{1-\beta_1}+\frac{d_2}{b^*+d_2}\cdot\frac{1}{1-\beta_2} = 1. \tag{2.21}$$

This value is indicated in Figure A2, which shows how a properly chosen inspection scheme (which selects p within the "Cone of Deterrence" introduced in [10]) can deter both states. Note that a state's penalty for detected violation must be sufficiently severe (at least b^*) in order for this common deterrence to be effective. The greater the penalty, the wider the latitude in choosing inspection schemes to induce legal behaviour.

It is important to note the relationship of the condition for simultaneous deterrence of both states, (2.9), with (1.2). Condition (1.2) indicates that an inspection deters a state from violating if and only if an index is small enough. This index reflects not only technical aspects of inspection $(1-\beta)$, but also the state's political values $(d/(b+d))$. Condition (2.9) indicates that two states can be deterred by (the threat of) one inspection if the sum of each state's index falls below the same threshold. Note that in Problem 1 [condition (1.2)] the state knows that it will be inspected, whereas in Problem 2 [condition (2.9)] both states know that only one state will be inspected — but nonetheless the threat of inspection is sufficient to deter.

Of course an inspection must be quite effective to do "double duty" in this way, as illustrated by Figure 2 in the text. Here, enforcement effectiveness against state i is taken to be proportional to $1-\beta_i$; any inspection opportunity which could be used in either state then corresponds to a point in the two-dimensional space shown. An inspection that can be used only against state

Figure A2: Probability of Inspecting State 1 vs. Common Penalty



i will deter state i from violating provided it is moderately effective against state i; but if the inspection could be used, and would be sufficiently effective, against either state, then it deters them both. The individual states' thresholds are determined by (1.2), and the joint threshold (curved line) by (2.9).

It is not difficult to generalize condition (2.9) to a model in which there are n states to be deterred by the threat of a single inspection. In analogy to (2.5) and (2.6), state i's unconditional expected payoff is

$$\{[-b_i(1-\beta_i) + d_i\beta_i]q_i\}p_i + d_iq_i(1-p_i) \equiv F_i(q_i, p_i). \tag{2.22}$$

Then it can be shown that, for any $i = 1, 2, \cdots n,$ the Nash condition

$$F_i(q_i^*, p_i^*) \geq F_i(q_i, p_i^*) \forall q_i \tag{2.23}$$

is satisfied with $q_i^* = 0$ if and only if

$$\frac{d_i}{b_i + d_i} \frac{1}{1-\beta_i} \leq p_i^* \tag{2.24}$$

This generalization of (2.10) describes the "Cone of Deterrence" for a single inspection spread over n states. It implies that a necessary and sufficient condition for n states to be deterrable using one inspection is

$$\sum_{i=1}^{n} \frac{d_i}{b_i + d_i} \frac{1}{1-\beta_i} \leq 1. \tag{2.25}$$

Condition (2.25) is a further extension of (1.2). Note that the same index,

$$\frac{d_i}{b_i + d_i} \frac{1}{1-\beta_i}$$

continues to measure "desirability of violating"; as long as the total desirability of violating is low enough and the inspection scheme (the collection of values of p_i) is well-chosen, no violation actually takes place. The threat of inspection is enough to guarantee compliance.

All of the calculations of Problem 2 have been carried out under the assumption that the IAEA must apply all its inspection effort in one state only. In other words, the IAEA selects which state is to be inspected, and then inspects only in that state — it cannot spread some inspection effort over other states. It is appropriate to end with a comment on this assumption.

The models analysed here are interesting because they are natural generalizations of the simple model of Problem 1. Furthermore, the all-or-nothing inspection policy does represent an optimal strategy sometimes (see Problem 3). However there are other cases when it is less than optimal; but if so, (2.9) and (2.25) are too strict, and inspections that are even less effective (in some sense) can deter all states. Nonetheless the main justification for these models is their simple form, and the clear and intuitive conclusions that they imply.

Problem 3

This problem addresses the distribution of inspection resources within a state. It assumes that a state which intends to violate may choose exactly how it will violate — and also that the most effective type of inspection depends on the type of violation.

Consider a model focusing on the behaviour of a state, that has a declared and an undeclared site for handling nuclear material. Assume that the state behaves illegally — in the sense of the NPT — in *at most* one of the two sites.

The IAEA, spends inspection effort ε_1 at site 1, the declared site, and $\varepsilon_2 = \varepsilon - \varepsilon_1$ at site 2, the undeclared site, where ε is its total available inspection effort. Let $1 - \beta_i(\varepsilon)$ be the probability of detecting an illegal action at site i, if it is inspected with effort ε_i $(i = 1, 2)$. Here $1 - \beta_1(\cdot)$ and $1 - \beta_2(\cdot)$ are detection probability functions as illustrated in Figure A1. The payoffs to (IAEA, state) are

$$\begin{aligned} &(0, 0) \text{ for legal behaviour of the state} \\ &(-a_i, -b_i) \text{ for detected illegal action at site } i \\ &(-c_i, d_i) \text{ for undetected illegal action at site } i. \end{aligned} \tag{3.1}$$

In this case, we assume

$$0 < a_i < c_i, \quad 0 < b_i, \quad 0 < d_i \quad \text{for } i = 1, 2. \tag{3.2}$$

The IAEA chooses its inspection effort at site 1, ε_1, according to a cumulative probability distribution $F(\cdot)$, with support in $[0, \varepsilon]$. The state behaves illegally with probability q_1 at the first site, and q_2 at the second, where

$$q_1 + q_2 \le 1. \tag{3.3}$$

Thus $q_1 = q_2 = 0$ means legal behaviour at both sites. The unconditional expected payoff to the IAEA is

$$E_I(F, q_1, q_2)$$
$$= \int_0^{\varepsilon} [q_1 \cdot (-a_1 - (c_1 - a_1)\beta_1(\varepsilon_1)) + q_2 \cdot (a_2 - (c_2 - a_2)\beta_2(\varepsilon - \varepsilon_1))] dF(\varepsilon_1) \tag{3.4}$$

while the state's expected payoff is

$$E_S(F, q_1, q_2)$$
$$= \int_0^{\varepsilon} [q_1(-b_1 + (b_1 + d_1)\beta_1(\varepsilon_1)) + q_2(-b_2 + (b_2 + d_2)\beta_2(\varepsilon - \varepsilon_1))] dF(\varepsilon_1). \tag{3.5}$$

We assume that the players do not cooperate. Thus, we model Problem 3 as a non-cooperative two-person game $(\{F\}, \{q_1, q_2\}, E_I, E_S)$ with strategies and payoffs as given above. The equilibrium solution (F^*, q_1^*, q_2^*) of this game is determined by the Nash conditions

$$E_I(F^*, q_1^*, q_2^*) \geq E_I(F, q_1^*, q_2^*) \; \forall F \tag{3.6}$$

$$E_S(F^*, q_1^*, q_2^*) \geq E_S(F^*, q_1, q_2) \; \forall q_1, q_2 \text{ satisfying (3.3)}, \tag{3.7}$$

where $E_I(F, q_1, q_2)$ and $E_S(F, q_1, q_2)$ are given by (3.4) and (3.5). Two equilibrium solutions are now preented, depending on the analytical forms of the detection probabilities $1 - \beta_i(\cdot)$. The first generalizes results previously obtained in [1] and [2].

Theorem 3.1 (Concentration of inspection effort)
Let $1 - \beta_1(\varepsilon_1)$ and $1 - \beta_2(\varepsilon - \varepsilon_1)$ have the properties

$$1 - \beta_1(0) = 1 - \beta_2(0) = 0 \tag{3.8}$$

$$\frac{d}{d\varepsilon_1}(1 - \beta_1(\varepsilon_1)) > 0, \; \frac{d}{d\varepsilon_1}(1 - \beta_1(\varepsilon - \varepsilon_1)) < 0 \text{ for all } \varepsilon_1 \text{ with } 0 \leq \varepsilon_1 \leq \varepsilon. \tag{3.9}$$

Furthermore, suppose that $1 - \beta_1(\varepsilon_1)$ and $1 - \beta_2(\varepsilon - \varepsilon_1)$ are strictly convex, i.e.,

$$\frac{d^2}{d\varepsilon_1^2}(1 - \beta_1(\varepsilon_1)) > 0, \; \frac{d^2}{d\varepsilon_1^2}(1 - \beta_2(\varepsilon - \varepsilon_1)) > 0 \text{ for all } \varepsilon_1 \text{ with } 0 \leq \varepsilon_1 \leq \varepsilon. \tag{3.10}$$

Define

$$1 - \beta_i(\varepsilon) = 1 - \beta_i \text{ for } i = 1, 2. \tag{3.11}$$

Then the equilibria (F^*, q_1^*, q_2^*) as well as the equilibrium payoffs E_I^* and E_S^* of the game described above are given by

$$F^*(\varepsilon_1) = \begin{cases} 0 & \varepsilon < 0 \\ 1 - p^* \quad \text{for} & 0 \leq \varepsilon_1 < \varepsilon, \\ 1 & \varepsilon_1 \geq \varepsilon \end{cases} \tag{3.12}$$

where p^*, q_1^* and q_2^* are as follows:

(i) If $\frac{d_1}{b_1+d_1} \cdot \frac{1}{1-\beta_1} + \frac{d_2}{b_2+d_2} \cdot \frac{1}{1-\beta_2} \leq 1,$ (3.13)

$$\frac{d_1}{b_1+d_1} \cdot \frac{1}{1-\beta_1} \leq p^* \leq 1 - \frac{d_2}{b_2+d_2} \cdot \frac{1}{1-\beta_2}, \quad q_1^* = q_2^* = 0 \tag{3.14}$$

$$E_I^* = E_S^* = 0. \tag{3.15}$$

(ii) If $\frac{d_1}{b_1+d_1} \cdot \frac{1}{1-\beta_1} + \frac{d_2}{b_2+d_2} \cdot \frac{1}{1-\beta_2} > 1$ and $d_1 > d_2$; (3.16)

If $d_1 - d_2 > (b_1+d_1)(1-\beta_1)$ (3.17)

$$p^* = 1, \; q_1^* = 1, \; q_2^* = 0 \tag{3.18}$$

$$E_I^* = -c_1 + (c_1 - a_1) \cdot (1-\beta_1) \tag{3.19}$$

$$E_S^* = d_1 - (b_1+d_1) \cdot (1-\beta_1)$$

If $d_1 - d_2 < (b_1+d_1) \cdot (1-\beta_1)$ (3.20)

$$p^* = \frac{d_1 - d_2 + (b_2+d_2)(1-\beta_2)}{(b_1+d_1) \cdot (1-\beta_1) + (b_2+d_2)(1-\beta_2)} \tag{3.21}$$

$$q_1^* = \frac{(c_2-a_2) \cdot (1-\beta_2)}{(c_1-a_1) \cdot (1-\beta_1) + (c_2-a_2) \cdot (1-\beta_2)} = 1 - q_2^* \tag{3.22}$$

$$E_I^* = \frac{-\frac{c_1}{c_1-a_1} \cdot \frac{1}{1-\beta_1} - \frac{c_2}{c_2-a_2} \cdot \frac{1}{1-\beta_2} + 1}{\frac{1}{c_1-a_1} \cdot \frac{1}{1-\beta_1} + \frac{1}{c_2-a_2} \cdot \frac{1}{1-\beta_2}} \tag{3.23}$$

$$E_S^* = \frac{\frac{d_1}{b_1+d_1} \cdot \frac{1}{1-\beta_1} + \frac{d_2}{b_2+d_2} \cdot \frac{1}{1-\beta_2} - 1}{\frac{1}{b_1+d_1} \cdot \frac{1}{1-\beta_1} + \frac{1}{b_2+d_2} \cdot \frac{1}{1-\beta_2}}. \tag{3.24}$$

Proof

(i) With (3.12) and (3.14) equilibrium condition (3.6) is identically fulfilled, whereas (3.7) becomes

$$0 \geq [q_1 \cdot (-b_1 + (b_1+d_1)\beta_1(\varepsilon) + q_2 \cdot (-b_2 + (b_2+d_2)\beta_2(0)] \cdot p^*$$
$$+ [q_1 \cdot (-b_1 + (b_1+d_1)\beta_1(0) + q_2 \cdot (-b_2 + (b_2+d_2)\beta_2(\varepsilon))] \cdot (1-p^*)$$

for all q_1, q_2 with $q_1 + q_2 \le 1$. Using (3.8) and (3.10) this is equivalent to

$$0 \ge [-(b_1+d_1)(1-\beta_1)\cdot p^* + d_1]\cdot q_1$$
$$+ [-(b_2+d_2)(1-\beta_2)\cdot(1-p^*)+d_2]\cdot q_2.$$

This inequality is true for all q_1, q_2 such that $q_1+q_2 \le 1$ if and only if

$$0 \ge -(b_1+d_1)(1-\beta_1)\cdot p^* + d_1$$
$$0 \ge -(b_2+d_2)(1-\beta_2)\cdot(1-p^*)+d_2,$$

which is equivalent to (3.14).

(ii) With (3.12) and (3.18), equilibrium conditions (3.6) and (3.7) become

$$-c_1 + (c_1-a_1)\cdot(1-\beta_1) \ge \int_0^{\varepsilon}(-c_1+(c_1-a_1)\cdot(1-\beta(\varepsilon_1))\,dF(\varepsilon_1)$$

$$-b_1+(b_1+d_1)\cdot\beta_1 \ge q_1\cdot(-b_1+(b_1+d_1)\cdot\beta_1)+q_2\cdot d_2$$

for all distributions F on $[0, \varepsilon]$ and all q_1, q_2 such that $q_1+q_2\le 1$. The first inequality is equivalent to

$$1-\beta_1 \ge \int_0^{\varepsilon}(1-\beta_1(\varepsilon_1))\,dF(\varepsilon_1) \text{ for all } F,$$

which is fulfilled because of (3.9). The second inequality follows from (3.17). With (3.12), (3.21) and (3.22), equilibrium conditions (3.5) and (3.7) reduce to

$$[q_1^*(-c_1+(c_1-a_1)(1-\beta_1))+q_2^*(-c_2)]p^* + [q_1^*(-c_1)+q_2^*(-c_2+(c_1-a_2)(1-\beta_2)](1-p^*)$$
$$\ge \int_0^{\varepsilon}[q_1^*(-c_1+(c_1-a_1)(1-\beta_1(\varepsilon_1))+q_2^*(-c_2+(c_2-a_2)(1-\beta_1(\varepsilon-\varepsilon_1))]\,dF(\varepsilon_1)$$

for all F, and

$$[q_1^*(-b_1+(b_1+d_1)\beta_1)+q_2^*d_2]p^* + [q_1^*d_1+q_2^*(-b_2+(b_2+d_2)\beta_2)](1-p^*)$$
$$\ge [q_1(-b_1+(b_1+d_1)\beta_1)+q_2d_2]p^* + [q_1d_1+q_2(-b_2+(b_2+d_2)\beta_2)](1-p^*)$$

for all q_1, q_2 such that $q_1+q_2\le 1$. The second inequality is always true by (3.21). The first inequality is equivalent to

$$E_I^* \ge \int_0^{\varepsilon}[q_1^*\cdot(c_1-a_1)\cdot(1-\beta_1(\varepsilon_1))+q_2^*\cdot(c_2-a_2)\cdot(1-\beta_2(\varepsilon-\varepsilon_1))]\,dF(\varepsilon_1)$$
$$+q_1^*\cdot(-c_1)+q_2^*\cdot(-c_2)$$

for all F. Because of (3.22) and (3.23) we have

$$q_1^* \cdot (-c_1) + q_2^* \cdot (-c_2) = E_I^* - \frac{(c_1 - a_1)(1-\beta_1) \cdot (c_2 - a_2)(1-\beta_2)}{(c_1 - a_1)(1-\beta_1) + (c_2 - a_2)(1-\beta_2)},$$

so the inequality is equivalent to

$$1 \geq \int_0^{\varepsilon} \left[\frac{1-\beta_1(\varepsilon_1)}{1-\beta_1} + \frac{1-\beta_2(\varepsilon-\varepsilon_1)}{1-\beta_2} \right] dF(\varepsilon_1)$$

for all F. Furthermore, because of (3.8), (3.9), and (3.10), the function

$$G(\varepsilon_1) = \frac{1-\beta_1(\varepsilon_1)}{1-\beta_1} + \frac{1-\beta_2(\varepsilon-\varepsilon_1)}{1-\beta_2} \quad \text{for } 0 \leq \varepsilon_1 \leq \varepsilon$$

satisfies

$$G(0) = G(\varepsilon) = 1$$

and is strictly convex in ε_1 for $0 \leq \varepsilon_1 \leq \varepsilon$. Therefore, the inequality is true for all F. ■

Part (i) of Theorem 3.1 defines necessary and sufficient conditions for all violation to be deterred, similar to Theorem 2.1(i). To illustrate, suppose that $b_1 = b_2 = b > 0$, and assume $d_1 > d_2$ and

$$d_1 - d_2 \leq (b + d_1) \cdot (1-\beta_1) \,\forall\, b \geq 0, \text{ i.e. } d_2 \geq d_1 \cdot \beta_1.$$

Then, according to (3.21), p^* is a function of b given by

$$p^* = \frac{d_1 - d_2 + (b + d_2)(1-\beta_2)}{(b + d_1)(1-\beta_1) + (b + d_2)(1-\beta_2)}$$

for $0 \leq b \leq b^*$, where b^* is defined by

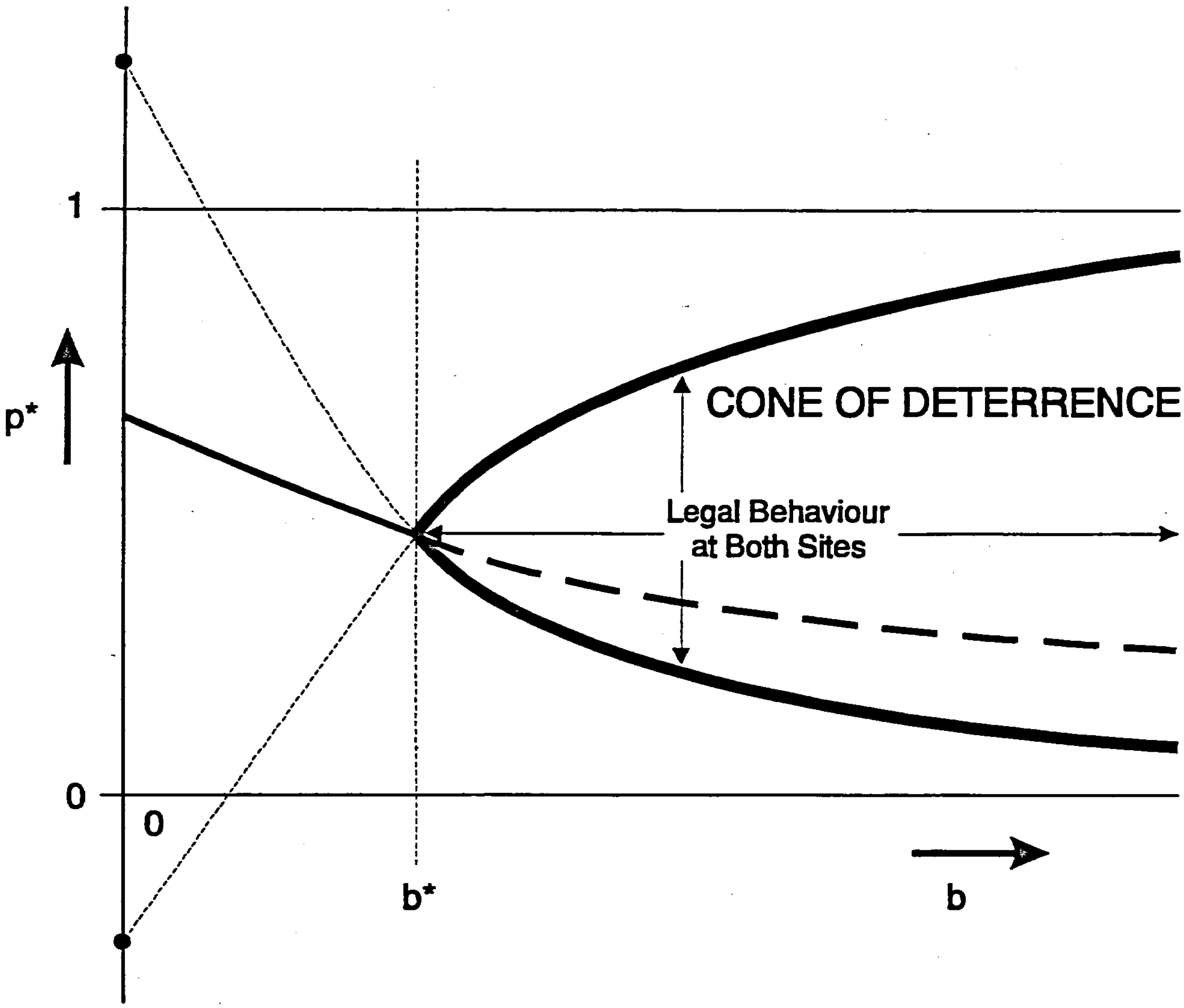
$$\frac{d_1}{b^* + d_1} \cdot \frac{1}{1-\beta_2} + \frac{d_2}{b^* + d_2} \cdot \frac{1}{1-\beta_2} = 1. \tag{3.25}$$

Furthermore, p^* satisfies (2.9), i.e.

$$\frac{d_1}{b + d_1} \cdot \frac{1}{1-\beta_2} \leq p^* \leq 1 - \frac{d_2}{b + d_2} \cdot \frac{1}{1-\beta_2}$$

whenever $b \geq b^*$. These calculations show that the optimal distribution of inspection effort achieves deterrence whenever deterrence is possible, as illustrated in Figure A3. Note also that, in general, p^* is determined only by the state's payoffs. For given "technical" parameters $1-\beta_1$ and $1-\beta_2$, and given values of d_1 and d_2, the common punishment given by (3.25), $b^* = b_1^* = b_2^*$, represents the minimum punishment level necessary to induce the state to behave legally.

Figure A3: Probability of Inspecting Declared Site vs. Penalty



We now present a second solution to this distribution of inspection effort problem, a solution which applies when detection probabilities have different properties.

Theorem 3.2 (Spreading of inspection effort)
Let $1-\beta_1(\varepsilon_1)$ and $1-\beta_2(\varepsilon-\varepsilon_1)$ have properties (3.8) and (3.9) and let them be strictly concave, i.e.,

$$\frac{d^2}{d\varepsilon_1^2}(1-\beta_1(\varepsilon_1)) < 0, \ \frac{d^2}{d\varepsilon_1^2}(1-\beta_2(\varepsilon-\varepsilon_1)) < 0 \text{ for all } \varepsilon_1 \text{ with } 0 \le \varepsilon_1 \le \varepsilon. \quad (3.26)$$

Then the equilibria (F^*, q_1^*, q_2^*) and the equilibrium payoffs E_I^* and E_S^* of the game described above are given by

$$F^*(\varepsilon_1) = \begin{cases} 0 & \text{for } \varepsilon < \varepsilon_1^* \text{ with } 0 < \varepsilon_1^* < \varepsilon \\ 1 & \text{otherwise} \end{cases} \quad (3.27)$$

where

(i) If $\beta_1^{-1}\left(\frac{b_1}{b_1+d_1}\right) + \beta_2^{-1}\left(\frac{b_2}{b_2+d_2}\right) < \varepsilon$ (3.28)

where $\beta_i^{-1}(\cdot)$ is the inverse of $\beta_i(\cdot)$ for $i = 1, 2$, then

$$\beta_1^{-1}\left(\frac{b_1}{b_1+d_1}\right) \le \varepsilon_1^* \le \varepsilon - \beta_2^{-1}\left(\frac{b_2}{b_2+d_2}\right), \ q_1^* = q_2^* = 0 \quad (3.29)$$

$$E_I^* = E_S^* = 0. \quad (3.30)$$

(ii) If $\beta_1^{-1}\left(\frac{b_1}{b_1+d_1}\right) + \beta_2^{-1}\left(\frac{b_2}{b_2+d_2}\right) \ge \varepsilon$ and $d_1 > d_2$; (3.31)

If $b_1 + d_2 > (b_1+d_1)\cdot\beta_1(\varepsilon)$, (3.32)

ε_1^* is the unique solution of the equation

$$-b_1 + (b_1+d_1)\cdot\beta_1(\varepsilon_1^*) = -b_2 + (b_2+d_2)\cdot\beta_2(\varepsilon-\varepsilon_1^*). \quad (3.33)$$

Furthermore,

$$q_1^* = \frac{G_2}{G_1+G_2} = 1 - q_2^*, \quad (3.34)$$

the equilibrium payoffs are

$$E_I^* = \frac{G_2}{G_1+G_2} \cdot (-c_1 + (c_1 - a_1) \cdot (1 - \beta_1(\varepsilon_1^*)) + \frac{G_1}{G_1+G_2} \cdot (-c_2 + (c_2 - a_2) \cdot (1 - \beta_2(\varepsilon - \varepsilon_1^*)), \quad (3.35)$$

where

$$G_1 = (c_1 - a_1) \cdot (1 - \beta_1(\varepsilon_1^*))', \; G_2 = -(c_2 - a_2) \cdot (1 - \beta_2(\varepsilon - \varepsilon_1^*))', \quad (3.36)$$

and

$$E_S^* = -b_2 + (b_2 + d_2) \cdot \beta_2(\varepsilon - \varepsilon_1^*) \;\; (= -b_1 + (b_1 + d_1) \cdot \beta_1(\varepsilon_1^*)). \quad (3.37)$$

If $b_1 + d_2 < (b_1 + d_1) \cdot \beta_1(\varepsilon)$, (3.38)

then

$$\varepsilon_1^* = \varepsilon, \; q_1^* = 1, \;\; q_2^* = 0 \quad (3.39)$$

$$E_I^* = -q_1 \cdot (1 - \beta_1(\varepsilon)) - c_1 \cdot \beta_1(\varepsilon) \quad (3.40)$$

$$E_S^* = -b_1 \cdot (1 - \beta_1(\varepsilon)) + d_1 \cdot \beta_1(\varepsilon). \quad (3.41)$$

Proof

(i) With (3.27) and (3.29) equilibrium condition (3.6) is identically fulfilled, whereas (3.7) becomes

$$0 \geq q_1 \cdot (-b_1 + (b_1 + d_1) \cdot \beta_1(\varepsilon_1^*)) + q_2 \cdot (-b_2 + (b_2 + d_2) \cdot \beta_2(\varepsilon - \varepsilon_1^*))$$

for all q_1, q_2 such that $q_1 + q_2 \leq 1$. This inequality is always fulfilled if and only if

$$-b_1 + (b_1 + d_1) \cdot \beta_1(\varepsilon_1^*) \leq 0$$

$$-b_2 + (b_2 + d_2) \cdot \beta_2(\varepsilon - \varepsilon_1^*) \leq 0$$

which is equivalent to (3.29).

(ii) Using (3.27), (3.32) and (3.33), equilibrium conditions (3.6) and (3.7) reduce to

$$q_1^*(-c_1 + (c_1 - a_1)(1 - \beta_1(\varepsilon_1^*)) + q_2^*(-c_2 + (c_2 - a_2)(1 - \beta_2(\varepsilon - \varepsilon_1^*))$$
$$\geq \int_0^{\varepsilon} [q_1^*(-c_1 + (c_1 - a_1)(1 - \beta_1(\varepsilon_1)) + q_2^*(c_2 + (c_2 - a_2)(1 - \beta_2(\varepsilon - \varepsilon_1))] \, dF(\varepsilon_1) \quad (3.42)$$

for all distributions F on $[0, \varepsilon]$, and

$$q_1^* \cdot \left(-b_1+(b_1+d_1)\beta_1(\varepsilon_1^*)\right)+q_2^* \cdot \left(-b_2+(b_2+d_2)\beta_2(\varepsilon-\varepsilon_1^*)\right)$$

$$\geq q_1 \cdot \left(-b_1+(b_1+d_1)\beta_1(\varepsilon_1^*)\right)+q_2 \cdot \left(-b_2(b_2+d_2)\beta_2(\varepsilon-\varepsilon_1^*)\right) \qquad (3.43)$$

for all q_1, q_2 and that $q_1+q_2 \leq 1$. The second inequality is satisfied by ε_1^* as defined by (3.33); it exists because of (3.31) and (3.32). Furthermore, using (3.34) and dividing both sides by $G_1 \cdot G_2/(G_1+G_2)$ shows that the first inequality is equivalent to

$$\frac{1-\beta_1(\varepsilon_1^*)}{(1-\beta_1(\varepsilon_1^*))'}-\frac{1-\beta_1(\varepsilon-\varepsilon_1^*)}{(1-\beta_2(\varepsilon-\varepsilon_1^*))'}$$

$$\geq \int_0^{\varepsilon}\left[\frac{1-\beta_1(\varepsilon_1)}{(1-\beta_1(\varepsilon_1^*))'}-\frac{1-\beta_2(\varepsilon-\varepsilon_1)}{(1-\beta_2(\varepsilon-\varepsilon_1^*))'}\right]dF(\varepsilon_1)$$

for all F. Now, because of (3.26), the function

$$H(\varepsilon_1) = \frac{1-\beta_1(\varepsilon_1)}{(1-\beta_1(\varepsilon_1^*))'}-\frac{1-\beta_2(\varepsilon-\varepsilon_1)}{(1-\beta_2(\varepsilon-\varepsilon_2^*))'} \quad \text{for } 0 \leq \varepsilon_1 \leq \varepsilon$$

is strictly concave in ε_1 for $0 \leq \varepsilon_1 \leq \varepsilon$, and satisfies

$$\left.\frac{dH(\varepsilon_1)}{d\varepsilon_1}\right|_{\varepsilon_1=\varepsilon_1^*} = 0.$$

Therefore, this inequality is fulfilled for all distributions F. If (3.38) holds, then it can be shown immediately that the solution (3.39) satisfies the Nash conditions.

∎

Consider now the general problem of guarenteeing legal behaviour of the state in equilibrium, i.e., $q_1^* = q_2^* = 0$. Whereas the Nash condition (3.6) is identically fulfilled, (3.7) is given by

$$0 \geq \int_0^{\varepsilon}[q_1 \cdot (-b_1+(b_1+d_1)\beta_1(\varepsilon_1))+q_2 \cdot (-b_2+(b_2+d_2)\beta_2(\varepsilon-\varepsilon_1))]\,dF^*(\varepsilon_1) \qquad (3.44)$$

for all q_1, q_2 such that $q_1+q_2 \leq 1$, where F^* is the equilibrium distribution of the IAEA's inspection effort. Now (3.44) is equivalent to

$$0 \geq q_1 \cdot \left[-b_1+(b_1+d_1)\cdot\int_0^{\varepsilon}\beta_1(\varepsilon_1)\,dF^*(\varepsilon_1)\right]+q_2 \cdot \left[-b_2+(b_2+d_2)\cdot\int_0^{\varepsilon}\beta_2(\varepsilon-\varepsilon_1)\,dF^*(\varepsilon_1)\right]$$

for all q_1, q_2 with $q_1+q_2 \leq 1$. This is true if and only if

$$\int_0^{\varepsilon} \beta_1(\varepsilon_1)\,dF^*(\varepsilon_1) \le \frac{b_1}{b_1+d_1} \quad \text{and} \quad \int_0^{\varepsilon} \beta_2(\varepsilon-\varepsilon_1)\,dF^*(\varepsilon_1) \le \frac{b_2}{b_2+d_2}. \tag{3.45}$$

The left-hand sides of these two inequalities are the *expected probabilities of no detection* with respect to the distribution F^*. This means that the state will behave legally if the IAEA selects any distribution F^* such that the expected probabilities of no detection satisfy the two inequalities (3.45). To illustrate, let $\overline{\beta_1} = \int_0^{\varepsilon} \beta_1(\varepsilon_1)\,dF^*(\varepsilon_1)$ and $\overline{\beta_2} = \int_0^{\varepsilon} \beta_2(\varepsilon-\varepsilon_1)\,dF^*(\varepsilon_1)$.

Then each type of violation is deterred provided

$$1-\overline{\beta_i} \ge \frac{d_i}{b_i+d_i}, \quad (i = 1, 2)$$

a condition strikingly similar to (1.2). This is the situation illustrated in Figure 3 of the text. Theorems 3.1 and 3.2 have identified the optimal $F^*(\cdot)$ in two special cases defined by the properties of $\beta_1(\cdot)$ and $\beta_2(\cdot)$.

Since (3.45) places no special requirements on the two detection probability functions, it follows that the solution (3.14) of Theorem 3.1 is also a solution under condition (3.26), and vice versa. For practical purposes, it is reasonable to use the theorems as presented here: Since the IAEA never knows precisely whether the conditions for legal behaviour are satisfied, it should use the effort distributions for the illegal case, according to the analytical forms of β_1 and β_2; if the legal case does apply, then no distribution of effort can improve on the distribution optimal for the illegal case.

Finally, some comments on the assumptions made in the foregoing analysis of Problem 3 are appropriate here. The modelling allows the players' values to be site-dependent; for instance, an undetected violation at a declared site may have a different value from one at an undeclared site. But it has been assumed that inspection processes at the two types of site are completely independent — no inspection ever provides any information about possible violations at the other type of site.

Certain aspects of the shapes of the functions linking detection probability and inspection effort (Figure A1) can only be guessed at. It can be assumed, at least to a first approximation, that all such functions are increasing, for additional inspection effort must always be useful. Some inspection problems are known for which the detection probability function is convex, at least for small values of inspection effort. However a law of diminishing returns is inevitable, so that in every case the detection probability function must become concave for sufficiently high levels of inspection effort. This suggests that for "severe" problems it is Theorem 3.2 that applies, so the spreading of inspection effort should be the norm rather than the exception when resources are constrained.