

ECONOMICS OF FORCE GENERATION AND STRUCTURE

Ugurhan G. Berkok
Department of Politics and Economics
Royal Military College of Canada

March 2010

If the first step in understanding the demand for defence is to understand the determination of the defence budget, surely the second step is to understand how the budget is allocated in order to generate a certain defence force structure (Treddenick [1999]). The defence budget as revelation of demand for defence is determined in competition with other government expenditure items. A government's ability to raise fiscal revenues or, in other words, taxpayer-citizens' willingness to finance government expenditures is one of the two major determinants of the defence budget, the other one being the perceived national security threat. This latter may not be exogenous in the sense that threats are co-determined strategically with allies and adversaries.

As for the defence budget allocation, force generation continues to remain on the demand side since the particular force structures chosen are responses to threats and allowed by given budgets (Smith [1995]). In fact, as we will see in detail below in the section where demand is linked to potential environments¹, defence is similar to a firefighting force in that it is not needed until a fire breaks out yet investment must be made and the force be readied in advance of any need for intervention. This chapter is hence organized with a view to describing the links between the force element generations all the way up to the defence budget.

1. Defence output production

A defence force element can be loosely defined as a collection of operational platforms. This collection may or may not be homogeneous. For instance, a tank, a fighter jet or a submarine together with their corresponding operating and support personnel constitute operational platforms. Evidently, new equipment like unmanned aerial vehicles and landmine disposal robots may have no operating personnel aboard yet they are operated, just as regular aircraft, by personnel stationed at a distance. The order of the day still being thus manned platforms, personnel may be located, aboard or at a distance.

A fundamental defence resource allocation problem is that of choosing the equipment intensiveness of platforms and force elements. The equipment intensiveness of a particular platform depends on what is called putty-clay technologies whereby, at the design stage, equipment can be made more or less equipment intensive. However, once manufactured, equipment follows a fixed proportions technology whereby a given personnel operates it. The equipment-intensiveness (capital-intensiveness) of a force element will depend on the composition of its range of platforms. If this composition tilts towards the use capital-intensive platforms the whole force element will be capital-intensive.

¹ Grimes & Rolfe [2002] introduced an innovative way to link force generation to environments. These latter are, of course, rooted in threats and threat assessments.

We will illustrate the equipment intensiveness choice with a landmine clearance example. Landmine clearance can be carried out by personnel equipped with properly-trained dogs, metal detectors, and sticks. However, any of these methods is slow and outright dangerous to personnel. Moreover, new generations of plastic mines are immune to detection by metal detectors. New equipment, from robots to remote controlled vehicles, are being built and used. Vehicles equipped with flails; vehicles taking air samples and locating mines by chemical analysis and using GPS (the global positioning system) or specially trained dogs identifying explosives; laying a bag filled with fuel-air explosives over a suspect area and then detonating it to set off mines; thermal and radar imaging; using chemical and biological sensors; and, finally, magnetic and sonar detectors. Some of these techniques may evolve to become sophisticated enough to permit their use from aircraft so as to revolutionize clearance in terms of speed and area coverage, both dimensions being fairly restricted at present.

For example, a mine detector requires one mine-disposal trained operator whereas an EROC (expedient route opening capability) team² is a complicated combination of personnel and sophisticated equipment producing the same output. Of course, the equipment intensiveness of the EROC team exceeds the simple combination of a metal detector. At the force element level, the equipment intensiveness may again become variable because, as defined above, a force element may be structured more or less equipment intensive by a choice of more or less equipment-intensive platforms. This may even be true in the short run as a task force can be structured by different platforms provided, of course, various platforms are available upon demand. For instance, the debate on whether to replace Canadian Air Force's CF-18 Hornets with the new Joint Strike Fighter (Doyon [2005]) or any fighter jet versus unmanned aerial vehicles (UAVs) falls into this context. The co-presence of platforms would allow a continuing choice of equipment intensiveness whereas a replacement of CF-18s by UAVs would imply a switch from current levels of equipment intensiveness with CF-18s (pilots and support personnel) to that associated with UAVs.

The implied resource allocation problem consists of choosing the cost-minimizing combination of equipment and personnel for every given level of expected output. Of course, this hypothesis has to be qualified in two respects. The first qualification relates to the time frame of decisions. Field commanders take decisions within such short periods of time that cost-minimization may not be the order of the moment. Also, "shock and awe" or overpowering the opponent as a tactic may require excessive use of force, beyond what would have otherwise been optimal. Second, the theoretically feasible choice set may not be available to field commanders due to equipment and/or personnel shortages or, alternatively, commitments in other theatres. This second qualification is taken up below in the analysis of US 7th Air Force example.

A basic analysis

² See description at <http://www.casr.ca/bg-husky-mdv.htm#mdt>. An EROC team is composed of three different vehicles (Cougar, Buffalo and Husky) and, beyond the vehicle drivers, includes equipment operators.

Of course, the mine clearance example is further complicated by the multidimensionality of output in that speed and area covered have to be synthesized into a single output measure. If, however, we simplify the problem to production by the two inputs, equipment and personnel, the production choices can be represented by bundles of inputs along various isoquants, as y^0 , y^1 and y^2 in Figure 1 below. Thus the isoquants represent the relationship between the output and the inputs in the following ways. First, the production function $y = G(L,K)$ is graphically represented by isoquants such that higher isoquants correspond to higher outputs. Second, along the same isoquant, many³ different input bundles allow the production of the same level of output and, hence, input substitutability is always a major preoccupation in military operations research as elsewhere in the analysis of production.

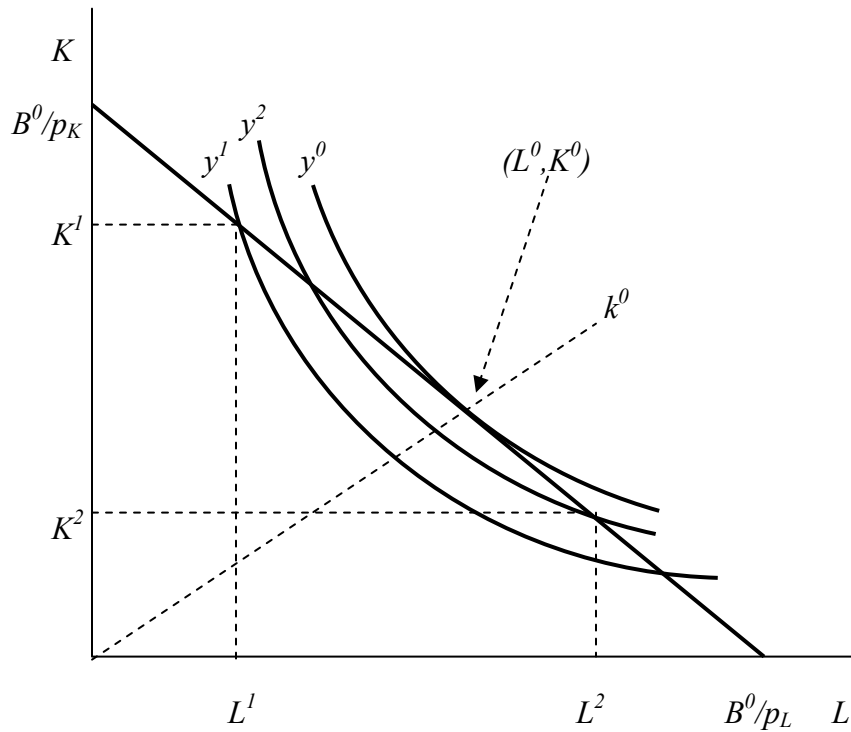


Figure 1 Output maximization and equipment-personnel tradeoffs

Especially noteworthy, in our running example, is the fact that the cost of personnel includes the actuarially-calculated risk premium to personnel life and limb. The diagram illustrates the choice of equipment-personnel combination for the given budget B^0 . Thus, given the budget B^0 , a defence planner's output maximization problem would be to maximize the output obtainable.⁴ Three input bundles, (L^0, K^0) , (L^1, K^1) and (L^2, K^2) , are

³ Although, for pedagogical purposes, the isoquants are drawn as smooth curves, in reality they consist of a finite number of bundles as implied by the given number of techniques of producing the same output.

⁴ Or, alternatively, given the output y^0 , the expenditure level B^0 is the cost-minimizing budget for y^0 .

represented on the diagram with the corresponding outputs $y^0 > y^1 > y^2$. The choice of input bundle (L^0, K^0) evidently results in the highest output from the given budget B^0 .

We note that the choice of (L^2, K^2) bundle would correspond to what is generally known as the *arm-the-man* approach⁵ where, typically, oversized personnel costs a big chunk of the budget, leaving the force with the low equipment-personnel ratio $k^2 = K^2/L^2$ due to residual funds left, hence a lower output $y^2 (< y^0)$ than under the output-maximizing ratio k^0 . The opposite case is where the input bundle (L^1, K^1) is what is known as *man-the-arm* approach⁶ with a high equipment-personnel ratio $k^1 = K^1/L^1$. This case typically leads to undermanning, undermining of training and education of personnel and a lower output $y^1 (< y^0)$ than under an output-maximizing choice (Treddenick [1998]).⁷

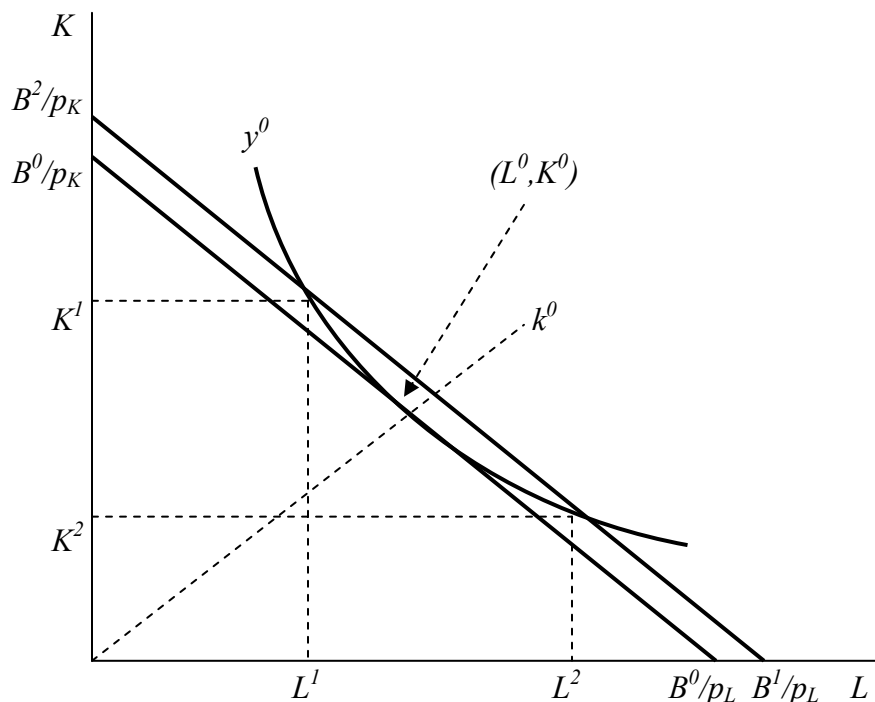


Figure 2 Cost minimization and equipment-personnel tradeoffs

Another perspective on the same phenomenon would be to minimize the cost of achieving the same output. As considered in Figure 2 above, the defence planner in our

⁵ This typically arises during force downsizing or budget crunch processes when governments let equipment purchases slip rather than sacrificing personnel.

⁶ One curious example is the observed Israeli phenomenon, possibly as a signaling distortion in that top generals, soon-to-become politicians, want to show off the shiny armour as evidence of strength in decision-making (Lipow & Feinerman [2001]).

⁷ As Treddenick [1998] pointed out, the Canadian defence resource allocation had typically consisted of leaving capital as residual rather than under-manning. However, the high tempo of peace support operations in the 1990s and the Afghanistan mission with speedy equipment acquisitions seem to have reversed the trend.

running example aims choosing that equipment-personnel ratio so as to achieve the minimal cost of producing the same output. Thus, on the diagram, the desired output y^0 is achieved at the minimum cost bundle (L^0, K^0) with an outlay of B^0 whereas the alternative bundles, (L^1, K^1) and (L^2, K^2) , achieving the same output, would cost more at $B^1 (> B^0)$.

A real example

Of course, beyond platform choices, defence output can be produced with different platforms combined. The following example (Hildebrandt [1999]) demonstrates, furthermore, choices imposed by constraints on the availability of inputs as well as by choices over the use of the same input over different tasks. The example is one of choosing the bundle composed of helicopter gunships and of fighter jets to be used by the American 7th Air Force for the interdiction campaign Commando Hunt against Vietnamese convoys in 1970. The output, measured by various truck-movement sensors, is the convoy netput (i.e. reduction in throughput) over the Ho Chi Minh trail system and

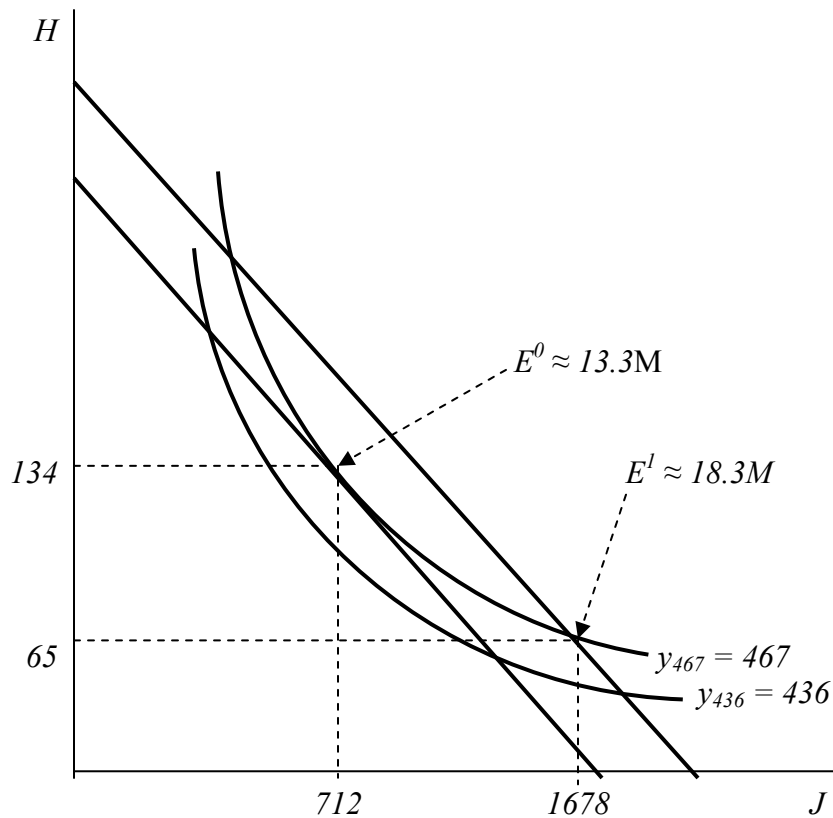


Figure 3 Helicopter gunships vs. fighter jets

supplying the troops in the South. The 7th Air Force used two types of interdiction sorties, gunship-team vs. fighter-attack sorties. Thus inputs J and H are the two types of sorties⁸ and the output y the netput reduction. The analysis conducted using real observations collected during the interdiction campaign over 6 months⁹ shows that, over a week on average, the 7th Air Force achieved an actual netput of 436 whereas with an optimum use of fighter-attack sorties, the netput would have risen to 467. Hildebrandt [1999] presents empirical estimates of the production function $y = G(J,H)$. The estimates are mapped onto the Figure 3 above. The input bundle (1678,65) was used the six-month period under consideration. Given that a helicopter-gunship sortie cost 52,300\$ and a fighter-attack sortie of 8,900\$, the output $y_{467} = 467$ would have required the total expenditure $E^1 \approx 18.3$ million dollars. However, alternatively, the same output would have been attained with the bundle (712,134) at the expenditure $E^0 \approx 13.3$ million dollars. Of course, this reduction in costs would have required a reallocation of fighter-attack sorties amongst their different uses. We note, in particular for use below, that isoquants also yield the relative value of inputs. For instance, in this case, the 7th Air Force would be prepared to give up 966 fighter jets in order to acquire 69 gunships for the same output. This imputes an average value of 14 fighter jets to a gunship.¹⁰

Beyond some econometric concerns¹¹, this particular analysis demonstrates the use of economic analysis in the derivation of efficient battlefield plans. In fact, the reason why the 7th Air Force didn't adjust towards more efficient bundles in Commando Hunt was because they had limited access to further helicopter gunships (Hildebrandt [1999]). A similar analysis is conducted in Rohlfs [2006] where World War II German battles are analyzed to estimate the production isoquants in order to infer the implicit valuations of soldiers' lives. The Rohlfs study estimates the isoquants based on personnel and equipment using the equipment-intensiveness variations in battle groups used during hundreds of battles. Since the slope of the isoquant yields the value of personnel in terms of equipment and the equipment values can easily be calculated, the slope of the isoquant at the actual equipment-intensiveness levels used in the battles yields the implicit value assigned to soldiers' lives. Rohlfs finds that soldiers' implicit life valuations thus obtained didn't differ significantly from their actuarial estimates. This does, ironically, suggest that American commanders didn't unduly risk human lives.

Efficiency vs. effectiveness

⁸ As explained in Hildebrandt [1999], the fighter sorties could be directed against various targets (e.g. against trucks and storage areas, against lines of communication, and as close air support). This target distinction effects the netput positively if fighter-attack sorties are efficiently combined for their three distinct uses. The Figure 2.d.3 shows the outcomes of the efficient case and the actual, inefficient case.

⁹ The author was an intelligence officer and personally collected the information.

¹⁰ Of course, a higher value when the Force is really short of gunships when they number a lean 65 and much lower when their numbers approach the optimum at 134.

¹¹ Econometric issues may be raised include the following. (a) The choice of variables (e.g. close air support could have been replaced by data on ground troop operations supported). (b) The daytime data may not have been appropriate as Vietnamese convoys moved mostly overnight. (c) Weapons dropped may have been better predictors. (d) Was the choice of terrain under consideration correct?

The two examples above, the replacement of CF-18s with UAVs and the substitution between fighter jets and helicopter gunships, show that defence planners do indeed carefully plan, at least ex ante when force elements are generated, for equipment intensiveness. It is interesting to note that, in the above UAV example, defence planners do indeed take into account such facts as UAVs' "... high accident rate--several orders of magnitude greater than that of manned aircraft." (Adams [2005]) In fact, these accidents would factor into the unit cost of UAV force element considered.¹² Of course, the reliability of the aircraft would appear in the output obtainable and, as such, affect decisions regarding platform and/or force element choices.

The **efficiency** requirement on ex ante force generation choices is, however, categorically superseded by battlefield **effectiveness** requirement. Regardless of what options are available to the field commanders, their decisions must be guided by the effectiveness of the force in order to win the next battle. In this sense, for instance, the fighter jet vs. helicopter gunship choice ceases to be a choice, the commanders having to do with what they have in order to achieve the output targeted if it is indeed achievable. In terms of Figure 3 above, this would have two implications. First, the available fighter jet sorties have to be correctly allocated to different tasks. Then, especially, helicopter maintenance has to be meticulously upheld so as to achieve the targeted number of sorties as the loss of helicopter sorties would be particularly harmful in terms of opportunity costs.

The efficiency vs. effectiveness distinction is rooted in the technological choices underlining investment in equipment with embodied technology. Once chosen, the technology is hard to disentangle from equipment throughout its useful life.¹³

This section developed the framework for understanding efficient equipment-personnel choices for platforms as well as force elements. In fact, platforms still constitute the building blocks of any force element. Despite the theorization that the new defence forces are network-centric¹⁴ rather than platform based, platforms continue to be the building blocks of the networks. What makes the network-centric warfare unique is the limitless range of communication tools and the resulting coordination provided by existing electronic technologies. The next section moves up to battle group and task force levels in order to understand the choice faced by a defence force facing various tasks and social preferences over the tasks.

2. Single-task defence output

Defence output is produced from force elements. These are the different task forces or battle groups within a defence force. Task forces can be integrated multi-service or just

¹² "In the USAF study, "mishap" implies damages ranging from more than \$20,000 through complete destruction of the air vehicle." (Adams [2005])

¹³ In growth literature, this would correspond to putty-clay technologies.

¹⁴ Just like the now-forgotten Revolution in Military Affairs and Global War on Terror, new descriptions tend to find their way, temporarily, into the defence language, just like in the fashion world, without users being able to see their poor descriptive and theoretical values.

based on a single service. This representation can even be taken to illustrate the case where force elements are the traditional services. We represent the defence production function as

$$d = D(F,f) , D_F, D_f > 0, D_{Ff} > 0$$

i.e. the defence output increases in force elements f and F and the contribution of a particular element is positively reinforced by an increase in the other one.

Moreover, the defence production function may exhibit the force multiplier effect¹⁵ which translates the idea of synergies. Thus, if two force elements are built together, the resulting defence output would exceed the total output of the two operated separately. A simple example would be to consolidate some military helicopter unit together with a search and rescue unit, especially if similar aircraft were in use. The consolidated support and maintenance would yield more air time to the helicopters than if the two units were set up independently. A rigorous definition of the force multiplier effect is given as

$$D(F,f) > D(F,0) + D(0,f), F,f > 0$$

where the stand-alone outputs are $D(F,0)$ and $D(0,f)$ whereas the combined operation's output is given by $D(F,f)$.

Of course, decisions are also constrained by budgets. Although a general formulation could be useful, we'll represent the defence budget constraint as follows:

$$C(F,f) = [K_F + c_FF] + [K_f + c_ff].$$

For a given defence budget B , the opportunity cost of expanding force element f at the expense of F is given by c_f/c_F , the slope of the defence budget constraint, as in Figure 4. We also note that force element costs exhibit scale economies¹⁶ due primarily to equipment intensiveness of any defence force.

Given the defence budget B , the maximum defence output can be achieved at

$$\frac{D_f(f,F)}{D_F(f,F)} = \frac{c_f}{c_F}$$

i.e. the marginal rate of technical substitution (i.e. the slope of the isoquant) or the marginal willingness to pay for f in terms of F is equal to the opportunity cost of f in terms of F (i.e. the slope of the isocost line or the budget line). This optimum is illustrated in Figure 4 below.

¹⁵ This phenomenon is similar to scope economies where two activities, if undertaken together, cost less than if undertaken separately.

¹⁶ Scale economies exists if average cost of an activity falls with the increase in the activity. Although the concept relates correctly to the organizational costs outweighing gains from specialization as the activity level increases and differs from spreading the overhead, it has come to be used in this latter context as well.

An intuitive way to explain the above condition is as follows. The left-hand side represents the defence planner's subjective opportunity cost, i.e. the amount of F it wants to sacrifice for an extra unit of f while preserving the output level. The right-hand side represents the objective opportunity cost of f in terms of F , i.e. how much F must be sacrificed in order to obtain one more unit of f . Since both sides are computed at a given feasible bundle (f, F) along the budget constraint, an equality would indicate optimality whereas if the subjective opportunity cost exceeds the objective one, it indicates a willingness to increase f by sacrificing F . This is illustrated at bundle (f_1, F_1) in Figure 4.

Another, perhaps more intuitive way of understanding this choice is to rewrite the condition as

$$\frac{D_f(f, F)}{c_f} = \frac{D_F(f, F)}{c_F}$$

where the two sides are, respectively, the marginal contribution per dollar spent of force element F and that of f . If, say, the left-hand side is bigger, as at (f_1, F_1) , then it would pay to transfer a dollar from spending on force element F to f because the fall in defence output due to a smaller F will be outweighed by an increase in output due to the extra dollar spent on f . This condition simply states the obvious in that if a dollar is better spent on a particular component of the defence force then it should be till it no longer is. That is the case at (f_0, F_0) .

In this framework (Berkok [2006]), the defence production corresponds to a single defence task d and resources are so allocated as to accomplish the task through the force elements f and F . Furthermore, we observe from the diagram that the balanced allocation of the budget B induces a higher output than any of the two specialization cases, i.e.

$$d_0 = D(f_0, F_0) > \max \{D(0, B/c_F), D(B/c_f, 0)\}.$$

Thus, as in Figure 4 below, the defence multiplier effect outweighs the scale economies emanating from potentially large fixed costs, K_f or K_F . The implication is that a balanced force is optimal.¹⁷

¹⁷ Figure 4 exhibits an interesting phenomenon due to the existence of significant fixed costs, K_f or K_F . We note that, at $(B - K_f)/c_f$, F is completely non-operational in the sense that even its infrastructure based on fixed costs isn't available. However, at $(B - K_f)/c_f$, the infrastructure for F is available yet the force element F is still non-operational.

otherwise if force elements are simultaneously operated and the specialization by shutting down one and avoiding the large fixed cost associated with the eliminated force element

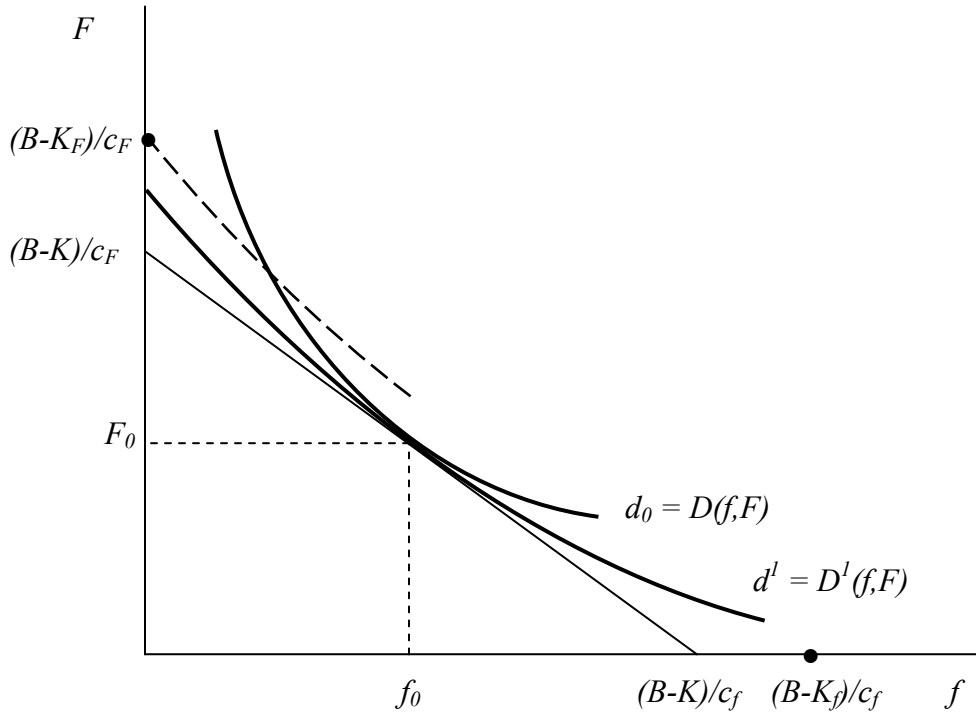


Figure 5 Specialized force optimal with technology D^1

whereas spreading the fixed costs of the operational force element.¹⁹ Theoretically, with a budget of B ,

$$D^0(f_0, F_0) > \max \{D(0, (B-K_F)/c_F), D((B-K_F)/c_f, 0)\}.$$

However, under technology D^1 , the force multiplier is weak and specialization with $(B-K_F)/c_F$ becomes optimal. As we will see below, there may be other considerations, such as a multi-task defence planning environment, complicating the above decision process.

¹⁹ The Joint Strike Fighter project may be seen in this perspective. The astronomical R&D costs associated with a versatile airframe made the new jet aircraft so attractive that many force elements may now be compressed into one, rather than diversifying to many aircraft. Of course, the resulting large scale economies has, hence, proven attractive. Similarly, of course, F-22 Raptor has numerous capabilities that allow its use in multiple tasks.

3. Multi-task defence output

We can now generalize the defence output concept to **multiple tasks** performed by several force elements (Grimes & Rolfe [2002]). This enlarged framework takes into account defence activities in many theatres, domestic and abroad, and addresses the defence resource allocation problems emerging as a result. For example, in the Canadian context, the recent reorganization of the defence force into Canada Command (for homeland security) and the Expeditionary Force (for peace support operations) is an excellent example of multi-tasking.

The defence output can now be represented as a composite good consisting of M , the domestic component or territorial defence, and N , the international component which could be a peace support operation or another task related to global stability such as disaster relief or emergency assistance abroad. These tasks can be performed using the two force elements as

$$m = M(f,F) \text{ and } n = N(F).$$

In this formulation, the force element F constitutes the sole force component for task N whereas both force components, f and F , have to be combined for the task M .²⁰ Moreover, the defence planner has preferences over the fulfillment of the two tasks. These preferences can be formulated by the defence objective function $d = D(N,M)$ with the usual properties of increasing in its arguments and also characterized with diminishing returns to each task. The defence budget constraint remains as in the above section, i.e.

$$B = [K_F + c_F F] + [K_f + c_f f]$$

and can, for a given budget, be expressed as

$$F = G(f) = \frac{B - K}{c_F} - \frac{c_f}{c_F} f$$

where c_f/c_F is the opportunity cost of force element f in terms of F and the total fixed investment into the defence force is given as $K = K_f + K_F$.

Using this linear relationship $F = G(f)$ we can rewrite the task production functions $m = M(f,F)$ and $n = N(F)$ as

$$m = M(f,G(f)) \text{ and } n = N(G(f))$$

and, eliminating f between the two production functions $m = M(f,G(f))$ and $n = N(G(f))$, the production possibility boundary $PPF(n,m)$ can be derived, as drawn below in Figure 6. The equation of the PPF can be simply denoted as $m = H(n)$.

²⁰ Of course, it could just as well be the converse, with international tasks requiring both force elements.

A few explanatory remarks on $PPF(n, m)$ are in order. First, the positively sloped section corresponds to existence of *synergies*²¹ for task forces n and m . This portion of the $PPF(n, m)$ corresponds to an initial reallocation of resources from force element f towards F . Since F is now just being built from scratch and f contracting from an overexpansion, F surely increases thus increasing n . However, a reduction in over-expanded f is more than compensated in terms of output m by initial increases in non-existent F . Thus, both task forces may expand. However, the marginal expansionary effect of F falls with further injections of F and the marginal expansionary effect of f increases with falling f . At point $(H^{-1}(m^{max}), m^{max})$, m reaches a maximum where the force element bundle is given as $[f(m^{max}), F(m^{max})]$. Second, in addition to these synergy economies, by virtue of the fact that task n only requires the force element F and that

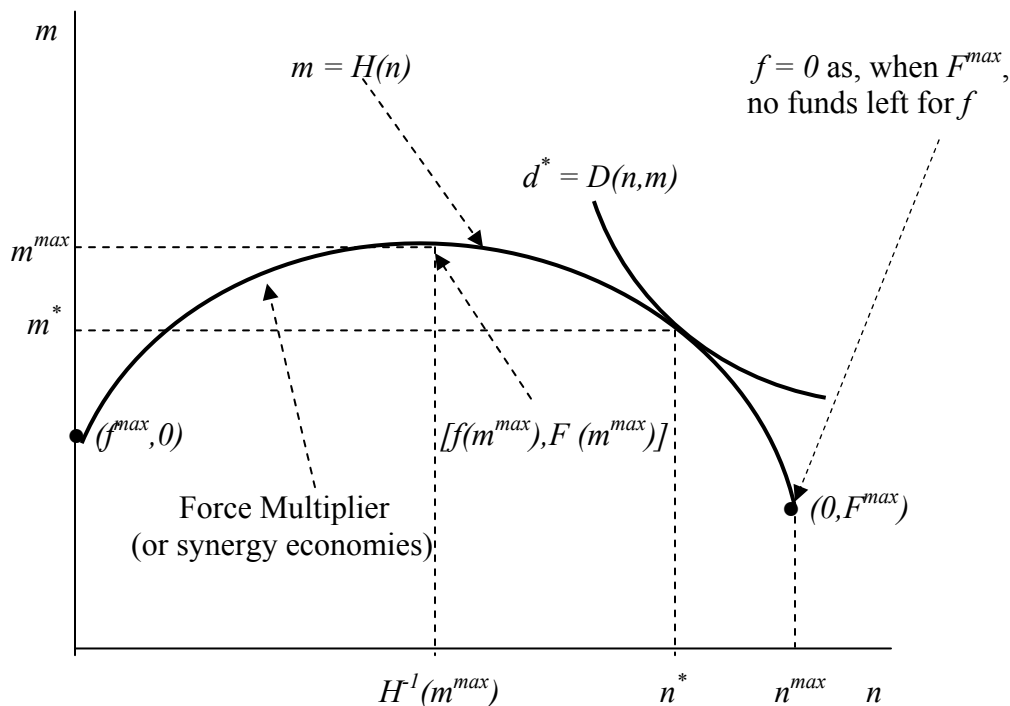


Figure 6 Force generation and organization

there are significant fixed costs (K_f), potentially there exist scale economies to the generation of task force n provided the international task function $n = N(F)$ doesn't exhibit strong diseconomies. Yet, if there are overall economies, they are exhausted at $(H^{-1}(m^{max}), m^{max})$ and a defence resource allocation tradeoff sets in. Third, the maximal international task force achievable may well be discretely to the right of n^{max} on Figure 6 if f is completely shut-down (i.e. $K_f = 0$).

We have now derived the feasible task force combinations for the country in question to obtain $PPF(n, m)$ or, simply, $m = H(n)$. Yet, every country, based on its foreign and

²¹ Or force multiplier (Hurley [2004]).

defence policies, makes choices over feasible defence force configurations to generate its specific force. Foreign and defence policies imply certain preferences over what the two force elements are assigned to in terms of domestic and international tasks. The relevant objective function representing such preferences is denoted as $d = D(n, m)$ in Figure 6 above. The general defence resource allocation problem can be written as

$$\begin{aligned} \max_{\{n, m\}} \quad & D(n, m) \\ \text{s.t.} \quad & m = H(n) \end{aligned}$$

and its solution is (n^*, m^*) and the underlying force elements are operated at (f^*, F^*) . If, however, the problem is modified with the domestic task becoming the overriding concern, i.e. the objective function simply becoming $D(n, m) = m$, the solution becomes $(H^{-1}(m^{max}), m^{max})$. Of course, if the sole concern is the domestic defence needs, then resources are so allocated that the domestic task force output is maximized.

4. Balanced vs. specialized force over security environments

The correspondence between strategic environments and force structures in terms of balanced vs. specialized forces is a complicated relationship. First, it depends on force generation per se, as we have seen above, depending on force generation technologies and associated costs²². Second, it depends on strategic environments and country preferences over tasks. And, finally, it depends on the risk-taking behaviour of the defence planner. This last issue will also be examined in this section.

As introduced above in the introduction to this chapter, defence force is quite similar to a firefighting force. For instance, a firefighting force in a rainy environment is structured quite differently than in a windy area with dry forests. Accordingly, a major investment like firefighting aircraft is not undertaken in the rainy area. The fundamental difference²³ between the firefighting and defence forces, however, is that the firefighting force faces a single task with a fairly certain occurrence over a given time period. Thus although it is structured for a particular environment, this latter is fairly stable. A defence force is typically built with the expected occurrence of certain environments. If potential strategic environmental shifts are taken into account then the similarity to a firefighting force ends because environmental uncertainty is invoked.

Since the force, just like the firefighting force, has to be generated in advance or, in other words, has to be ready, defence planners have to take into account two major factors. The first is the set of probable environments. For example, the Cold War is gone and, by all

²² In this context, force surge-ability may at times be pivotal. When regular forces are strained under operational tempo, reserve forces may be drawn upon. However, this surge-ability depends on the combat readiness level of reserves.

²³ Of course there is an order of magnitude between investments into defence force equipment and into particular firefighting forces but, by and large, such a difference remains a clear and simple problem.

accounts, isn't coming back. A Cold War force structure is not a reasonable defence policy alternative. Just the same, a force structured to defend the homeland is typically not suitable as an expeditionary force. We note that a particular environment requires the corresponding force structure. The second factor is the estimation of the likelihood that a particular, reasonable environment may arise. Given the information available, the defence planner must form subjective probabilities. For example, though highly improbable, Cold war would have necessitated a corresponding force structure but its improbability implies that a decision to generate a force for Cold War is highly unlikely.

These two factors, the discernment of reasonable environments and the estimation of their likelihoods, combine to invoke corresponding force structures. That is, the force structure that maximizes the ex ante expected security benefits obtainable ought to be chosen. The choice is ex ante precisely for the reasons stated at the beginning of the chapter. A force cannot be built upon an environment arising. Rather, the force must be combat ready for environments.

We will now concentrate on the interaction of technological constraints, environments and budgets for force generation. The technological constraints, beyond the availability of platforms, refer to the minimum outlays necessary to build viable force elements. And, those force elements have to correspond to emerging environments.

Since the force generation determinant of force specialization derives from technological requirements, force elements can not be operated at all levels for all technologies. Often, military technology imposes a minimum. For instance, two fighter jets do not constitute a squadron, neither do two ships form a viable naval task force nor two tanks do a squadron. Thus, lower bounds exist for a number of platforms to constitute a force element. Figure 7 below illustrates the effect of this constraint. Given the defence budget B_0 , the defence planner would have chosen the balanced force (f^0, F^0) . However, the force element f^0 isn't feasible as it falls below the technological minimum f^{min} . This constraint imposes a choice between the specialized force structure $(0, B_0)$ and the over-budget balanced force $(f^{min}, B_1 - f^{min})$. Thus the option value of the balanced force is equal to $(B_1 - B_0)$, i.e. if the planner is averse to a specialized force, the defence budget must be supplemented by $(B_1 - B_0)$ to generate a balanced force.

Thus, in general, shrinking (expanding) budgets reduce (increase) options. The technological minimum f^0 constrains force generation especially in small countries either to spend more to preserve options with a balanced force or to downsize to a specialized force if defence budgets are hard.

Moreover, the strategic environment naturally affects the choice between generalist (balanced) and specialist (specialized) forces. It is common knowledge that operational readiness is key for a defence force to be effective just as for firefighters, search & rescue teams, police, ambulance and similar emergency services. However, the defence force differs significantly in terms of investment required and the speed at which it can adapt to a new strategic environment due to its sheer relative size. Thus force structures must be readied for a set of environments as there may not be sufficient response time to activate

or transform a force for an environment outside the original set. Although, in general, readiness is geared to the likelihood of a particular environment, there may be a case for a more safety-first approach in force generation. Whereas the former is a Bayesian approach to decision-making, the latter is rooted in minimax criteria.

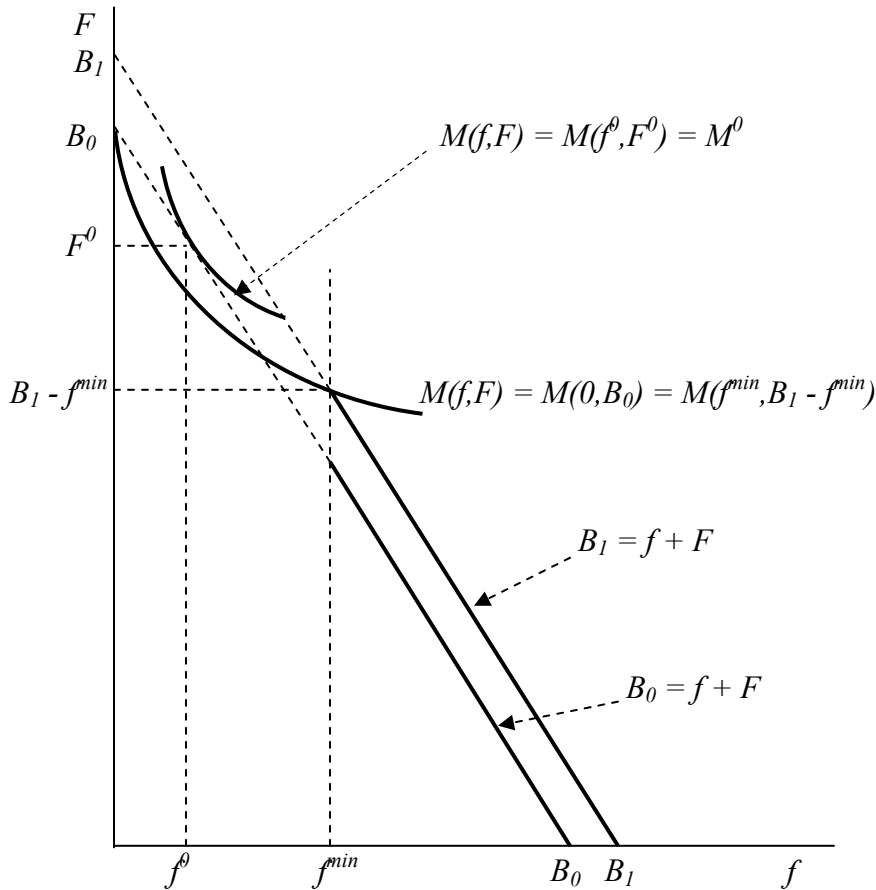


Figure 7 Specialized vs. balanced force

The minimax approach corresponds to very high risk aversion. To illustrate this interaction between environments and force generation, we continue with the simple objective function $M(f, F) = (af)^{1/2} + (bF)^{1/2}$ dealt with in the Appendix and with the minimum size requirement denoted as f^{min} for force element f . The actual values of parameters a and b reflect the environmental requirements. For example, an environment where a is high and b low would boost the marginal product of force element f thus requiring a strong f .

Let us consider two environments, E1 and E2, where, had it not been for the minimum size force element constraint $f > f^{min}$, the optimum would entail a balanced force structure with $f = f^0$ as in Figure 7 above, i.e. the force element f is desired to be operated

at a fairly low but positive level. However, given certain values of parameters a and b , the environment E1 is assumed to yield the optimal force structure $(0, B_0)^{24}$ implied by

$$M^1 = M^1(0, B_0) > M^1(f^{min}, B_0 - f^{min}).$$

This means that, given the parameters a and b , the defence production isoquants would have been flatter than in Figure 7. Alternatively, in environment E2, parameters a and b , with different values, yield steeper isoquants and the force structure $(f^{min}, B_1 - f^{min})$ implied by

$$M^2 = M^2(f^{min}, B_0 - f^{min}) > M^2(0, B_0).$$

The emerging decision problem can be summarized in the following decision matrix

	$f = 0$	$f = f^{min}$
E1	$M^1(0, B_0)$	$M^1(f^{min}, B_0 - f^{min})$
E2	$M^2(0, B_0)$	$M^2(f^{min}, B_0 - f^{min})$

Table 1

The relationship between environments and force generation is, in general, such that the defence planner normally faces environmental uncertainty whereas the force must be structured regardless. The problem is identical to investment under uncertainty. The minimax criterion would have the planner compare the two off-main-diagonal entries, i.e. safest bets in environments. Thus, if it so happens that $M^2(0, B_0) > M^1(f^{min}, B_0 - f^{min})$ then, ironically, the specialized force is optimal across environments. For instance, if the environment required a balanced force, the minimum force element requirement combined with the impossibility of speedy adjustment, the planner would have to do with a specialized force. Therefore environmental uncertainty does not necessarily imply a balanced force even when the planner is completely risk averse.²⁵

Finally, if the defence planner is a Bayesian decision-maker, then its evaluation of the likelihood of each environment²⁶ will guide the choice between specialized and balanced force structures. For instance, if the probability of E1 is given by a fraction α and the planner's risk-aversion is quantified by a utility function $U(m)$ then its objective function is given as

$$U^e(M(f, F)) = \alpha U(M^1(f, F)) + (1-\alpha) U(M^2(f, F)).$$

²⁴ For simplification purposes, we set $K_F = K_f = 0$ and $c_F = c_f = 1$.

²⁵ That the defence planner is completely risk averse is, of course, too strong an assumption but the framework illustrates the irony that a completely risk-averse decision-maker puts all its eggs in one basket.

²⁶ This must, of course, be complemented by a critical listing of possible environments. More on Bayesian decision-making in defence will be discussed below.

The problem is illustrated in Figure 8 below with the value to the planner of defence output m is given as $U(m)$. We maintain the assumption $M^2(0, B_0) > M^1(f^{min}, B_0 - f^{min})$ that drove the result that a specialized force dominated a balanced one across environments when the planner adopted a minimax approach. However, when it adopts a Bayesian approach with $\alpha \approx .4$, the decision is reversed and a balanced force is chosen across environments. To see this, consider the planner's expected value functions, the first for the specialized and the second for balanced force respectively:

$$U^e(M(0, B_0)) = \alpha m^1 + (1-\alpha) U(M^2) \text{ and } U^e(M(f^{min}, B_0 - f^{min})) = \alpha U(M^1) + (1-\alpha) m^2$$

where

$$m^2 = M^2(f^{min}, B_0 - f^{min}) > M^2(0, B_0) = M^2$$

$$m^1 = M^1(0, B_0) > M^1(f^{min}, B_0 - f^{min}) = M^1.$$

These values are all illustrated below in Figure 8.

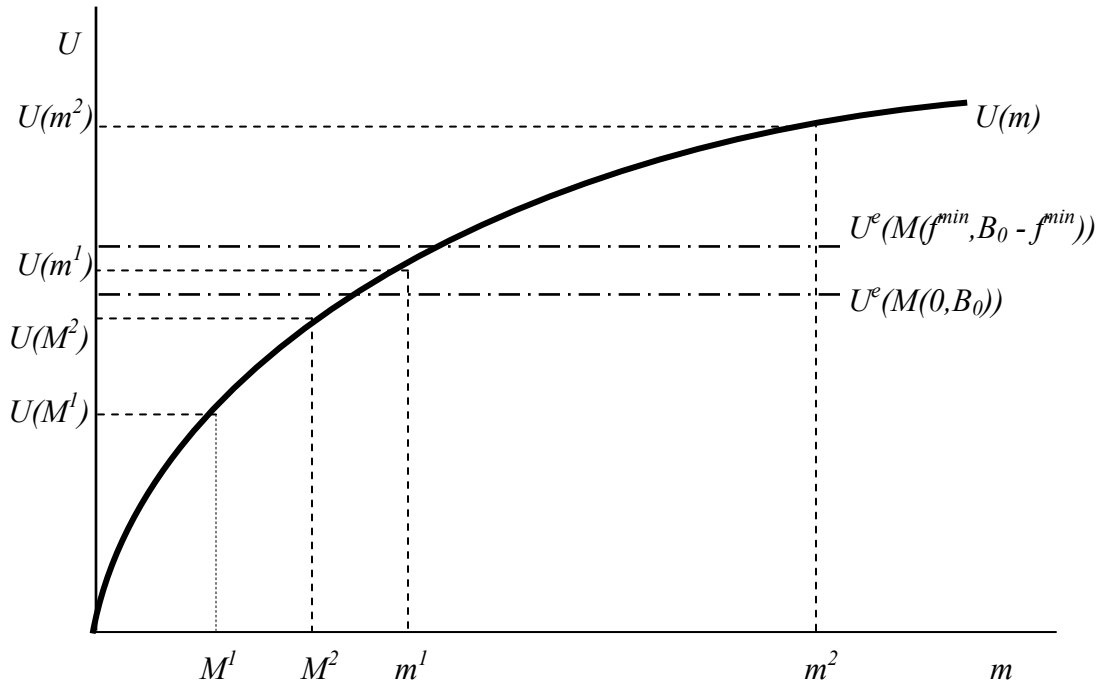


Figure 8 Bayesian defence planner

Thus, as can be seen in Figure 8, the balanced force is optimal over all environments as

$$U^e(M(f^{min}, B_0 - f^{min})) > U^e(M(0, B_0)).$$

This happens because, unlike its extreme risk aversion in the minimax approach, the planner now weighs the likelihoods of environments. Given that E2 is likelier and the balanced force that would be optimal in E2 yields a high output at m^2 , the expected value of choosing the balanced force outweighs that of the specialized force. To reiterate, although the specialized force produces a fairly stable output across environments, the higher likelihood of E2 combined with the balanced force's exceptionally high output in E2 outweighs the balanced force's wide output variability and, hence, the associated risk. Of course, a lower likelihood of E2 and/or a smaller balanced force output in E2 might reverse the decision.

The likelihood of an environment used in this type of decision-making derives from strategic assessments combined with planners' subjective views on future environments. Moreover, the future environments themselves are scenarios that may or may not materialize. The pessimistic outlook represented by the minimax criterion, above in Table 1, is thus not a realistic decision-making tool. Rather, realism imposes Bayesian decision-making criteria on defence planners. As explained in the example above (with Figure 8), in a very simple context, defence planners have to construct future scenarios and assign likelihoods to scenarios in order to make force generation decisions today for readiness tomorrow, should any of the scenarios materialize.

Common to any defence decision-making framework is the realization that force generation decisions are taken *ex ante*, without a precise knowledge of the impending environment. The first important consideration, consistent with risk aversion, is that the current force generation ought to be consistent with as many future scenarios as possible. This is, of course, an economic efficiency requirement in that such a multi-use force element would exploit would correspond to the exploitation of scope economies. Needless to add, such scenarios ought to be relevant ones. In addition, the defence planner must assure that particular force elements generated are consistent with various scenarios, take into account the fact that certain platforms acquired may be consistent with many force elements and thus create synergies into the future (Camm et al. [2009], Staker [2003]).

Scenarios about the future are really demands for defence or, in other words, future risks assessed. The risk in question is the threat inherent in any scenario and the defence force's preparedness for the particular threat. Thus the risk would be lower if the force elements built can respond to more threats. We must emphasize that threat assessment logically precedes force generation. That is, threat assessment determines demand to which the defence planner responds by allocating resources to force generation.

In the presence of a future threat, some questions become relevant. First, can one affect the likelihood of a scenario? A positive answer to this first question typically means that the resource allocation from defence to, say, diplomacy can avert the threat over time and hence reduce the demand for defence. Second, if the scenario is realized and the threat becomes reality, how likely is open conflict? (Camm et al. [2009]) The question addresses the fact that, if parties to the threat may not want to escalate the conflict to open conflict, demand for defence subsides. Third, what are the consequences of not

responding to the threat? Capitulation may be more acceptable than escalation and conflict. If so, demand for defence is lower. Therefore, answers to these three questions will affect defence planner's assessment of threats emerging and the damage to national interests given the policies in place and the resources available. Under scarcity of resources, the planner may be forced to trade off one type of potential damage to the national interest for another, as in $(D(n,m))$ above in this chapter (Camm et al. [2009]).

5. The economics of readiness

Readiness is a key element of force generation. Consistent with the assessment of threat, readiness has to pass the two criteria of capability and intent. Both criteria have to be met by a force element to become credible. Capability is generated when force-in-barrack is translated into actual force-in-theater. However, actual force's readiness and, essentially, intent are crystallized in actual training exercises. Therefore, training and exercising against potential threat scenarios constitute the final phase of force generation, with the pre-deployment training being the culmination of the process. Force readiness can be coined as force projection and, more in line with strategic analysis, as signaling intent where signaling means communicating intent credibly.

The intrinsic economic problem associated with readiness is not so much with the above process of force generation but, rather, with choices with budgetary implications. Readiness pitches equipment (K) and human resources (L) expenditures against operations and maintenance expenditures (O&M), the first two generating the force-in-barrack and the latter raising capability into force-in-theater or readiness. Despite some linguistic overlap between capability and readiness, the distinction can be encapsulated into force-in-existence heightened to force-in-readiness by the precise definition of perceived threat and the set of training exercises required to generate readiness.

We illustrate the readiness problem below in Figure 9. The allocation of defence budgets to generate the force-in-existence F can be construed as combining platforms with specialized personnel. However, the transition towards force-in-theater requires the training input T . The precise combinations of T and F derive from military technology, in the sense that, for a given defence budget, readiness $r = R(T,F)$ can be maximized by a proper choice of the two inputs.

The readiness isoquants represented below in Figure 9 convey the tradeoff between capabilities generated and training exercises required. They are required inputs into readiness although the relationship allows substitution.

Figure 9 illustrates the choice.²⁷ When an extra bit of training per budget allocation generates as much readiness as that lost by the diversion of that budget allocation away from force-in-existence, the optimal budget allocation is reached and no further increase in readiness is feasible with the given budget. This level of readiness r^0 is given at (T^0, F^0) .

²⁷ This graph is drawn following the same principles as above in Figures 1 through 5. Substitutions between inputs are discreet in reality. Yet, graphs above treat them as continuous for expositional ease.

The same graph can be used to illustrate the constrained choices defence forces may face at times training may be restricted, not so much for budgetary reasons but for insufficiency of lead times. Thus if, for the given budget B^0 , training is restricted to T^l , the resulting readiness falls to r^l at best if force-in-existence can be boosted to F^l . If force-in-existence cannot, for lack of short-run flexibility, be adjusted and remains at F^0 then readiness may fall to r_{LOW} .

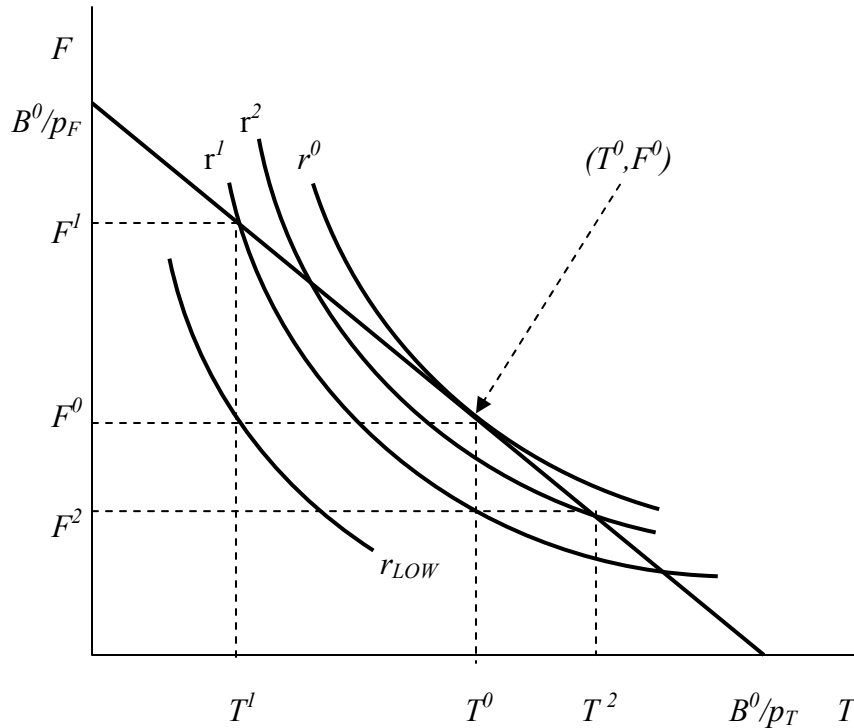


Figure 9 Training-capability tradeoffs

Of course, one can argue that readiness is generated by the very existence of capabilities built due to a need of deterrence. To use the firefighting analogy, whereas the frequently and easily tested water-pumping equipment is sufficient to generate readiness against ordinary house fires, readiness against chemical fires requires somewhat different and more involving training exercises. Defence force readiness exhibiting considerably higher complexity, if not in terms of equipment at least in terms of strategic adversaries, training gains in prominence to generate readiness.

A related force generation requirement is force rotation. One of the consequences of up-tempo operations is the regular rotation of troops in theater. Recuperation of military personnel is a necessary component of force generation and sustainment. The direct resource implication of this requirement is that for a given size deployment, the total active troop numbers to sustain such deployments easily amount to a multiple troops in deployed. By implication, the deployable troops must also exhibit readiness. This requires resources increase with operational tempo.

6. Summary and discussion

This paper sheds light on the recent restructuring of the Canadian Forces. Canada had not fundamentally restructured its defence force since 1960s until recently when various commands were created beyond the traditional services. “In spite of Ottawa's desire to promote international peace and stability alongside the United States and the United Nations, Canada's minimalist approaches to defence spending and capital expenditures are undermining the long-term viability of the Canadian Forces' (CF) expeditionary and interoperable capabilities. Two solutions to this dilemma present themselves: increased defence spending or greater force structure specialization.” (Lagassé [2005]) In fact, the restructuring of the Canadian defence force has followed a specialization path towards strengthening the force's expeditionary component. True, this restructuring was partly rushed by Canada's involvement in Afghanistan but it is consistent with Canada's chosen role in committing to peace support operations. There exist various difficulties associated with such restructurings such as organizational and traditional service based resistance to change as well as intrinsic coordination problems. However, a further problem is generated by the fact that inter-service forces impose different readiness problems, from training to budgetary allocation problems. These are deferred to later analysis.

References

- Adams, C. [2005], “Input/Output: Learning from UAV mishaps”, Avionics Magazine, October 1
- Berkok, U.G. [2005], “Specialization in defence forces”, Defence and Peace Economics 16, 191-204
- Camm, F., L. Caston, A.C. Hou, F.E. Morgan & A.J. Vick, Managing Risk in USAF Force Planning, RAND Corporation, 2009
- Chun, C. [2003], “Some comments on ‘The Military Production Function’”, unpublished Note
- Doyon, C. [2005], “Replacing the CF-18 Hornet: Unmanned combat aerial vehicle of Joint Strike Fighter?”, Canadian Military Journal 6(1), 33-39
- French, B. [2006], “The Business of Land-Mine Clearing”, Economics of Peace and Security Journal 1(2), 54-57
- Grimes, A. & J. Rolfe [2002], “Optimal defence structure for a small country”, Defence and Peace Economics 13(4), 271-286
- Hildebrandt, G.G. [1999], “The military production function”, Defence and Peace Economics 10, 247-272
- Hurley, W. [2004], “A clarification of the concepts of force multiplier and returns to force scale”, unpublished note
- Lagassé, P. [2005], “Specialization and the Canadian forces”, Defence and Peace Economics 16, 205-222

- Lipow, J. & E. Feinerman [2001], "Better weapons or better troops?", Defence and Peace Economics 12, 271-284
- Owen, N. [1994], "How Many Men do Armed Forces Need?", Defence and Peace Economics 5, 269-288
- Rohlf, C. [2006], "The Government's Valuation of Military Life-Saving in War: A Cost-Minimization Approach", American Economic Review, May 2006 (Papers and Proceedings), 96(2), 39-44
- Smith, R. [1995], "The Demand for Military Expenditure", H&S [1995]
- Staker, R.J. [2003], "Stochastic Simulation Methods for Force Level Defence Systems Design", SimTecT Papers, Simulation Industry Association of Australia, http://www.siaa.asn.au/library_simtect_2003.html#category-2391114409
- Treddenick, J.M. [1999], "Distributing the Defence Budget: Choosing Between Capital and Manpower", in D.L. Bland, (ed.), Issues in Defence Management, Queen's Policy Studies Series #12, 1999

Appendix 1

If the domestic defence is the overriding concern then $d = D(n,m) = m$ and the resource allocation problem reduces to

$$\begin{aligned} \max_{\{n, m\}} \quad & m = M(f, F) \\ \text{s.t.} \quad & F = G(f) \end{aligned}$$

where the constraint becomes, simply, $B = [K_F + c_F F] + [K_f + c_f f]$. Further simplification of the problem by assuming the fixed costs away and by eliminating the coefficients by properly choosing the force element units yields the very simple constraint $B = F + f$.

Case 1 If defence planner's objective function is given as $M(f,F) = (af)^{1/2} + (bF)^{1/2}$ (or $M(f,F) = f^a F^b$) then the solution yields $f = [a/(a+b)]B$ and $F = [b/(a+b)]B$. Thus the defence budget is allocated in proportion to the relative marginal products of force elements in homeland security production. This solution yields a balanced force.

Case 2 If defence planner's objective function is given as $M(f,F) = af + bF$ (perfect substitutability of force elements) then the solution yields $f=B$ if $a > b$ and $F = B$ if $a < b$. Thus the defence budget is wholly allocated to the force element whose constant marginal product is higher. Thus a specialized force results in each case.

Case 3 If defence planner's objective function is given as $M(f,F) = \min \{ af, bF \}$ (perfect complementarity of force elements) then the solution yields $f = [a/(a+b)]B$ and $F = [b/(a+b)]B$. Thus a balanced force results.

Appendix 2

This appendix analyzes an interesting facet to the optimal air force generation example considered earlier in the article. As jet aircraft can be used for various tasks, the issue arises as to how to allocate the potential number of sorties amongst tasks. Evidently, if this use is not optimal, the potential output obtainable from force elements isn't achieved.

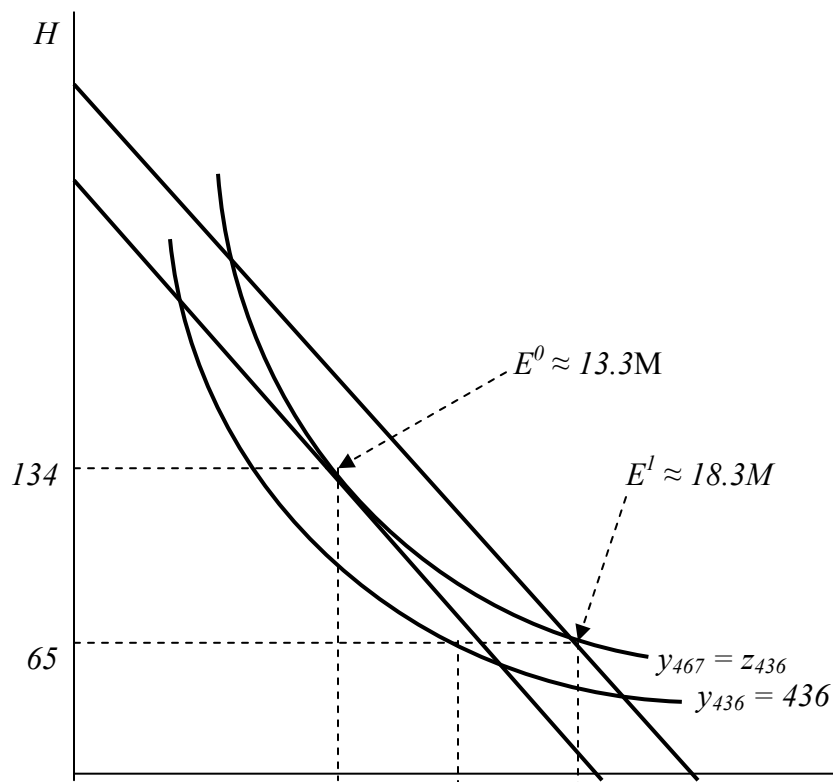
Below, in Figure A2-1, the output along the isoquant $y_{436} = 436$ represents what can be achieved upon a misallocation of available jet aircraft that flew 1678 sorties. Yet, the same output is achievable with $(1678-m)$ sorties when the aircraft is optimally used. These can be expressed using the production function $y = F(J,H)$ as follows:

$$y_{467} = 467 = F(1678, 65)$$

where J^* designates the optimal combination of various tasks assigned to jet aircraft while $j^* (< 1678)$ designates the optimal combination of a smaller number of sorties yielding

$$y_{436} = 436 = F(j^*, 65) = G(1678, 65) = z_{436}$$

where the production function $G(J,H)$ derives from an suboptimal combination of jet aircraft for different tasks.



712 j^* 1678 J

Figure A2-1 Helicopter gunships vs. fighter jets